AUTOMATIC MOTION FEATURE EXTRACTION WITH APPLICATION TO QUANTITATIVE ASSESSMENT OF FACIAL PARALYSIS

Shu He University of Strathclyde

John J Soraghan University of Strathclyde

Brian F O'Reilly Institute of Neurological Sciences

ABSTRACT

This paper presents a robust, objective, automated and quantitative assessment system for Facial Paralysis using artificial intelligence analysis of biomedial video data. Facial feature localization and prescribed facial movements detection are discussed. Optical flow is used to obtain the motion features in the relevant facial regions. Radial Basis Function (RBF) Neural Network is applied to provide quantitative evaluation of Facial Paralysis based on the House-Brackmann Scale. The results from 197 videos of 87 subjects are encouraging with a Mean Squared Error (MSE) of 0.013 (training) and 0.0169 (testing).

Index Terms— Facial Paralysis Measurement, Optical Flow, RBF Neural Network, House-Brackmann Scale

1. INTRODUCTION

Facial Paralysis is a condition where damage to the facial nerve causes weakness of the muscles on one side of the face. Traditional assessment of facial paralysis is by the House-Brackmann (H-B) grading system [1]. Grading is achieved by asking the patient to perform certain movements and then using clinical observation and subjective judgment to assign a grade of palsy ranging from grade I (normal) to grade VI (no movement). The advantages of the H-B grading scale are its ease of use by clinicians and that it offers a single figure description of facial function. The drawbacks are its subjective judgment with significant inter & intra observer variation and its insensitivity to regional differences of facial function [2]. Several objective facial grading systems have been recently reported. These predominantly involve the use of markers on the face requiring trained technicians. The success of the uptake of any such system will hinge on the ease of use of the technology [3].

In this paper, we present an automated, objective and robust facial grading system. The patient is videotaped using a front face view with a clean background. The video sequence begins with the patient at rest, followed by the five facial movements, which are raise eyebrows, close eyes gently, close eyes tightly, screw-up nose and smile, going back to rest in between each movement. A robust face feature localization method is employed in the reference frame (at rest). The image of the subject is stabilized by using block matching techniques. Image subtraction is then employed to identify the period of each facial movements. Our proposed method uses optical flow to compare the symmetry of the facial movements between each side of the face. The extracted features are fed into RBF neural networks to quantitatively estimate the degree of damage for each facial region. Finally, the regional results are fed into to a RBF neural network to provide an overall quantitative evaluation of facial paralysis based on H-B Scale.

The paper is organized as follows. In Section 2 the face feature localization process is presented. In Section 3 the algorithms for extraction of motion features are developed. In Section 4 the results obtained from the artificial neural networks are presented and Section 5 concludes the paper.

2. LOCALIZATION OF FACIAL REGIONS

Identifying the different regions of the face on each side is crucial for an automatic grading system. In order to do this the pupils are first localized and then the inter-pupil distance is used to scale the size of each facial region. The mouth corners are then localized. The forehead and nasal regions are initially assigned by the positions of the pupils and the mouth corners and are calibrated. Finally, a face region map is assigned. This method was tested using 266 front view faces and reached 95.11% accuracy. The following sections describe the algorithms for pupil and mouth corners detection.

2.1 Detection of the ROI of eyes and mouth

The features of a face (eyebrows, eyes, nostril, mouth) are generally darker than the normal skin color. Hair may be darker than facial features. A Gaussian filter is used to centre weight the head area. The Gaussian weights can be expressed as

$$w(x,y) = e^{-\frac{(x-x_o)^2+(y-y_o)^2}{2*((x_{right}-x_{left})/3)^2}} \quad (1)$$

where $w(x,y)$ denotes the weight at the pixel with coordinates (x,y). x_{right} and x_{left} are the horizontal position of right and left face boundaries, which can be

identified by vertical projection of a Sobel filtered image followed by thresholding and smoothing. Similarly, horizontal projection of the binary image is used to find the top boundary. (x_o, y_o) is the centre of the face, which can be roughly estimated by the boundaries of the face. An example of an inverted, thresholded, Gaussian weighted image is shown in Fig.1.(a). The ROI of the eyes and mouth can be determined by its horizontal projection as shown in Fig.1.(b) and the boundaries of the face.

