Automated Representation of Non-Emotional Expressivity to Facilitate Understanding of Facial Mobility: Preliminary Findings

Kathy Clawson[1], Louise S Delicato[2], Sheila Garfield[1]

[1]School of Computer Science, [2]School of Psychology

University of Sunderland

Sunderland, United Kingdom

kathy.clawson@sunderland.ac.uk; louise.delicato@sunderland.ac.uk; sheila.garfield@sunderland.ac.uk

Abstract—We present an automated method of identifying and representing non-emotional facial expressivity in video data. A benchmark dataset is created using the framework of an existing clinical test of upper and lower face movement, and initial findings regarding automated quantification of facial motion intensity are discussed. We describe a new set of features which combine tracked interest point statistics within a temporal window, and explore the effectiveness of those features as methods of quantifying changes in non-emotional facial expressivity of movement in the upper part of the face. We aim to develop this approach as a protocol which could inform clinical diagnosis and evaluation of treatment efficacy of a number of neurological conditions including Parkinson's Disease.

Keywords—smart healthcare; facial function; motion detection; KLT; motion tracking; Parkinson's Disease

I. INTRODUCTION

Understanding the factors that influence ability to recognize facial expressions is essential for social communication. The representation and recognition of facial expressions is therefore an important task and has attracted extensive attention from research communities across multiple disciplines, including psychology, neural and cognitive sciences [1], and computer science [2]. Within clinical contexts, evaluation of facial expressivity can assist with diagnosis and understanding of a number of neurological and neuropsychiatric conditions, including: Facial masking in people with Parkinson's; Facial Palsy; Apraxia; and Schizophrenia. Evaluation of facial mobility within such contexts can facilitate enhanced understanding of condition progression, and the impact of interventions.

There is evidence to suggest that an individual's ability to recognise emotions is impaired in a range of clinical disorders including depression and anxiety [3], post-traumatic stress disorder [4], dementia [5], and Parkinson's Disease [6]. People with Parkinson's Disease (PwP) can exhibit facial masking described as a loss in the movement of facial muscles leading to a reduction in ability to express emotions. Patients have a reduced ability to express spontaneous [7] and voluntary [8] emotions. As a result of this, clinical assessment methods typically focus on emotion-based criteria. For example, the

Facial Action Coding System (FACS) has been used to compare a subject's facial 'action unit' patterns against expected patterns (and their intensity) for specific spontaneous emotions including happiness, anger, fear and disgust [9]. However, Simons et al., [7] propose that facial masking in PwP is not limited to spontaneous and voluntary facial expressions of emotion, but that diminished non-emotional facial movement may be contributory. Using FACS as a basis for analysis, they showed a reduction in non-emotional expressivity in PwP.

Marneweck and Hammond [1] have identified the need for a systematic and comprehensive clinical investigation into non-emotional expressivity in the context of PwP. They acknowledge that, despite being a well-established and suitable approach for assessing facial musculature motion, FACS analysis is time consuming, highly specialised and requires significant training. As an alternative, they propose that voluntary facial musculature control may be measured in PwP using an adapted version of the upper and lower face apraxia test originally defined by Bizzozero et al., [10] for investigation of right- and left- hemisphere damaged patients. Although automated systems have been developed which replicate assessment of emotional indicators and FACS-based indicators of facial masking in PwP, we can find no evidence of computational frameworks which use the test proposed by Bizzozero et al., [10] to assist clinical investigation of non-emotional expressivity.

The impairment in ability of PwP to recognise emotions may be related to their reduction in ability to express emotions. Therefore, being able to quantify facial movement in PwP may lead to a greater understanding of the relationship between facial masking and impaired emotion recognition. In addition to supporting clinical diagnosis and treatment, understanding this relationship could support the development of an intervention to alleviate the extent of the impairment. Our aim, as the first stage in this process, is to develop an accurate automated method of quantifying facial movement.

This paper presents a video dataset which replicates the upper and lower face apraxia test [1], [10], and which is used for Automated Representation of Non-Emotional Expressivity (ARNEE). This could be developed to evaluate facial masking

in people with Parkinson's. We propose a new set of features which quantify facial motion using tracked interest point statistics, and illustrate the applicability of these features for quantifying the magnitude of linear translation within image sequences. An overview of related literature is presented in Section II. Our methodology, including ARNEE dataset description, and experimental overview is offered in Section III. Preliminary findings are summarised in Section IV, and conclusions are offered in Section V.

II. RELATED LITERATURE

While automated recognition of facial expressions constitutes a widely research topic, due to its applicability across a variety of domains, quantification of low-level facial function within image and video data is less well researched. There are few attempts at designing automated frameworks to quantitatively analyse facial expressivity in Parkinson's disease [11, 12].

