# Automatic Facial Analysis for Objective Assessment of Facial Paralysis

Shu He, John J. Soraghan, Brian F O'Reilly and Stewart McGrenary

**ABSTRACT**

**Facial Paralysis is a condition causing decreased movement on one side of the face. A quantitative, objective and reliable assessment system would be an invaluable tool for clinicians treating patients with this condition. This paper presents an approach based on the automatic analysis of patient video data. Facial feature localization and facial movement detection methods are discussed. An algorithm is presented to process the optical flow data to obtain the motion features in the relevant facial regions. Three classification methods are applied to provide quantitative evaluations of regional facial nerve function and the overall facial nerve function based on the House-Brackmann Scale. Experiments show the Radial Basis Function (RBF) Neural Network to have superior performance.**

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

## I. INTRODUCTION

Facial Paralysis is a condition where damage to the facial nerve causes weakness of the muscles on one side of the face resulting in an inability to close the eye and dropping of the angle of the mouth. The commonest cause of facial palsy is a presumed herpes simplex viral infection, commonly referred to as Bell's palsy, which causes temporary damage to the facial nerve. Treatment of such viral infections has been the source of controversy in the past partly because it has been difficult to audit the effectiveness of treatment. Facial paralysis may also occur as a result of malignant tumors, Herpes Zoster infection middle ear bacterial infection, following head trauma or during skull base surgical procedures, particular in the surgical removal of acoustic neuroma, [1]. As the facial nerve is often damaged during the neurosurgical removal of these intracranial benign tumours of the hearing nerve facial nerve function is a commonly used indicator of the degree of success of the surgical technique. As most methods of assessing facial function are subjective there is considerable variability in the results between different assessors.

Traditional assessment of facial paralysis is by the House-Brackmann (H-B) grading system [2] which was proposed in 1983 and has been adopted as the North American standard for the evaluation of facial paralysis. 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 that it relies a subjective judgment with significant inter & intra observer variation [3-5] and it is insensitive to regional differences of function in the different parts of the face.

Several objective facial grading systems have been reported recently. These predominantly involve the use of markers on the face [5-7]. As the color of the physical markers are a contrasting color to that of skin, then simple threshold methods can be applied to locate the markers throughout the subjects facial movements. This makes the image processing simpler but there are negative implications as a trained technician has to accurately placing the markers on the same part of the face. The success and uptake of any automatic system will hinge on the ease of use of the technology [8]. Neely et al [9-11] and McGrenary et al [8] measured facial paralysis by the differences between the frames of a video. Although their results correlate with the clinical H-B grade, this method can not cope with irregular or paradoxical motion in weak side. Watchman et al. [12,13] measured facial paralysis by examining the facial asymmetry on static

images. They define the face midline by manually labeling three feature points: the inner canthus of each eye and philtrum and then measuring the intensity difference and edge difference between the two sides of the face. However this method cannot separate the intrinsic facial asymmetry caused by facial nerve dysfunction from the extrinsic facial asymmetry caused by orientation, illumination, shadows and the natural bilateral asymmetry.

In this paper, we present an automated, objective and reliable facial grading system. In order to assess the degree of movement in the different regions of the face the patient is asked to perform five separate facial movements, which are :- raise eyebrows, close eyes gently, close eyes tightly, screw-up nose and smile. 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 movements, going back to rest between each movement. A highly efficient face feature localization method is employed in the reference frame that is grabbed at the beginning of the video during the initial resting phase. The image of the subject is stabilized to compensate for any movement of the head by using block matching techniques. Image subtraction is then employed to identify the period of each facial movements. Optical flow is calculated to identify the direction and amount of movement between image sequences. The optical flow computation results are processed by our proposed method to measure the symmetry of the facial movements between each side of the face. These results combined with the total pixel intensity changes and an illumination compensation factor in the relevant facial regions are fed into classifiers to quantitatively estimate the degree of movement in each facial region using the normal side as the normal base line. Finally, the regional results are then fed into to another classifier to provide an overall quantitative evaluation of facial paralysis based on H-B Scale. Three classification methods were applied. Experiments show the Radial Basis Function (RBF) Neural Network has superior performance.

The paper is organized as follows. In Section II the face feature localization process is presented. In Sections III and IV image stabilization and key movements detection are introduced. In Section V the algorithms of the extraction of motion features are developed. In Section VI, the quantitative results obtained from three classification methods are compared and Section VII concludes the paper.

