Quantifying Hypomimia in Parkinson Patients Using a Depth Camera

Nomi Vinokurov[1], David Arkadir[2], Eduard Linetsky[2], Hagai Bergman[2], and Daphna Weinshall[1]

[1] Hebrew University, School of Computer Science and Engineering, Jerusalem 91904, Israel
[2] Hebrew University, Hadassah Medical School, Jerusalem 91120, Israel
Nomi.Vinokurov@mail.huji.ac.il

Abstract. One of Parkinson's disease early symptoms is called hypomimia (masked facies), and timely detection of this symptom could potentially assist early diagnosis. In this study we developed methods to automatically detect and assess the severity of hypomimia, using machine learning tools and a 3D sensor that allows for fairly accurate facial movements tracking. To evaluate our prediction of hypomimia score for participants not included in the training set, we computed the score's correlation with hypomimia scores provided by 2 neurologists. The correlations in 4 conditions were $0.84, 0.69, 0.71, 0.70$. This should be compared with the correlation between the somewhat subjectives scores of the two neurologists, which is 0.78. When training classifiers to discriminate between people who suffer from hypomimia and people who do not, the area under the curve of the corresponding Receiver Operating Characteristic curves in the same 4 conditions is $0.90 - 0.99$. These encouraging results provide proof of concept that automatic evaluation of hypomimia can be sufficiently reliable to be useful for clinical early detection of Parkinson-related hypomimia.

Key words: Parkinson's disease, hypomimia, 3D camera, facial expressions, affect prediction

1 Introduction

Parkinson's Disease (PD) is the second most common neurodegenerative disorder with a prevalence rate exceeding 100/100,000 among all American population and 1,588/100,000 among population over the age of 65 . This statistic might underestimate the problem because PD diagnosis is complicated. Since age is the single most important factor for PD and population is growing older, the prevalence rate could further increase in the not too distant future [1, 2]. PD symptoms include tremor, rigidity and loss of muscle control in general, as well as cognitive impairment. The difficulty in reliable PD diagnosis has inspired researchers to develop decision support tools relying on algorithms aiming to differentiate healthy controls from people with PD [3, 4].

Hypomimia is a cardinal sign of the disease often presented in its early stages. The syndrome is characterized by a marked diminution of expressive gestures of the face, including brow movements that accompany speech and emotional facial expressions. Punctuation brow movements - very brief ($\sim$ 50 ms) contractions of the muscles of the upper face that occur during speech and appear to add semantic emphasis, are often absent. Additionally, hypomimia is commonly manifested in only one side of the face [5].

Research has shown that PD patients have lower expressivity ratings than the normal population while watching video clips, and that PD patients differ also in the frequency of smiles while watching a series of cartoons and in the degree of mouth opening while smiling [6, 7]. In those studies ratings were performed with the assistance of human judges, while in our study we aim to measure these quantities automatically from video recordings, in order to compute an accurate prediction of hypomimia. [8] investigated hypomimia in patients suffering from depression using ultrasonic markers placed on participants faces. New wearable technology could enable home monitoring of patients, but since the number of sensors that can be put on the patients face is limited, the quality of assessing hypomimia severity is reduced.

The Unified Parkinson's Disease Rating Scale (UPDRS) is the most common scale used in clinical studies in order to follow the longitudinal course of PD. UPDRS defines hypomimia levels as follows:

0. Normal
1. Minimal hypomimia, could be called "poker face" (healthy subjects might get this score)
2. Slight but definitely abnormal diminution of facial expression
3. Moderate hypomimia; lips parted some of the time
4. Masked or fixed facies with severe or complete loss of facial expression; lips parted 1/4 inch or more

Quantitative assessment of hypomimia could assist early diagnosis of the disease, which could in the future (and with the development of new procedures) enable better treatment and slow down the progression of the disease. A home stationed application that enables quantitative assessment of hypomimia, with no need to meet the neurologist, would allow for better evaluation and monitoring of the treatment and could improve life quality of Parkinson's patients. Recent advances in computer vision allow for the reliable tracking of facial movements using simple devices that could be stationed in one's home. In this study we used a common depth camera - PrimeSense Carmine 1.09, which delivers depth video and audio information.