Early investigations of facial expression in PwP, used a mathematical model of the face to quantify smiling behaviour based on 12 facial measures corresponding to the eyes, lips, and brows [11]. After digitisation of smiling expressions, and automatic determination of facial outlines, statistics were calculated for each facial measure. Results supported the use of the mathematical model as a basis for studying facial activity and expression in PwP. In particular, when smiling, the "mouth opening" facial measure was the only facial measure to produce a significant difference between the PwP and a control group.

More recently, Joshi and Betke [12] extracted six geometric features from temporal signals associated with facial landmark coordinates, and evaluated their discriminant capability for predicting facial expressivity in PwP. Specifically, eye, eyebrow, and mouth distances are characterised over time and used as inputs within a random forest regression framework. Evaluation using 9-fold cross validation suggested that, contradictory to the work of Katsikitis and Pilowsky [11], eyes and eyebrows were more influential and better predictors of facial expressivity than the mouth.

Vinkurov *et al.,* [13] evaluate facial expression through application of machine learning and a 3D depth camera. Real time face tracking is used to generate features from the raw signal data, which are then used as inputs for linear regression training. The Pearson Correlation Coefficient (PCC) is applied to assess predicted scores against expert ground truth. Despite results showing that all correlation scores are statistically significant, this approach requires participant training prior to use. Furthermore, although 3D and sensor-based methods for facial analysis allow more accurate determination of motion across x, y and z planes, their practical application is constrained due to system set up requirements.

Leonard *et al.,* [14] proposed facial motion may be quantified by calculating the *entropy* of changes in greyscale image intensity over time. After subtracting corresponding pixel values between consecutive frames, and generating histograms of difference $h(s)$, entropy is derived using:

$$entropy = -\Sigma_{i:n} \, h(s_i) * \log(h(s_i)) \qquad (1)$$

where $j_s = 1, 2, \ldots n$ and n represents the number of difference images extracted from a sequence. It was concluded that entropy values characterised the volume of facial change between frames, and that patterns of entropy change over time could effectively represent changes in facial expressions.

The foundations of this approach have been used within Parkinson's research [16], and within research on asymmetry of facial mobility [15]. On a practical level, the implementation details of 'entropy' vary between researchers. Bowers *et al.,* [16] evaluate entropy as a method for quantifying micro-expressivity and bradykinesia (delayed onset of expression) during voluntary facial expressions in PwP. Using entropy and the time taken for an expression to reach peak entropy as features, it was demonstrated that less motion occurred over the face of PwP when compared against control participants. Richardson *et al.,* [15] use entropy to demonstrate that the majority of facial expressions are characterised by greater magnitudes of motion in the right hand side of the upper face (regardless of emotionality). Here entropy is calculated as:

$$entropy = \sum_{t=2}^{T} ID_t \qquad (2)$$

Where ID_t constitutes the mean intensity difference between image frames at time t and t-1.

Despite promising results, these approaches assume facial texture relates directly to facial musculature activation and do not consider motion magnitudes or directions. It is likely that large individual differences exist in skin texture, not least those defined by the ageing process. Furthermore, when entropy is used to assess an entire facial region globally, the spatial relativity of changes in expression is lost. Calculation of entropy is based on the assumption that changes in intensity are representative of motion properties within the image. However, if the underlying background distribution of a region of interest is stronger that the motion properties inherent within it then this assumption will not hold.

III. ARNEE: METHODS

A. Video Processing

The framework for video processing is demonstrated in Figure 1. After transformation into grayscale format and face detection using the Viola-Jones algorithm [17], face regions of interest are sub-divided into upper (*UROI)* and lower (*LROI)* segments, defined arbitrarily by the location of the nose. *UROI* is defined as the top two thirds of the original face region of interest and *LROI* is the remaining region. Feature extraction is performed independently for both *UROI* and *LROI*, and consists of: neutral face subtraction; interest point detection and tracking; and trajectory statistic calculation.

Facial expressions exhibit variability between individuals, due to differences in facial shape, texture, and appearance, the degree of natural facial elasticity, and frequency of expression. It can be difficult to quantify the strength of expressivity, both
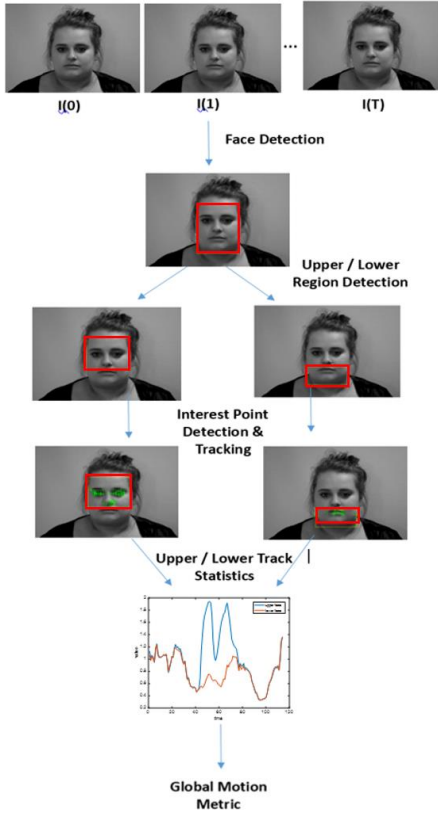
and pixels are ranked according to:

$$R = \det(M) \tag{5}$$

Finally, a threshold is applied so that the top *n* points of *R* are retained within upper and lower regions (arbitrarily set to *n* = 50 for *UROI* analysis, and *n* = 25 for *LROI* analysis). Identification of interest points in this manner assumes that good features to track are those whose motion can be estimated reliably, and eliminates the requirement for facial landmark detection. It is less computationally expensive than dense tracking, for example when motion properties are calculated for every pixel in the frame.

Interest points are tracked frame by frame using the KLT optical flow algorithm [19]. KLT compares the spatial intensity gradient between windowed image regions and unlike many point trackers does not require segmented image input data. The KLT algorithm utilises the fact that, in many instances, consecutive time frames within image datasets are already approximately registered. Optical flow calculation via KLT is equivalent to quantifying displacement across two images, *A(x)* and *B(x) = A(x + v)*, where *v* is an n-dimensional vector. To achieve this, the behaviour of *A(x)* is approximated across the neighbourhood of *x*. It is assumed that the following linear relationship exists:

$$A(x+v) \approx A(x) + vA'(x) \tag{6}$$

The success of Equation (6) assumes that the displacement vector, *v*, is small enough for an adequate approximation to be obtained (*h*). Based on this, we may then solve displacement by minimizing the energy function [20]:

$$E_x = [A(x+h) - B(x)]^2 \tag{7}$$

The result of point tracking in each face region of interest is a set of *N*T* vectors where N_n (*n* = 1,2, … number of trajectories), is a single trajectory with (x, y) coordinates tracked over all timeframes *T*. A distance measure *DM* is calculated for each time frame *t* as:

$$DM_t = \frac{\sum_{n=2}^{N} ED_n}{N} \tag{8}$$

where *t = 2, 3, … T* corresponds to a single frame within a video segment, ED_n is the Euclidean Distance between a trajectory's coordinates at times *t* and *t-1*, and *N* is the total number of (interest point) trajectories in the frame. To compensate for variations in scale, each Euclidean Distance measure ED_n is normalised to assume a standardised face region of interest.

Fig. 1.   Proposed Framework for Video Processing.

visually and using automated measures, as changes are often subtle in nature. To eliminate the effect of inter-person variances, we perform neutral face subtraction prior to image frame analysis [18]. Specifically, for a given sequence *Fi*, comprised of *T* frames, we derive normalised sequences S_i where:

$$S(i) = F(x_i, y_i) - F(x_1, y_1) \tag{3}$$

for *i* = 1, 2,… *T*. Interest points and motion trajectories are subsequently calculated using *S,* the set of subtracted frames corresponding to the original sequence minus the neutral face frame.

To initialise tracking, we adopt the corner point detection methodology defined in [19]. Specifically, for image frame *I* = *S(1)*, the second moment matrix of each pixel is calculated within a windowed region, *w*:

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \tag{4}$$

Finally, to allow comparison across signals with varying durations, *DM*, is scaled to express the total cumulative value of *DM* across a 1/10th of a second period. It has been reported that true signal amplitudes evolve smoothly as a function of time, and that rapid variations in amplitude may be indicative of signal noise [21]. In order to reduce the impact of potential noise variations, and to allow comparison across signals with varying durations, an unweighted sliding window is applied to the distance metric, *DM,* such that the value of *DM* for each frame becomes a function of surrounding frames:

$$DM(t) = \sum_{i=t-w}^{i=t+w} DM(i) \qquad (9)$$

Where *t = 1,2...* number of frames, and *w* = 4 is the half-width of the sliding window applied.

B. Experimental Overview

For the purposes of this investigation, data constitutes video recordings of 25 participants recruited from the University of Sunderland staff and student population, with age categories of 18 – 34 (11 participants), 35 to 50 (9 participants) and over 50 (5 participants). Of the 25 participants included, 9 were male and 16 were female. Subjects with any self-reported history of learning disability, depression, or neurological condition were excluded, as were individuals with facial jewellery or whose hair occluded the forehead region. The research was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). The authors obtained informed consent from participants, including consent to share video data with the wider research community. Data collection is ongoing.