## II. Localization of Facial Regions

Many techniques to detect faces have been developed. Yang [14,15] classifies them into four categories: Knowledge-based, Feature-based, Template-based and Appearance-based. Template-based and Appearance-based methods can be extended to detect faces in cluttered background, different poses and orientation. However they need either lot of positive and negative examples to train the models or they need to be initialized manually and their computation is either time or memory intensive [15]. Our main objective is to develop an automatic assessment of facial paralysis for clinical use by measuring facial motion. In order to localize the facial features quickly, accurately and without any manual interaction, the patient is videotaped using a front face view with a clean background. Knowledge-based methods are designed mainly for face localization in uncluttered background but a method is proposed for facial feature localization. It processes a 720x576 image in 560 ms on a 1.73 GHz laptop. It was tested using 266 images in which faces have the in-plane rotation within $\pm 35$ degrees and achieved a 95.11% accuracy for all eight facial regions as shown in Fig.1 to be localized precisely.

The face area is segmented, the pupils are localized and the inter-pupil distance is then used to scale the size of each facial region. The middle point between the two pupils is used as a fulcrum to rotate the inter-pupillary line to the horizontal so that the face is made perpendicular in the image. Since most subjects and especially those with a facial palsy do not have bilateral symmetrical faces, the mouth may not be symmetric on the line of the pupil middle point. The mouth corners are therefore separately localized and the middle point of the mouth is assigned. The nasal regions are initially assigned by the positions of the pupils and the middle point of mouth. They are calibrated by minimizing the difference between the left and right side of nose. Finally, a face region map is assigned as shown in Fig.1.

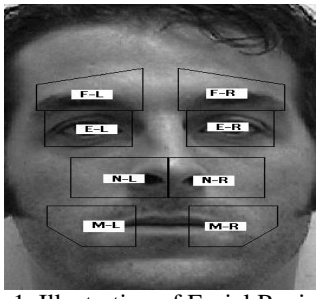


Fig.1. Illustration of Facial Regions.
F-forehead region; E-eye region; N-nasal region; M-mouth region, L-left, R-right

### *A. Face Boundary Search*

The face area has to be identified before starting the search for the face features. In our approach the subjects face is viewed frontally and is the only object in the frame. The face boundary can be detected by horizontal and vertical projections of an edge detected image. Fig.2 demonstrates that the left and right face boundaries are identified by vertical projection of a Sobel filtered image. Similarly, horizontal projection of the binary image is used to find the top boundary of face.

(a) Original Frame (b) Sobel filering (c) Vertical Projection
Fig.2. Face boundary detection using Sobel filter and Vertical projection

### *B. Detection of the ROI of eyes and mouth*

All the features of a face (eyebrows, eyes, nostril, mouth) are generally darker than the normal skin color [16] however hair may also be darker than facial features. A Gaussian filter is used to centre weight the head area. The intensity values of Gaussian weighted image can be expressed as

$$I(x,y) = I_{original}(x,y) * w(x,y) \quad (1)$$

where $I_{original}(x,y)$ denotes the intensity value of original image at pixel (x,y) , and $w(x,y)$ is computed as

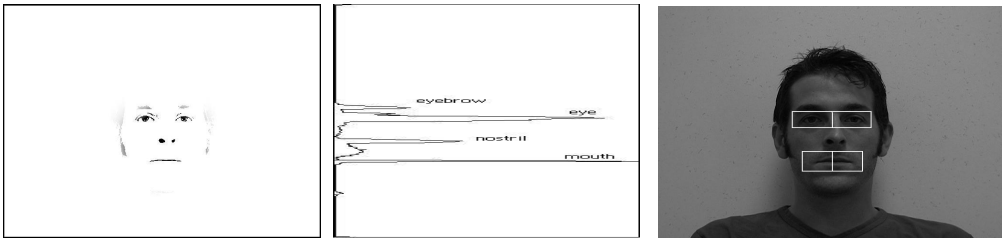
$$w(x,y) = e^{\frac{(x-x_o)^2+(y-y_o)^2}{2*((x_{right}-x_{left})/3)^2}} \quad (2)$$

where $x_{right}$ and $x_{left}$ are the horizontal position of right and left face boundaries. The centre of the face $(x_o, y_o)$ can be estimated as

$$x_o = x_{left} + (x_{right} - x_{left})/2 \quad (3)$$

$$y_o = y_{top} + (x_{right} - x_{left}) * 3/4 \quad (4)$$

since the height of face is approximately 1.5 times of the width. The ROI (region of interest) of the head is assigned with the $x_{right}$ , $x_{left}$ , $y_{top}$