In the following we present an algorithm that utilizes depth sensor data to detect and scale hypomimia. In Section 2 we describe the data collection with some details of the recording procedure. In Section 3 we describe an algorithm that will assess hypomimia severity. In Section 4 we present the results, discussing the correlation between our algorithm's predictions and both neurologists scores, and presenting Receiver Operating Characteristic (ROC) curves for hypomimia detection.

2 Methods and Materials

Technology Overview. In recent years $3D$ sensors developed as part of the growth of the gaming market. Some of these sensors come with software support to track facial movements fairly reliably. In our study we chose to use the Carmine 1.09 camera developed by PrimeSense. The sensor depth acquisition is based on the "light coding" technology. The process codes the scene with near-IR light, light that returns distorted depending upon where things are. The solution then uses a standard off-the-shelf complementary metal-oxide-semiconductor image sensor to read the coded light back from the scene using various algorithms to triangulate and extract depth. The product analyzes scenery in 3 dimensions [9, 10].

To generate features we used $Faceshift$©, which is a commercial software that performs real time face tracking. The software gets as input data from the depth sensor with sampling rate of 19 Hz, and tracks points of interest on the participant's face. (see example in Fig. 1) After tracking it further analyzes facial movements and describes them as a mixture of basic expressions, as well as head orientation and gaze. Specifically, $Faceshift$ outputs the intensity level of 51 facial Action Units ($fs\text{-}AU$ signals) over time, including eyes (blink, squint, up, down, in, out), brows (up, down), jaw (forward, left, right, open), mouth (left, right, frown, smile, dimple), lips (stretch, close, open, up, down, funnel, pucker), chin (raise), cheek (squint), sneer, and puff [11, 12]. In the analysis below we use the $fs\text{-}AU$ signal as raw data to generate features for our prediction algorithm.

Data Collection Protocol. The study was approved by the ethical committee of Hadassah Medical center. 14 Patients ages 58 to 84 with varying levels of hypomimia, and 15 Controls ages 48 to 84, were recruited at Hadassah Medical Center in Jerusalem. Each participant was given a short overview of the experiment, and then gave a written informed consent for study participation in accordance with the Helsinki Declaration. Each participant also indicated their consent for the whole procedure to be videotaped. The $3D$ camera was positioned 50 cm from the subject's face, 10 cm above eyes level.

First, the participant went through a short training stage (for the benefit of $Faceshift$) which included presenting different facial expressions to the camera. Afterwards each participant was recorded during 5 different sessions. The sessions were: Answering 5 interview questions, watching photographs as slide show, watching a funny short movie involving cats, watching a short movie involving humans, and staring at the camera for 60 seconds. The most discriminantive results were obtained when using the recordings of the $4th$ session (watching a short movie involving humans), and therefore only these results will be presented.

Two Movement Disorder specialists (denoted DA and ED) rated each participant for hypomimia using the recordings of the $1th$ session (an interview with 5 questions). DA scored participants with integer and half values in the range $[0, 4]$; he stated that as a matter of procedure he would round up the hypomimia score when his perception is that the hypomimia level is not an integer value.

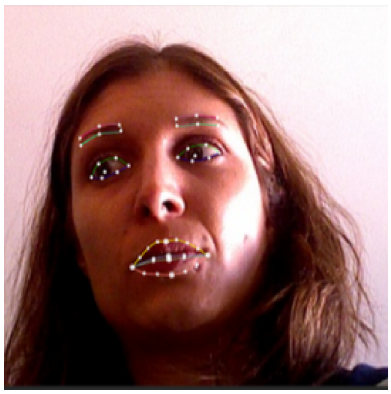


Fig. 1: Example of points tracking by $Faceshift$©

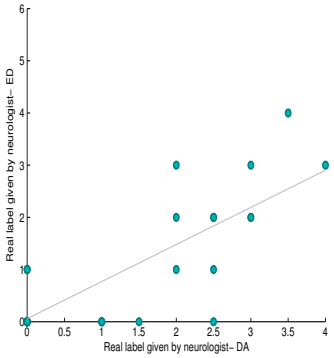


Fig. 2: Correlation of labels given by two neurologists, denoted DA and ED.