Facial movements were captured at 60 fps using a Nikon D5500 digital SLR camera with 35mm lens. The camera was positioned at a fixed distance (2 metres) from participants, and lighting was standardized through the use of industry-standard backlighting. Each video frame was captured at 448 * 797 resolution, in RGB colour. Subjects were informed they were participating in a study of facial movement and that they would be video recorded by a researcher throughout the session. One researcher operated the digital camera, while a second asked participants to pose each facial gesture. Participants were asked to begin and terminate each sequence in a neutral face expression. When time permitted, each participant was recorded completing the full set of gestures twice.

The 38 gestures captured constituted those listed within the upper- and lower- apraxia test [10]. For the purposes of this study, only gestures within the upper apraxia test are investigated, specifically: (a) close eyes, (b) look down, (c) look right, (d) wrinkle nose, (e) wrinkle forehead, (f) look left, (g) look up, and (h) blink right. Example lower apraxia test gestures include: open your mouth, move the jaw left to right three times, puff out your cheek, and stick out both lips.

Initial analyses evaluated which facial expressions resulted in the greatest overall amount of movement by examining *DM* for each face region of interest and across the entire sequence.

Specifically, we calculate: *UDM,* which corresponds to the average distance metric *DM* calculated across the entire sequence over *UROI*; *LDM,* which corresponds to the average distance metric *DM* calculated across the entire image sequence over *LROI;* and *GDM = UDM + LDM,* which we define as a global measure of facial motion. To compare this approach with existing methods, entropy (as defined in [16]) is also calculated. To evaluate the statistical significance of calculated metrics, one way analysis of variance (ANOVA) is implemented using Matlab.

IV. RESULTS & DISCUSSION

The objectives of this study were:

- to validate the capability of tracked motion descriptors for measuring translations in image sequences;

- to investigate how motion displacements can describe facial gesture evolution over time;

- to compare motion descriptors for a given facial gesture between individuals; and

- to identify which facial gestures result in the greatest overall movement.

A. Results using Synthetic Sequences

To validate our approach, we generate sequences of synthetic images (300*300 pixels per frame) with known motion and noise properties, and calculate (per frame) *DM*. Specifically, we generate sequences of greyscale images, as illustrated in Figure 2 (a), with a single circular foreground object (radius = 25 pixels), and known foreground / background distribution (background mean = 50, foreground mean = 200, with signal to noise ratio, SNR = 10, 8, … 4). At each frame, the foreground object is translated by a random value within a known uniform distribution (+/- *n* pixels).

We compute *DM* values for each frame (number of frames = 20) and compare frame level *DM* against ground truth motion. Each simulation is repeated for 100 iterations, and mean *DM* recorded. It can be seen from Figure 2 (b) that *DM* remains an accurate measure of linear displacement, across all signal to noise ratios. *DM* decreases slightly in the existence of noise, but the overall trend remains stable- specifically, as displacement increases the value of *DM* increases linearly. Figure 2 (c) and (d) illustrate frame-level displacement for varying motion steps and SNRs. Again, it can be seen that our approach is able to accurately capture global motion patterns across a sequence.

B. Results using Control Participants

When applied to the ARNEE dataset, per-frame displacements calculated using Equation (9) allow visualisation of gestures over time. For example, Figure 3 illustrates upper face track displacement for a single individual, for both *wrinkle nose* and *close eyes* gestures (signals aligned).

(a)   (b)
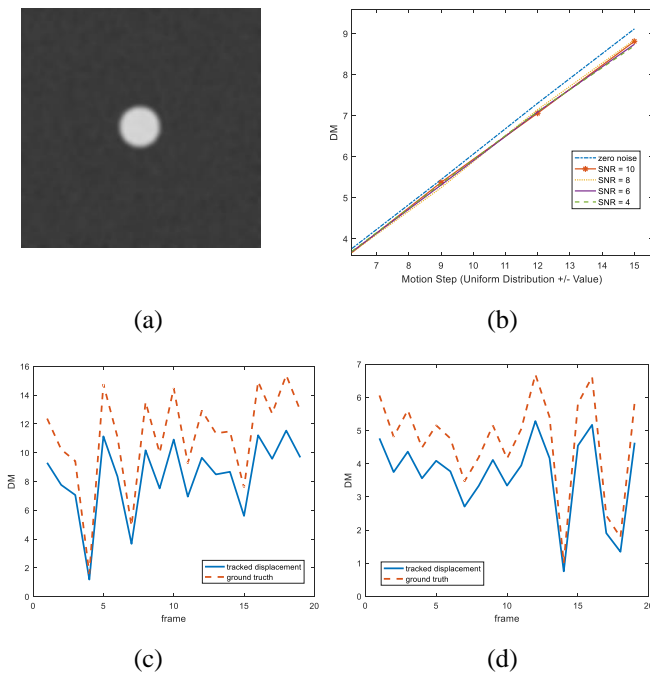

(c)   (d)