(a). Gaussian filtering of face (b). Horizontal projection (c) ROI of eyes and mouth

Fig.3. Detection of the vertical position of facial features

Due to varied skin color and lighting conditions, a dynamic threshold is applied to the image such that only those facial features information is included for analysis. It is obtained by the solution of Eq. 5.

$$\frac{1}{N} * \sum_{i=Threshold}^{M} C(i) = 0.1 \qquad (5)$$

Here the threshold is set to a value that only 10% of pixels present since the irises, nostrils and the mouth border occupy no more than 10% of the ROI of head. N is the number of pixels in the ROI of head. $C(i) = Histogram\ (ROI_{head})$. $M = 255$ if working on a 8-bit images.

An example of an inverted, thresholded, Gaussian weighted image is shown in Fig.3.(a). The vertical position of eyebrow, eye, nostril, mouth can be determined by its horizontal projection as shown in Fig.3.(b). In some cases the eyebrow or nostril may not be identified but only the pupils and mouth corners are the essential key points necessary to assign the facial map. With the eyes and mouth vertical position and the face borders, the ROI of eyes and mouth on each side can be set to allow refining of the pupils and mouth corners positions.

### *C. Pupil Search*

This approach is based on the characterization of the iris and pupil. The iris-pupil region is dark compared to the white of the sclera of the eye ball and to the luminance values of the skin color. The iris 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 [17]. The eye width to eye height relation can be expressed as approximately 3:1 and the eye height is interpreted as the iris diameter [18]. Therefore, the eye template can be created as shown in Fig.4.(a). This eye template is scaled automatically depending on the size of the face area. The iris is roughly localized by searching the minimum difference between the template and the ROI of the eye. The pupil is darker than the iris and therefore its position can be determined by searching the small circle with the lowest intensity value within the iris area. Here the diameter of small circle is set to be 1/3 of the iris diameter.

(a) eye template (b) detected pupil centre

Fig.4. Pupil centre detection

### *D. Mouth Corner Search*

The mouth corners are detected by applying the SUSAN (Smallest Univalue Segment Assimilating Nucleus) algorithm for corner extraction [19] 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 around the nucleus. The points from the neighbourhood whose brightness are approximately the same as the brightness of the nucleus form the area referred to as USAN( Univalue Segment Assimilating Nucleus). The point (nucleus) with smallest USAN area indicates the corner. In Fig.5, the USANs are shown as grey parts and the upper left one is SUSAN. Usually, more that one point is extracted as a corner and these points are called mouth corner candidates. Three knowledge-based rules are applied to these points. First, the left corner candidates

are eliminated if their horizontal distance from the middle of the pupil line is greater than 70% of the width of the search region and a similar rule is employed to the right candidates. Second, the candidates are eliminated if the horizontal distance between a left and right corner candidate is greater than 150% of the inter-pupil distance or less than 50% of the inter-pupil distance. Third, among the remaining left candidates the one located furthest to the left is considered to be the left mouth corner and a similar rule is employed to the right candidates [20]. An example of detected the mouth corners is shown in Fig.6.

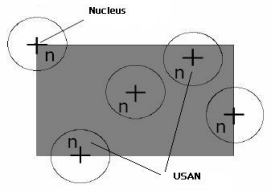


Fig.5. SUSAN corner detector

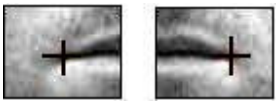


Fig.6. detected the mouth corners

## III. IMAGE STABILIZATION

Subjects will raise their head spontaneously when asked to raise there eyebrows and also shake their head while smiling. Before measuring facial motion, these rigid global motions need to be removed so that only the non-rigid facial expressions are kept in the image sequences for analysis. Feature tracking is normally considered to help solve this problem. A set of features are tracked through the image sequence and their motion is used to estimate the stabilizing warping [21]. However in our work there are no key features in the face which do not change in structure when the movements are carried out. Therefore, all facial features are used for tracking. A ROI of the face encompassing the eyebrows, eyes, nose and mouth in the reference frame is defined by the position of the pupils, mouth corners and inter-pupils distance, as shown Fig.7. The image is stabilized by finding the best matched ROI of the face between the reference and the subsequent frames. The affine transformation given by Eq.6 is performed on the subsequent frame. Image stabilization can be formulated as a minimization problem given by Eq.7.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} dx \\ dy \end{bmatrix} \tag{6}$$