Neurologist ED scored all participants with only integer values in the range $[0, 4]$ as is the custom in such neurological evaluations.

Data Representation. The *fs-AU* signal was used to generate the following representation for each recording, whose components included: (i) 4 moments of each *fs-AU* (mean, variance, skew and kurtosis). (ii) 4 moments of bilateral *fs-AU* differences (left vs. right side of the face). These comparisons were employed since hypomimia is also characterized by asymmetrical facial expressions [5]. (iii) Correlations between every pair of *fs-AU* signals (50*51/2). (iv) Quantization of the *fs-AU* signal to 4 discrete values, which were chosen using k-means for each *fs-AU* signal. Using this discretization, we could compute for each *fs-AU* signal the number of changes (a transition from one discrete value to another discrete value) and the number of fast changes (a change of 2 discrete values or more).

Methodology. To learn a predictor, the data was divided into train and test sets using the Leave One Out (LOO) procedure, where the data of each left out participant was kept for testing, while the data of all other participants was used to train the linear regressor. This was repeated for each participant. Given labeled data by two different neurologists, for each learning session we trained two predictors, one for each neurologist. We then tested the predictor on the test data from both movement Disorder specialists, giving us 4 different results (see below in Section 4). For lack of objective hypomimia score, by defition our gold standard is the Pearson Correlation Coefficient (PCC) between the subjective scores of the two expert neurologists when based on the same recordings as our algorithm, which is 0.78 (see Fig. 2).

3 Predicting Hypomimia Level

We shall now describe the procedure to obtain a prediction for hypomimia from each recording, using the data representation described above.

Learning Method: Supervised Learning Using Linear Regression. Linear regression is typically used to model the relationship between a scalar dependent variable Y and one or more explanatory variables X. Since our challenge is to predict an ordinal value in the range $[0, 4]$, we used linear regression slightly modified to construct such a predictor.

We start by noting that the data representation described in Section 2 lies in high dimension, while the size of our training data is rather small. Therefore training always started with greedy feature selection as described next. Subsequently we modeled the relationship between the selected features and the given labels in each training set. Finally, this model was used to predict a continuous hypomimia score for the left out recording of the test participant.

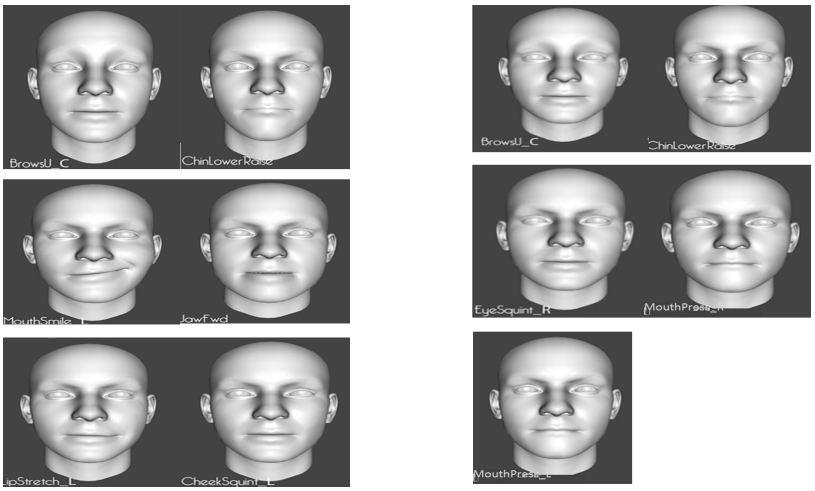
Note that this procedure outputs a continuous number, while the variable we aim to predict is integer following the UPDRS guidelines. Thus in the final step of the procedure, the output of the predictor is rounded to integer or half integer values. When the prediction value is lower than 0 or higher than 4, it is truncated to 0 or 4 respectively.