Fig. 2.   (a) Example synthetic frame (SNR = 8, Motion step +/- 9 pixels), (b) *DM* as a function of linear displacement, (c) *DM* per frame (motion step +/-15 pixels, SNR = 8) and (d) *DM* per frame (motion step +/-8 pixels, SNR = 8)

Fig. 3.   Comparison of upper face displacement for 'wrinkle nose' and 'close eyes' gestures (single participant)

It can be seen in Figure 3 that both signals have two peaks. The first peak leads to the expression apex, while the latter constitutes the face returning back to neutral. It is also apparent that, for the given individual, *wrinkle nose* induces greater overall displacement than *close eyes.*

Figure 4 summarises mean upper and lower face displacement, and entropy [15] , for a single participant, during *blink right*. It can be seen that both upper and lower displacement are similar for frames 1 to 40 and 80 to 120. When looking at frames 40 to 80 it is observed that the relationship between upper and lower motion facilitates identification of the gesture boundary. Specifically, the point at which the difference between *LDM* and *UDM* is maximum facilitates temporal isolation of the gesture. The peak in upper face displacement, evident after frame = 40 (Figure 4 (a)), corresponds to the beginning of a 'blink' gesture. Similarly, the trough at frame = 54 (Figure 4 (b)) corresponds to the participant having fully closed her eye and paused briefly before re-opening it (Figure 4 (c)). No similar observation exists within the entropy signal.

When comparing gesture manifestation across individuals, it is apparent that despite neutral face subtraction (Equation 3), there exists variability in expression durations (Figure 5) and displacement magnitudes (Figure 6). The standard deviation of gesture duration is largest for *look up* and *blink right*. The average *UDM* for *blink right* across all participants is 13.4, with a range of 16.5 and standard deviation of 3.5. Similarly, the average *LDM* for *blink right* across all participants is 11.1, with a range of 16.5 and standard deviation of 3.9. One way ANOVA indicates that the signals for upper and lower face for *blink right* are statistically different ($p = 0.04$).
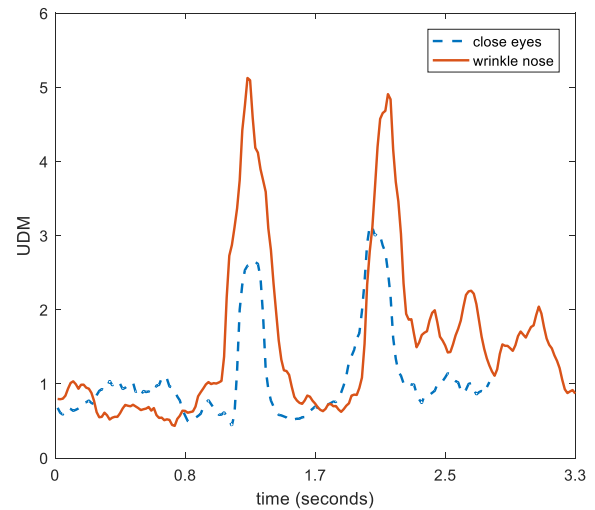
Despite variability in signal durations and magnitudes, many signals exhibit similar patterns across individuals, irrespective of magnitude differences. Figure 7 illustrates *blink right* upper displacement for two participants, with the signal's peaks aligned. It can be seen that, although participant b (red) exhibits greater magnitudes of motion during their expression, the overall pattern of motion remains similar across the two participants. Such observations highlight the degree to which levels of expressivity vary naturally across populations, including healthy control participants.

Table I illustrates mean entropy and mean *DM* scores (all participants) for each gesture in the upper face apraxia test. It can be seen from Table I that *wrinkle nose* (*GDM* = 25.72), *wrinkle forehead* (*GDM* = 23.71), *blink right* (*GDM* = 24.54), and *blink left* (*GDM* = 24.34) all display higher global scores than *closing* and *looking* gestures. It is also evident from Table 1 that the right hemisphere expression *look right* is characterised by marginally greater *DM* scores than it's left hemisphere counterpart. Left versus right hemiface asymmetry is a naturally occurring manifestation which has been discussed elsewhere [15]. The standard deviation across all gestures is 0.44 and 1.6 for entropy and *GDM,* respectively.