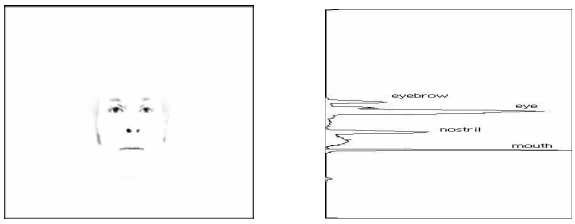

Fig.1 Detection of the vertical position of facial features
(a) Gaussian filtering of face, (b) Horizontal projection

2.2. Pupil Search

The iris-pupil region localization is based on an eye template, which is a filled circle surrounded by a box. The filled circle represents the iris and pupil as one part. The eye width to eye height relation can be expressed as 3:1 and the eye height is interpreted as iris diameter [4]. This eye template is scaled automatically depending on the size of the face area. The iris is localized by searching the minimum difference between the template and the ROI of the eyes. The pupil is darker than the iris in most cases and therefore can be determined by searching the small circle with lowest intensity value within the iris area.

2.3 Mouth Corner Search

The mouth corners are detected by applying the SUSAN (Smallest Univalue Segment Assimilating Nucleus) algorithm [5] for corner extraction to the ROI of the mouth. The decision whether or not a point (nucleus) is a corner is based on examining a circular neighbourhood centred round the nucleus. The points from the neighbourhood whose brightness is approximately the same as the brightness of the nucleus form the area referred to as USAN(Univalue Segment Assimilating Nucleus). The point (nucleus) with the smallest USAN areas indicates the corner. Usually, more than one point is extracted as a corner. Therefore knowledge-based rules, such as relative position with the pupils, the distances from the middle face to left and right mouth corner, etc., are applied to reliably extract the mouth corners.

3. FACIAL MOVEMENT MEASURE

The face images are stabilized firstly to remove unwanted head motion by an error minimization algorithm which finds the minimum displacement between each frame and the facial window encompassing the eyebrows, eyes, nose and mouth. The five key movements are then detected. The motion is extracted from the appropriate time in the videos for the appropriate facial region movement.

3.1 Movement Recognition

The video sequence begins with the patient at rest, followed by five movements, returning to rest in between each movement. The five movements are detected by totalling several smoothed and varying thresholded pixel displacement values until five peaks and four troughs of sufficient separation can be extracted. The magnitude of each key movement can be extracted by examining the pixel displacement information in the specific facial region at the corresponding sequence of frames. The graphs shown in Fig.3 demonstrate the full displacement results for an almost recovered patient with mild weakness at right side of eye and mouth. Five plots in Fig.3 show the magnitude of the five movements in the relevant facial region. The broken line indicates the detected movement on the patients' right side of the face and the solid line indicates the movement detected on the left. The normal side is standardized to 1.

3.2 Illumination

The results of videos taken in nonhomogeneous lighting conditions may be skewed. In our work, once the facial map has been defined in the reference frame, the ratios of the intensity mean values between left side and right side in the relevant regions are calculated as an illumination compensation factor to adjust the subsequent frames. Fig.2 illustrates the results without illumination compensation, indicating the detected motion on the right side is significantly less than the left side. In Fig.3 the results with compensation shows similar movement magnitude for both sides except for the eye and mouth, which is in keeping with the clinical situation.

The movement magnitude in the relevant region can be computed as follows:

$$mag(n) = \sum_{(x,y)\subset R}\sum \left[I_n(x,y) - I_{ref}(x,y)\right] * w(x,y) * lum \qquad (2)$$

where $I_n(x,y)$ is the intensity of pixel (x,y) at the nth frame, $I_{ref}(x,y)$ is the intensity value at the reference frame, w is the Gaussian weight, similar to the equation (1), but set (x_o, y_o) to be the centre of the region and lum denotes the luminance compensation factor, which is set to

$$\sum_{(x,y)\subset left}\sum I_{ref}(x,y) \Big/ \sum_{(x,y)\subset right}\sum I_{ref}(x,y) \qquad (3)$$

for the right side and $lum = 1$ for the left side.

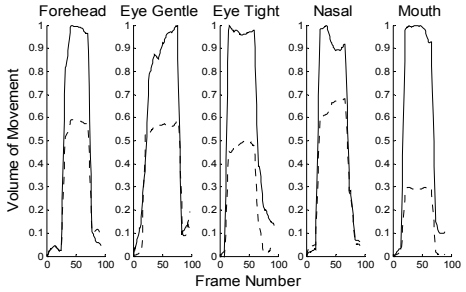

Fig.2. Motion magnitude without illumination compensation

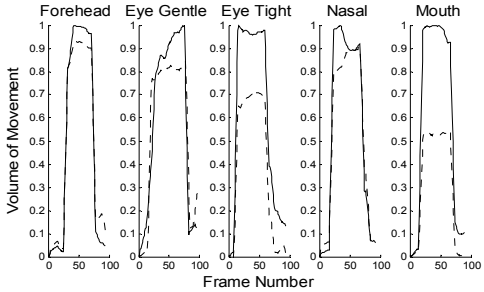

Fig.3. Motion magnitude with illumination compensation

3.3 Optical flow

The Lucas and Kanade algorithm[6], including the pyramid approach is employed to compute the optical flow on five pairs of images, i.e. the reference frame and the frame with maximum motion in each movement. Using the pyramid method with reduced resolution allows us to track the large motion while maintaining its sensitivity to subtle facial motion and allows the flow computation to converge quickly. Fig.4 shows the results of optical flow estimation for a left palsy subject.