Feature Selection. The first step of the learning procedure reduced the dimensionality of the signal using forward greedy feature selection coupled with least squares regression [13]. The greedy selection procedure works as follow: in iteration i, the feature that mostly reduces the residual sum of squares (RSS) between the algorithm prediction and real labels is chosen, and improvement from the last iteration ($\Delta RSS = RSS_i - RSS_{i-1}$) is calculated. Features are added until the improvement is no longer statistically significant (under the null assumption that ΔRSS has a chi-square distribution with one degree of freedom).

When learning a predictor based on neurologist DA who typically used higher scores, we used all recordings in the train set with hypomimia score of 0 and all recordings with hypomimia score > 2, and ran forward greedy features selection using linear regression as the matched predictor. Similalry, when learning a predictor based on neurologist ED who used somewhat lower subjective scores, we used all recordings in the train set with hypomimia score of 0 and all recordings with hypomimia score ≥ 2.

Anecdotally, the Features that were selected by the greedy procedure using the data of each neurologist separately gave different results. When learning from DA, the features most often selected included the correlation between brows up movement and lower chin raise, the correlation between left side mouth smile and forward jaw movement, and the correlation between left side lips stretch and left side cheek squints (see Fig. 3a). When learning from ED, the features most often selected included the correlation between brows up movement and lower chin raise as above, the correlation between right eye squints and right side mouth press, and the mean value of the left side mouth press. In both cases,

between 2 and 3 features were selected at each iteration of the algorithm (see Fig. 3b).



(a) (b)

Fig. 3: Most frequently selected features: (a) when learning from DA, and (b) when learning from ED.

3.1 Learning Algorithm

We use the following notations: Let n denote the number of participants (in our case 29), $\mathcal{K}$ denote the set of participants whose scores were used to train the greedy feature selection method, and $|\mathcal{K}| = m$. Let s_i denote subject i, y_i denote the label given by the neurologist to participant i, $\hat{y}_i$ the prediction of hypomimia score generated by our algorithm, and r_i the quantized hypomimia integer score generated by our algorithm. Let F_i denote the set of features that were selected using $\mathcal{K} \backslash s_i$ as the train set. Let V_{ji} denote the features vector of subject j according to the features in set F_i.

Algorithm 1 Predict Hypomimia Severity

1: **for** $i = 1..n$ **do**
2: F_i = forward greedy features selection using $\mathcal{K} \backslash s_i$
3: Generate V_{ji} for $j = 1 \dots n$
4: W_i = Linear regression$(V_{ji}, y_j$ for $j \in \mathcal{K} \backslash s_i)$
5: $\hat{y}_i = W_i^T * V_{ii}$
6: $\hat{r}_i = \text{round}(\hat{y}_i)$
7: **end for**