When feature scores are aggregated into high level class of gesture, specifically, *blink, wrinkle, look* and *close* (Table II), it is apparent that *blink* and *wrinkle* gestures have similar *DM* scores, as do *close* and *look* gestures. There is a statistically significant difference between gesture class means for *GDM,* specifically $F(3, 221) = 3.72$, $p = 0.012$. When evaluating the statistical significance of entropy across gesture classes, the same pattern does not hold. Specifically, we accept the null hypothesis where $F(3, 221) = 1.17$, $p = 0.342$. The lack of statistical significance when using entropy suggests that it is a less reliable measure of movement when utilised in high resolution and high frame rate environments (such as the ARNEE data capture scenario).
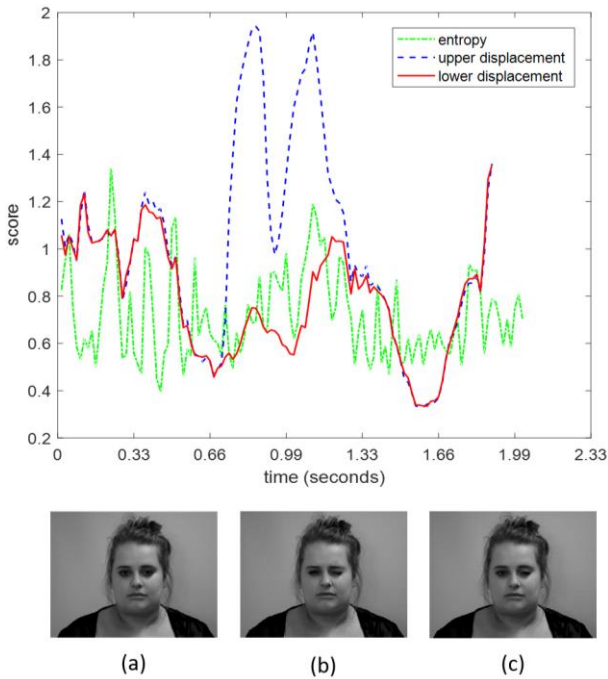
Fig. 4. Entropy versus mean upper / lower track displacement during blink gesture, and corresponding video frame for frame = 40 (a), frame = 54 (b), and frame = 61 (c)

Table I also shows that, while upper motion scores are similar across all classes of gesture in the upper face apraxia test, some gestures induce more lower face motion than others. *Close* and *look* are typified by lower mean *LDM* across all participants than *blink* and *wrinkle*. Results of a one-way ANOVA for *LDM* scores shows a significant main effect for lower facial motion by gesture type $F(3, 221) = 9.33$, $p = 0.00001$. Post hock multiple comparisons show that *closing eyes* produced less lower face movement than *wrinkle* ($p = 0.0046$) and *blink* ($p = 0.033$) gestures, respectively. Similarly, *look* gestures were characterised by less lower face motion than *wrinkle* ($p = 0.0001$) and *blink* ($p = 0.0022$) gestures. There was no significant difference in means for *close* versus *look* ($p = 0.99$) or *wrinkle* versus *blink* ($p = 0.87$).

V.   CONCLUSIONS

In this study, we calculate the displacement of automatically detected two-dimensional interest points as a measure of facial mobility. This method overcomes a limitation of existing approaches such as entropy, which assumes that global changes in pixel intensity directly correspond to facial mobility through facial muscle activation. The presence of strong background noise or strong facial texture degrades the performance of entropy. Interest point analysis considers only regions in the face whose movement properties may be reliably tracked. It is a more selective and robust approach, and less computationally intensive than dense tracking.
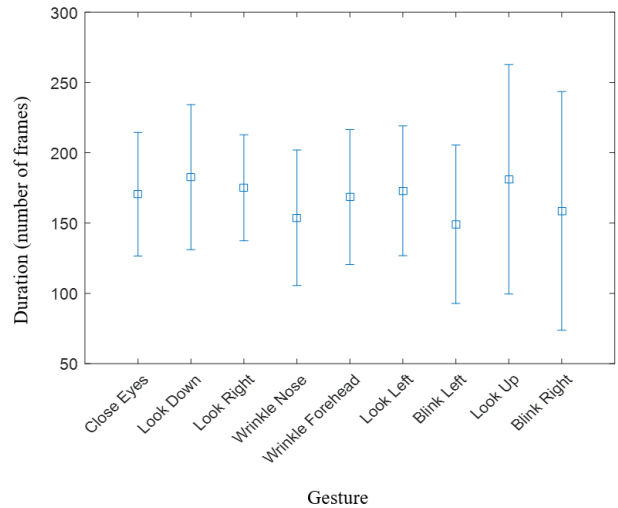