Here $(x', y')$ is the original image coordinates, which is mapped to the new image at $(x, y)$. $dx, dy$ are the horizontal and vertical displacements. $\theta$ is the rotation angle. The scaling factor is not included as the distance between the subject and the camera is fixed and the face maintains constant size through images sequences in our application.

$$(dx_n^*, dy_n^*, \theta_n^*) = \arg\min_{dx,dy,\theta} \sum_{(x,y)\subset ROI} \sum \left| T_n(x, y) - I_{ref}(x, y) \right| \tag{7}$$

$dx_n^*, dy_n^*, \theta_n^*$ are the optimal transformation parameters for the frame n. $I_{ref}(x, y)$ is the intensity of pixel at $(x, y)$ in reference frame. $T_n(x, y)$ is the intensity of pixel at $(x, y)$ in the warped frame n. ROI denotes the ROI of face. $dx_n, dy_n, \theta_n$ are initialized to the optimal values in the last frame $dx_{n-1}^*, dy_{n-1}^*, \theta_{n-1}^*$.

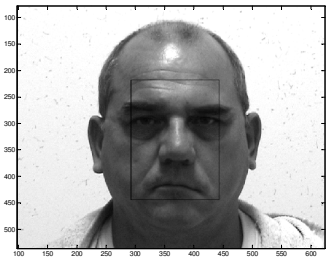


Fig.7. the ROI of the face in the reference frame

## IV. KEY MOVEMENTS DETECTION

To examine the five key movements in the relevant regions the timings of the five movements are identified. An algorithm based on image subtraction is proposed to determine the start and end of each movement so that information is only extracted from the appropriate time in the videos and from the appropriate facial region. The video sequence begins with the subject at rest followed by the five key movements and going back to rest inbetween each movement. Therefore, the rest frames between movements have to be detected as splice points. This is achieved by totalling up several smoothed and varying thresholded pixel changes until five peaks and four valleys of sufficient separation can be extracted. The equation to produce the line from which the splice points can be detected is given in Eq.8 as follows:

$$Y(n) = smooth \sum_{m=0}^{4} \sum_{(x,y)\subset ROI} \sum thresh\left(\left|I_n(x,y) - I_{ref}(x,y)\right|, (0.1 + 0.02m)\right) \quad (8)$$

Here $I_n(x,y)$ and $I_{ref}(x,y)$ is the intensity of pixel $(x,y)$ at the n$^{th}$ frame and the reference frame. The ROI is the face region, defined in session III. m is index for threshold level. 0.1 is an empirical threshold bias to keep the high intensity changes and remove the small pixel changes those may be produced by noise. The varying intensity of motion can be detected by changing m. By summing the different intensity of motions, the peak of motion is obvious and the splice points are easy to detect.

An example of $Y(n)$ is the highest curve in Fig.8 while the rest five curves from up to bottom are the plots at m=0 to 4 respectively. The splice points are shown as the dotted lines in Fig.8 The five displacement peaks of movement correspond to the five key movements in the exercise: raising eyebrows, closing eyes gently, closing eyes tightly, scrunching nose, big smile.