In each facial feature region, the flow is thresholded to reduce the effect of small computed flow which may be either produced from textureless parts. The overall flow vector $\vec{v} = (u, v)$ of each region can be calculated by $u = \sum_i threshold(u_i) * w_i$ and $v = \sum_i threshold(v_i) * w_i$, where (u_i, v_i) denotes the flow vector and w_i is the Gaussian weight. The results show that even for a normal person looking at both sides there is no consistency in the direction of the horizontal displacement. Also the motion on the horizontal direction does not contribute much information when measuring the symmetry of the motion direction during the raising eyebrows or closing eyes movements. Therefore the horizontal displacements are given less weighting than the vertical displacements which are more relevant when measuring the symmetry of motion. The symmetry of the facial motion is quantified as:

$$Sym_y = 1 - \frac{v_{left} - v_{right}}{|v_{left}| + |v_{right}|} \tag{4}$$

$$Sym_r = 1 - \frac{\|\vec{v}_{left}\| - \|\vec{v}_{right}\|}{\|\vec{v}_{left}\| + \|\vec{v}_{right}\|} \tag{5}$$

where the v_{left} and v_{right} are the overall vertical displacements on the left side and the right side, $\vec{v}_{left}$ and $\vec{v}_{right}$ are the overall flow vector on the left side and the right side. Sym_y represents the symmetry relative to the vertical component of the total amount of displacements from the resting face to the peak of movement. Sym_r represents the symmetry relative to the strength of the total amount of displacements from the resting face to the peak of movement. Sym_y and Sym_r will be within the range 0-1. The motions on each side of the face are symmetrical when both approximate 1. When both approximate 0 indicates one is normal and the other side has no movement at all. While Sym_y = 0 and Sym_r = 1 indicates the motion on each side are same amplitude but opposite direction. The overall flow $\vec{v} = (u, v)$ for smile in the mouth region is refined by $u_{left} = \sum_{u_i<0} threshold(u_i) * w_i$, $v_{left} = \sum_{u_i<0} threshold(v_i) * w_i$ $u_{right} = \sum_{u_i>0} threshold(u_i) * w_i$, $v_{right} = \sum_{u_i>0} threshold(v_i) * w_i$ This allows us to remove the motion on one side that is drawn by the other side during the 'smiling' movement in the presence of severe facial palsy patients.

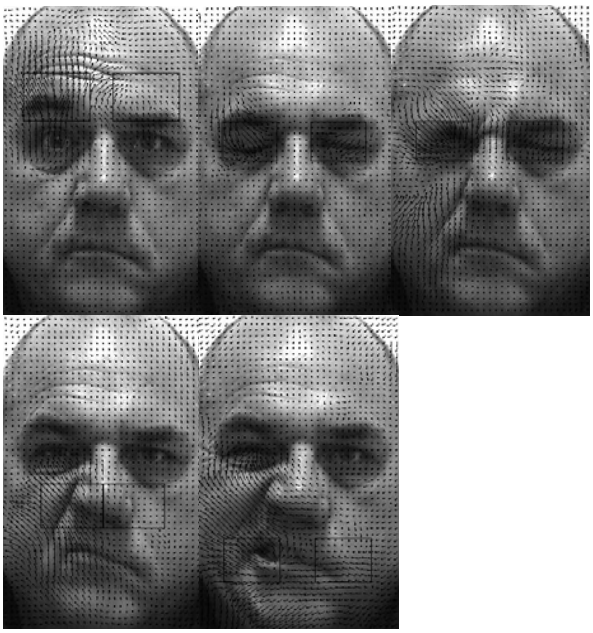

Fig.4. Results of optical flow estimation on five frames with peak motion for a right palsy case. (a) raise eyebrows, (b) close eye gently, (c) close eye tightly, (d) screw-up nose, (e) big smile.