4 Results

Predicting the Severity of Hypomimia

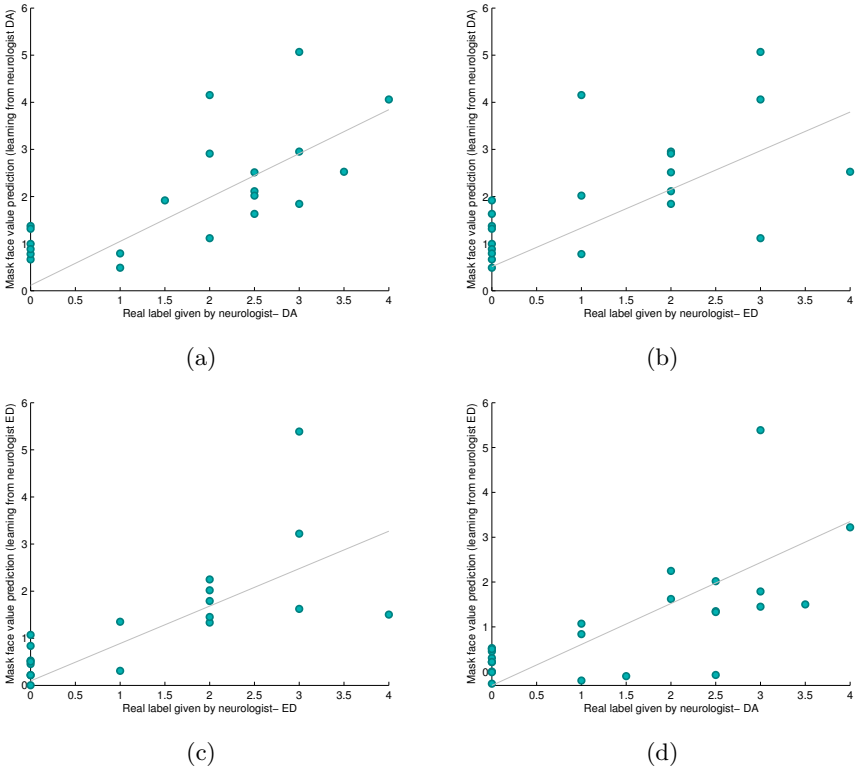


Fig. 4: Correlations between the scores given by each of 2 neurologists and our algorithm predictions in 4 conditions: (a) Scores of neurologist DA in the train data used for training, scores of neurologist DA in the test data used for testing. (b) Scores of neurologist DA in the train data used for training, scores of neurologist ED in the test data used for testing. (c) Same as (a), training with ED and testing with ED. (d) Same as (b), training with ED and testing with DA.

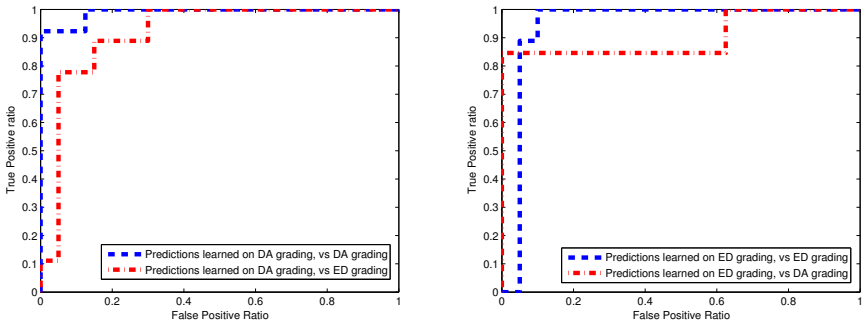
Using the algorithm described above, we predict a hypomimia score for each participant. We then test these predictions by correlating them with model scores. Specifically, recall that for each recording we have 2 predictions, based on two separate predictors trained for each neurologist. These 2 predictions are correlated with the scores of both neurologists (the one whose scores were used for training, and the other one), giving us 4 prediction graphs (see Fig. 4) and 4 correlation scores (see Table 1).

Table 1: Pearson Correlation Coefficient of our method's predictions with neurologists' scores in 4 conditions, each marked by a pair of initials. The first set of initial denotes the neurologist whose scores in the train data were used for training, while the second set of initials denotes the neurologist whose scores in the test data were used for correlation. In either case, correlation was computed with predicted values for unseen data from the test set. All correlation values are very significant $p < 3 * 10^{-4}$; The first row shows correlations based on the raw predictor values, while the second row shows correlations based on the integer predictor values. Since ED scored hypomimia with whole integer values, only the integer prediction values are shown when training on ED scores.

	DA - DA	DA - ED	ED - DA	ED - ED
Prediction	**0.784**	**0.623**		
Integer Prediction	**0.836**	**0.686**	**0.711**	**0.707**

Binary Detection of Hypomimia

The predictors can be used to discriminate between healthy individuals (score ≤ 1) and people who suffer from hypomimia (score > 1). This is a binary classification task. We plot Receiver Operator Characteristic (ROC) curves to evaluate our algorithm in this discimination task, see Fig. 5. The area under the curve (AUC) is used to measure success, see Table 2.