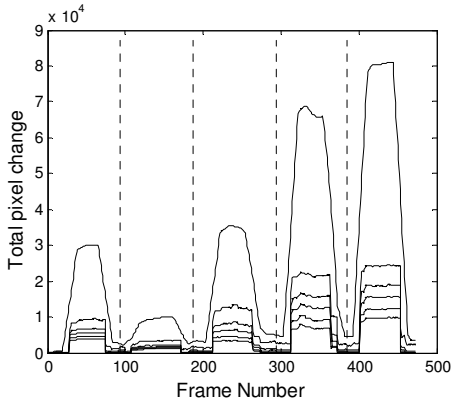



Fig.8. Total of Thresholded, Smoothed, Pixel Displacements

## V. REGIONAL FACIAL MOVEMENT ANALYSIS

### *A. Motion magnitude by image subtraction*

Neely et al showed that image subtraction is a viable method of quantifying facial paralysis [9,10]. This method is therefore used to measure the motion magnitude of each key movement in the relevant region. Fig.9.(a) shows a reference frame grabbed with the subject at rest. Fig.9.(b) shows the frame with the subject raising eyebrows. Fig.9.(c) is the difference image between (a) and (b). The pixel is bright if there have been pixel changes and it is black if there has been no change. From Fig.9.(c) it is clear that there are some changes in the forehead with no difference in the areas of the nose or mouth.

It has been observed that in general the more light that falls on a region of the face, the more that changes will be detected and that the results of video taken in nonhomogeneous lighting conditions may be skewed. In our work after the facial map is defined in the reference frame, the ratios of the intensity mean values between left side and right side in the relevant regions are calculated and then used as illumination compensation factors to adjust subsequent frames. Fig.9 illustrates a frame taken in nonhomogeneous lighting conditions. The original image lighting conditions are shown in Fig.9.(a). This subject has almost completely recovered except for a mild weakness of the eye and mouth on the right side. Fig.9.(c) shows difference between image (a) and (b). Note that the left side of the image is the subject's right side. Here it is obvious that more changes are detected on the left side of the forehead than on the right side. Fig.9.(d) is the difference between images after the illumination compensation for the forehead region has been applied. The highlighted areas have the similar intensity, i.e. similar movement magnitude. The movement magnitude in the relevant region can be computed by Eq.9 as:

$$mag(n) = \sum_{(x,y)\subset R}\sum \left| I_n(x,y) - I_{ref}(x,y) \right| * w(x,y) * lum \qquad (9)$$

where $w$ is the Gaussian weights, similar to the Eq. 1 , but set $(x_o, y_o)$ to be the centre of the region, $x_{right}$ and $x_{left}$ are right and left boundaries of the region and $lum$ is the illumination compensation factor, which is set to

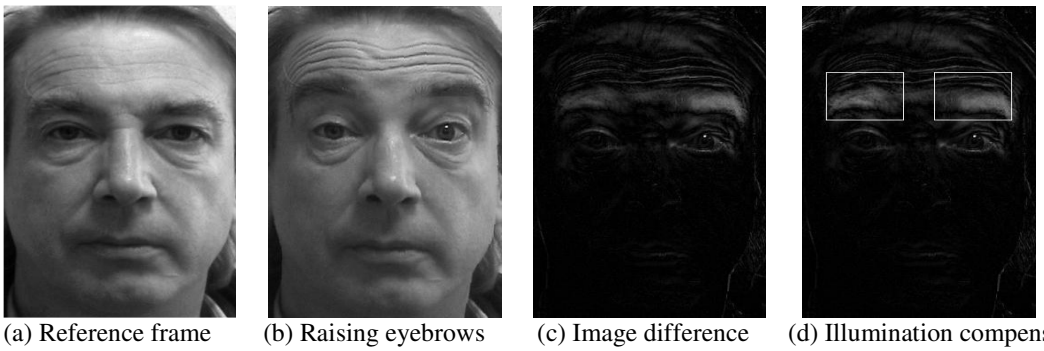
$$lum = \sum_{(x,y)\subset left}\sum I_{ref}(x,y) / \sum_{(x,y)\subset right}\sum I_{ref}(x,y) \qquad (10)$$

for right side and lum = 1 for left side.