4. RESULTS AND DISCUSSION

The Radial Basis Functions (RBF) network is used in our work to quantify the facial nerve function. There are 197 subject videos in our database with the H-B and regional grade evaluated by clinical expert. 118 subjects are used to train the networks and 79 to test the networks. Five RBF networks are trained for the five movements respectively. Each has four inputs: (i) the ratio of the maximum motion magnitude between of two sides, (Eq 2); (ii) the illumination compensation factor (Eq 3); (iii) Sym_y (Eq 4); (iv) Sym_r (Eq 5). The outputs from each regional network are the estimated palsy grade for each facial feature region. These are then used as the inputs for the overall RBF network to generate the H-B overall palsy grade.

Outputs are graded from 1 to 6, with 6 representing severe palsy and 1 being normal. The performance of the method for training and test data is given in Table1 and Table 2 respectively. The first columns in the tables give the percentage of the estimated values which are the same as the expert's assessments. The other columns show the percentages where the disagreement is from 1 to 5 grades respectively. For the forehead and overall H-B grade, there over 60% agreement between estimated value and the expert's assessment and less than 10% with a difference of more than 1 grade. In the other region, the agreement is between 40% and 60%.

Table 1 Training Data Performance

Disagreement	0	1	2	3	4	5
Forehead	64.41	28.81	5.08	1.69	0	0
Eye gentle	47.46	42.37	8.47	0.85	0.85	0
Eye tight	46.61	41.53	10.17	0.85	0.85	0
Nose	48.31	37.29	12.71	1.69	0	0
Mouth	46.61	38.14	11.86	3.39	0	0
H-B	61.02	33.05	5.08	0.85	0	0

Table 2 Test Data Performance

Disagreement	0	1	2	3	4	5
Forehead	62.03	29.11	6.33	2.53	0	0
Eye gentle	46.84	45.57	7.59	0	0	0
Eye tight	40.51	49.37	7.59	2.53	0	0
Nose	58.23	26.58	15.19	0	0	0
Mouth	51.9	36.71	10.13	1.27	0	0
H-B	64.56	31.65	3.8	0	0	0

The most encouraging aspect of these results is that the agreement within one grade between the estimated value and the expert's assessment was around 90% for regional grading and 95% for the H-B overall grading. The best that clinical assessment alone can achieve is usually an inter or intra observer variation of at least one grade. The system is objective and stable. It provides the same regional results and H-B grade during analyzing of different videos taken from the same subjects on the same day whereas clinicians have inconsistent assessments. The results show the best agreement in the forehead region as in this region the optical flow can be estimated with a high degree of accuracy. The error of optical flow estimation in the other regions is the major reason for the disagreement being greater than 1. Another reason is that the subjects who can not finish the prescribed movements correctly have introduced errors. The disagreements between the clinical and the estimated H-B values are greater than 1 grade only when the regional results introduce a higher average error.

The proposed algorithms have been implemented in Java, with JMF (Java Media Framework) and ImageJ. The average video with 500 frames on a 1.73GHz laptop can be processed in 3 minutes. The overall processing time should satisfy the real-time requirement.

5. CONCLUSION

In this paper we have proposed an automatic system that combines facial feature detection, face motion extraction and facial nerve function assessment by RBF networks. The results of regional evaluation in forehead and the overall H-B grade are more reliable. The errors are mainly introduced by nonstandard facial movements and the incorrect estimation of the optical flow. Therefore encouraging patient to perform the key movements correctly and a more accurate estimation of optical flow should improve the performance of the system. The present results are very encouraging in that they indicate that it should be possible to produce a reliable and objective method of measuring the grade of a facial palsy in the clinical setting.

6. REFERENCES

[1] JW. House, "Facial nerve grading systems," *Laryngoscope*, 93: pp. 1056-1069, 1983.

[2] C. H. Beurskens, P. G. Heymans, "Positive Effects of Mime Therapy on Sequelae of Facial Paralysis: Stiffness, Lip Mobility, and Social and Physical Aspects of Facial Disability," *Otology & Neurotology,* 24(4):pp677-681, July 2003.

[3] S. McGrenary, B. F. O'Reilly and J. J. Soraghan , "Objective Grading of Facial Paralysis Using Artificial Intelligence Analysis of Video Data", *18th IEEE Symposium on Computer-Based Medical Systems,* pp. 587-592, 2005.

[4] J. Rurainsky, P. Eisert, "Template-based Eye and Mouth Detection for 3D Video Conferencing," *Lecture Notes in Computer Science, Springer,* vol. 2849, pp. 23-31, September 2003.

[5] S. M. Smith and J.M. Brady, "SUSAN - A New Approach to Low Level Image Processing," *International Journal of Computer Vision*, 23(1): 44-78, May 1997.

[6] S. Baker, I. Matthews. "Lucas-Kanade 20 years on:A unifying framework," *International Journal of Computer Vision*, 56(3):pp221–255, 2004.