Fig. 5.   Average gesture duration (number of frames), and standard deviation.
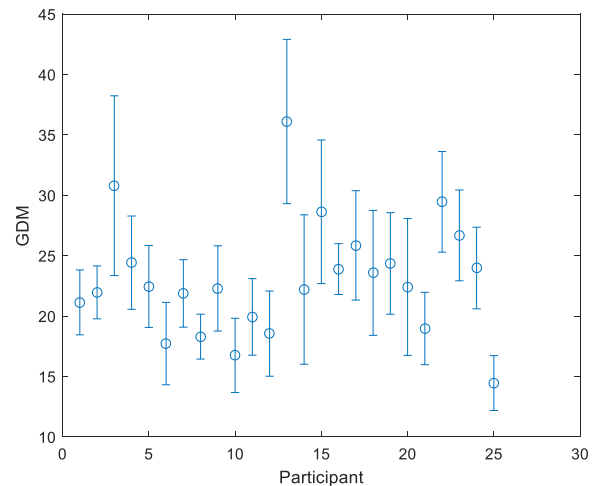

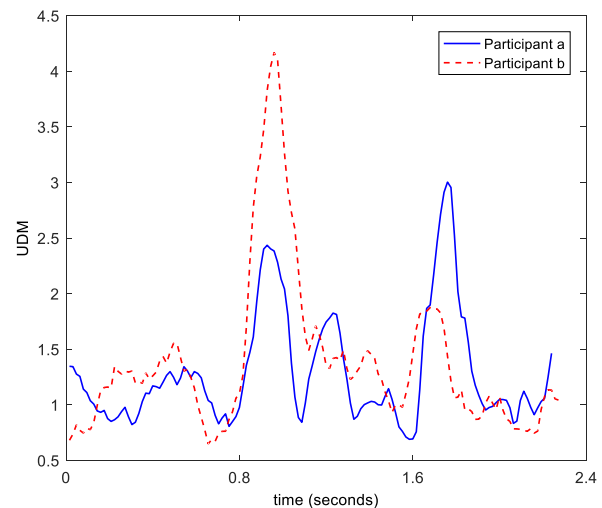Fig. 6.   Average GDM and standard deviation, by participant

Fig. 7.   'Blink Right' upper face displacement, aligned signals (truncated)

TABLE I.   COMPARISON OF FEATURE SCORES BY GESTURE

Gesture	Feature Score			
	Entropy	*UDM*	*LDM*	*GDM*
Close Eyes	13.33	13.07	8.79	21.86
Look Down	13.08	13.30	9.38	22.68
Look Right	13.76	12.42	8.91	21.33
Wrinke Nose	14.09	12.44	13.28	25.72
Wrinkle Forehead	13.97	13.76	9.95	23.71
Look Left	13.54	12.11	8.70	20.81
Blink Left	13.72	13.33	11.01	24.34
Look Up	12.99	13.85	8.87	22.72
Blink Right	14.25	13.37	11.17	24.54

TABLE II.   COMPARISON OF FEATURE SCORES BY CLASS OF GESTURE

Gesture	Feature Score			
	Entropy	*UDM*	*LDM*	*GDM*
'Close'	13.33	13.07	8.79	21.86
'Look'	13.34	12.92	8.96	21.88
'Blink'	13.98	13.35	11.09	24.44
'Wrinkle'	14.03	13.10	11.62	24.71

To validate tracked features, we generate synthetic image sequences with known motion properties, and create a benchmark dataset which replicates the upper and lower face apraxia test defined by Bizzozero et al. [10]. Validation experiments using synthetic image sequences illustrates the capability of tracked motion descriptors for measuring translations in image sequences. *DM* proves a reliable indicator of motion displacement across a variety of signal to noise ratios and across differing magnitudes of displacement.

Comparison of *UDM*, *LDM* and *GDM* between expressions demonstrates that motion displacements can describe facial gesture strength and evolution over time, and can allow comparison of gestures between individuals. *Look up* and *blink right* have larger average signal duration than other gestures, whilst *wrinkle nose*, *wrinkle forehead*, *blink right* and *blink left* induce greater magnitudes of displacement than *close* and *look* gestures. Comparison of displacement across four categories of gesture type (*close, look, blink, wrinkle*) shows there is a statistically significant difference between gesture class means for *GDM*. No statistical significance was found for entropy scores. Further analysis of *LDM* suggests that classes of gestures which demonstrate higher mean displacement magnitudes induce more motion in the lower face than those with lower mean motion magnitudes. This highlights the complexity involved when isolating analysis of 'upper face' motion. All elements of the upper face test involved some degree of lower face muscle activation.

Future investigation should evaluate the lower face apraxia test gestures and consider validating it against artificially generated face images with known movement properties. We anticipate that the final ARNEE dataset will comprise 100 participants, thus facilitating an extended and more comprehensive study. It is also desirable to perform more extensive comparison against other methods for facial motion characterisation. We anticipate that future investigation will seek to compare facial motion strength for control participants (both upper and lower face apraxia tests) with motion properties of clinical populations, for example people with Parkinson's.