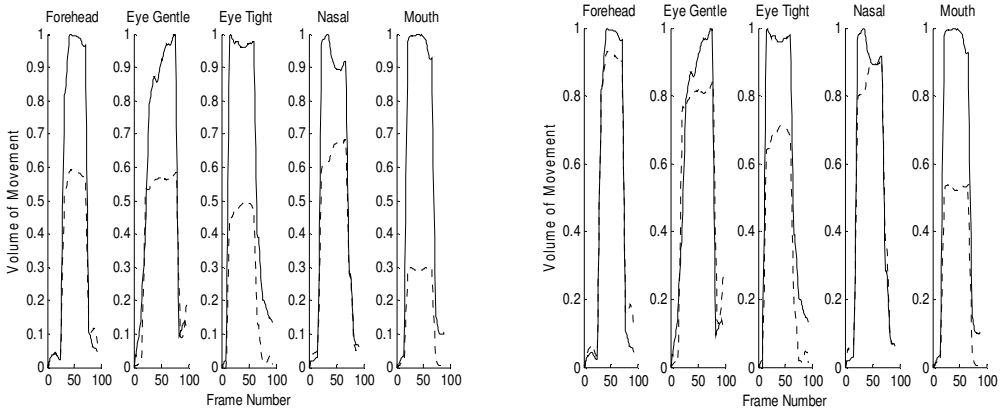
The graphs shown in Fig.9.(e) demonstrate the full displacement results for an almost recovered subject with mild weakness at the right side of the eye and mouth. Five plots in Fig.9.(e) show the magnitude of the five movements in the relevant facial region. The broken line indicates the detected movement on the subjects' right side of the face and the solid line indicates the movement detected on the left. The x axis shows the frame count for each movement and the y axis indicates the proportional volume of movement from the reference frame; the normal side being standardized to 1. The output from the forehead and nose show similar responses for the left and right side but the movement amplitude for the eye and mouth region for right side is weaker than left side.

Fig.9.(e) and (f) compare the results with and without illumination compensation. Fig.9.(e) indicates that detected motion on the right is significantly less than the left while Fig.9.(f) shows similar movement magnitude for both sides except for the eye and mouth, which is in keeping with the clinical situation.

The illumination compensation factors, which are the ratios of the intensity mean values between the left and right side for each region, are between 0.56 and 1.8 for all the subjects videos in our study. This illumination compensation method is very effective in correcting the magnitude but it needs to be investigated further whether the illumination compensation factors can be used linearly to adjust the intensity for those videos with ratios out of this range.



(a) Reference frame (b) Raising eyebrows (c) Image difference (d) Illumination compensation




(e) without illumination compensation (f) with illumination compensation

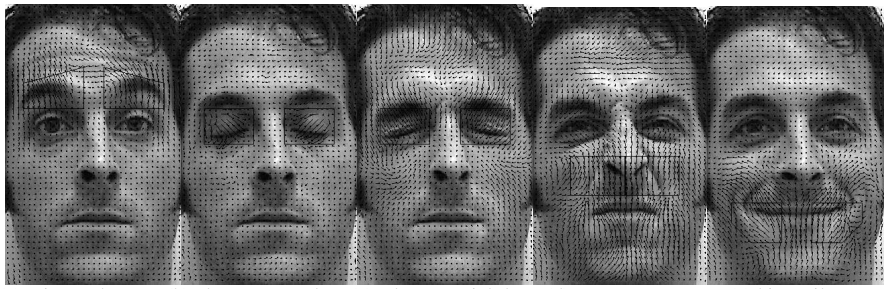
Fig.9. Illustration of the solution of varying illumination

### *B. Motion measurement by Optical Flow*

The magnitude of the movement on each side of the face (i.e. Fig.9(f)) is a very effective way to compare the motion intensity between the normal and the weak side of the face. However it does not take into account the direction of motion. For a normal subject, the amount of motion in the relative directions on each side of the face is similar. As shown in Fig.10.(e), for a normal subject producing a smile, the amount of motion in the up-left direction on the left side of the image is close to the amount in the up-right direction on the right side of the image. Fig.10.(e) shows a left palsy subject asked to smile. Although the left side has a severe paralysis, motion on the left side of the mouth is detected as the left side is drawn to the right by the movement of the right. Therefore, not only should the motion intensity be measured but the direction should also be taken into account when assessing the degree of palsy.

Optical flow is an approximation of the velocity field related to each of the pixels in an image sequence. Such a displacement field results from the apparent motion of the image brightness in time [22]. In a highly textured region it is easy to determine optical flow and the computation converges very fast because there are high gradients in many directions at every location. Optical flow to track facial motion is advantageous because facial features and skin have a great deal of texture. There are many methods for the estimation of optical flow. Barron et al. [23] classify optical flow algorithms by their signal-extraction stage. This provides four groups: differential techniques, energy-based methods, phase-based techniques and region-based matching. They compared several different methods and concluded that the Lucas and Kanade algorithm is the most accurate.