(a) (b)

Fig. 5: ROC curves of hypomimia detection using our algorithm's predictions. (a) Scores of neurologist DA in the train data used for training.(b) Scores of neurologist ED in the train data used for training.

Table 2: AUC of ROC curves from Fig. 5. The 4 conditions are described in the caption of Table. 1. All correlation values are very significant $p < 10^{-4}$.

	DA - DA	DA - ED	ED - DA	ED - ED
AUC	**0.990**	**0.917**	**0.904**	**0.944**

5 Discussion

We described a learning algorithm that detects and scores hypomimia with relatively high accuracy, when trained on other subjects. This kind of work may contribute to the goal of early detection of Parkinson, by providing an automatic tool which can be combined with other such tools to produce automatic scores correlated with symptoms of Parkinson. One attractive feature of the approach is its potential to provide a home-stationed diagnostic aid not requiring a trip to the neurologist's clinic. One drawback of our method is its reliance on the availability of a depth camera, which is less readily available to most people, and the use of the $Faceshift$© software which requires pre-training by all participants, a step which is not always straightforward for Parkinson patients.

Acknowledgements

This work was supported in part by the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI), and the Gatsby Charitable Foundations.

References

1. Wright Willis, A., Evanoff, B.A., Lian, M., Criswell, S.R., Racette, B.A.: Geographic and ethnic variation in parkinson disease: a population-based study of us medicare beneficiaries. Neuroepidemiology **34** (2010) 143–151
2. von Campenhausen, S., Bornschein, B., Wick, R., Bötzel, K., Sampaio, C., Poewe, W., Oertel, W., Siebert, U., Berger, K., Dodel, R.: Prevalence and incidence of parkinson's disease in europe. European Neuropsychopharmacology **15** (2005) 473–490
3. Tsanas, A., Little, M.A., McSharry, P.E., Spielman, J., Ramig, L.O.: Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease. Biomedical Engineering, IEEE Transactions on **59** (2012) 1264–1271
4. Guo, P.F., Bhattacharya, P., Kharma, N.: Advances in detecting parkinsons disease. In: Medical Biometrics. Springer (2010) 306–314
5. Rinn, W.E.: The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. Psychological bulletin **95** (1984) 52
6. Simons, G., Pasqualini, M.C.S., Reddy, V., Wood, J.: Emotional and nonemotional facial expressions in people with parkinson's disease. Journal of the International Neuropsychological Society **10** (2004) 521–535

7. Katsikitis, M., Pilowsky, I.: A study of facial expression in parkinson's disease using a novel microcomputer-based method. Journal of Neurology, Neurosurgery & Psychiatry **51** (1988) 362–366
8. Mergl, R., Mavrogiorgou, P., Hegerl, U., Juckel, G.: Kinematical analysis of emotionally induced facial expressions: a novel tool to investigate hypomimia in patients suffering from depression. Journal of Neurology, Neurosurgery & Psychiatry **76** (2005) 138–140
9. He, G.F., Kang, S.K., Song, W.C., Jung, S.T.: Real-time gesture recognition using 3d depth camera. In: Software Engineering and Service Science (ICSESS), 2011 IEEE 2nd International Conference on, IEEE (2011) 187–190
10. Zhang, Z.: Microsoft kinect sensor and its effect. MultiMedia, IEEE **19** (2012) 4–10
11. Ltkebohle, I.: BWorld Robot Control Software. `http://www.faceshift.com` (2015) [Online; accessed 11-6-2015].
12. Bouaziz, S., Wang, Y., Pauly, M.: Online modeling for realtime facial animation. ACM Transactions on Graphics (TOG) **32** (2013) 40
13. Shalev-Shwartz, S., Ben-David, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press (2014)