REFERENCES

[1] M. Marneweck and G. Hammond, "Voluntary control of facial musculature in Parkinson's Disease," *J. Neurol. Sci.*, no. 347, pp. 332–336, 2014.

[2] C. Wagenbreth, L. Wattenberg, H.-J. Heinze, and T. Zaehle, "Implicit and explicit processing of emotional facial expressions in Parkinson's disease.," *Behav. Brain Res.*, vol. 303, pp. 182–190, 2016.

[3] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. M. Mavadati, Z. Hammal, and D. P. Rosenwald, "Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses," *Image Vis. Comput.*, vol. 32, no. 10, pp. 641–647, 2014.

[4] E. H. F. Poljac, E., Montange, B. & de Haan, "Reduced recognition of fear and sadness in post-traumatic stress disorder.," *Cortex*, vol. 47, no. 8, pp. 974–980, 2011.

[5] A. W. Keane, J., Calder, A.J., Hodges, J.R., Young, "Face and emotion processing in frontal variant frontotemporal dementia.," *Neuropsychologia*, vol. 40, no. 6, pp. 655–665, 2002.

[6] S. Zhao, F. Rudzicz, L. G. Carvalho, C. Marquez-Chin, and S. Livingstone, "Automatic detection of expressed emotion in Parkinson's Disease," *2014 IEEE Int. Conf. Acoust. Speech Signal Process.*, no. May, pp. 4813–4817, 2014.

[7] G. Simons, H. Ellgring, and P. Smith, "Disturbance of spontaneous and posed facial expressions in Parkinson's disease," *J. Cogn. Emot.*, vol. 17, no. 5, pp. 759–778, 2003.

[8] G. Simons, M. C. S. Pasqualini, V. Reddy, and J. Wood, "Emotional and nonemotional facial expressions in people with Parkinson's disease.," *J. Int. Neuropsychol. Soc.*, vol. 10, no. 4, pp. 521–535, 2004.

[9] P. Wu, I. Gonzalez, G. Patsis, D. Jiang, H. Sahli, E. Kerckhofs, M. Vandekerckhove, P. Wu, I. Gonzalez, G. Patsis, D. Jiang, H. Sahli, E. Kerckhofs, and M. Vandekerckhove, "Objectifying Facial Expressivity Assessment of Parkinson Patients: Preliminary Study," *Comput. Math. Methods Med.*, vol. 2014, no. Article, 2014.

[10] I. Bizzozero, D. Costato, S. Della Sala, C. Papagno, H. Spinnler, and A. Venneri, "Upper and lower face apraxia: role of the right hemisphere," *Brain*, vol. 123, pp. 2213–2230, 2000.

[11] I. Katsikitis, M., and Pilowsky, "A study of facial expression in Parkinson's disease using a novel microcomputer-based method," *J. Neurol. Neurosurg. Psychiatry*, vol. 51, no. 3, pp. 362–366, 1988.

[12] M. Joshi, A., and Betke, "Predicting Active Facial Expressivity in People with Parkinson's Disease," in *Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '16)*, 2016.

[13] N. Vinokurov, D. Arkadir, E. Linetsky, H. Bergman, and D. Weinshall, "Quantifying hypomimia in Parkinson patients using a depth camera," *Commun. Comput. Inf. Sci.*, vol. 604, pp. 63–71, 2016.

[14] C. M. Leonard, K. K. S. Voeller, and J. M. Kuldau, "WHEN'S A SMILE A SMILE? Or How To Detect a Message by Digitizing the Signal," *Psychol. Sci.*, vol. 2, no. 3, pp. 166–172, 1991.

[15] C. K. Richardson, D. Bowers, R. M. Bauer, K. M. Heilman, and C. M. Leonard, "Digitizing the moving face during dynamic displays of emotion," *Neuropsychologia*, vol. 38, no. 7, pp. 1028–1039, 2000.

[16] D. Bowers, K. Miller, W. Bosch, D. Gokcay, O. Pedraza, U. Springer,

and M. Okun, "Faces of emotion in Parkinsons disease: micro-expressivity and bradykinesia during voluntary facial expressions.," *J. Int. Neuropsychol. Soc.*, vol. 12, pp. 765–773, 2006.

[17]   O. M. Way and M. J. Jones, "Robust Real-Time Face Detection," 2004.

[18]   and M. V. L. Bazzo, Juliano J., "Recognizing facial actions using gabor wavelets with neutral face average difference.," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[19]   & T. Jianbo, S, "Good features to track," in *Proceedings of IEEE*

Conference on Computer Vision & Pattern Recognition, 1994, p. 593.

[20]   B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 2. pp. 674–679, 1981.

[21]   B. Boashash, *Time-frequency signal analysis and processing: a comprehensive reference.* Academic Press, 2015.