The Lucas and Kanade algorithm [24], including the pyramid approach is employed to compute 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.10-12 show the results of optical flow estimation for a normal subject, a left palsy subject and a right palsy subject. In Fig.10 the motion flows are approximately symmetrical between two highlighted regions. There is almost no motion in the left side of the forehead and the nose in Fig.11.(a) and (d), whereas there is obvious flow towards right on the left side of mouth in Fig.11.(e). Note that the right side of image is the subject's left side. Fig.12 shows a subject who cannot close his eye but when attempting to do so his iris moves upward. Although this movement of the iris is detected by the image subtraction method it should be discriminated from the motion of the eyes closing and removed from the calculation of the degree of movement. Fig.12.(b) and (c) shows little flow detected in the right eye confirming the severe palsy in the eye region.

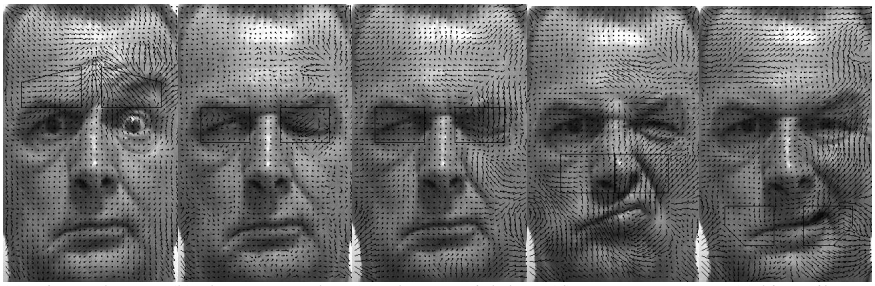


(a) raise eyebrows, (b) close eye gently, (c) close eye tightly, (d) screw up nose, (e) big smile

Fig.10 Results of optical flow estimation on five frames with peak motion for a normal case

(a) raise eyebrows, (b) close eye gently, (c) close eye tightly, (d) screw up nose, (e) big smile

Fig.11 Results of optical flow estimation on five frames with peak motion for a left palsy case



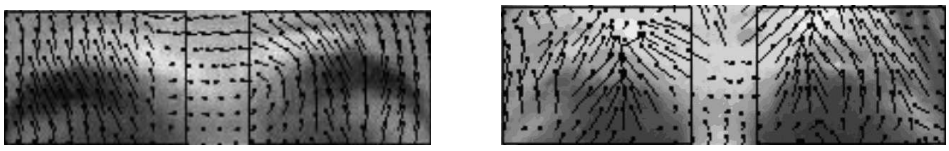
(a) raise eyebrows, (b) close eye gently, (c) close eye tightly, (d) screw up nose, (e) big smile

Fig.12 Results of optical flow estimation on five frames with peak motion for a right palsy case

In each facial feature region, the flow magnitude is thresholded to reduce the effect of small computed motions which may be either produced from textureless areas or affected by illumination and the flow magnitude is centre weighted by a Gaussian filter. Given the thresholded flow vector $\vec{v}_i = (u_i, v_i)$ in the region, the overall flow vector of each region can be expressed as $\vec{v} = (u, v)$, the components of vector, $u$ and $v$, denote of overall displacement on the horizontal and vertical direction. $u = \sum_i u_i * w_i$, $v = \sum_i v_i * w_i$ here $w_i$ is the Gaussian weights, similar to the Eq. 1 , but set $(x_o, y_o)$ to be the centre of the region, $x_{right}$ and $x_{left}$ are right and left boundaries of the region.

When subjects raise their eyebrows, close their eyes or screw up their nose, the muscles in each relevant region move mainly in the vertical direction. Studies have show that even for normal subjects neither the amplitude nor the orientation of horizontal displacements on each side are consistently symmetrical. Fig.13 shows two normal subjects raising their eyebrows. In Fig13.(a), the mean horizontal displacements are negative for both sides, i.e. in the same direction , while in Fig13.(b), the mean horizontal displacements on each side are opposite. In Fig.13.(a), the amplitude of the mean horizontal displacements in the left side is

larger than that in the right side, while in Fig.13.(b) they are similar. The movement in the horizontal direction does not contribute much information when measuring the symmetry of the eyebrow, eyes or nose movements. Therefore, the displacements strength and the vertical displacements only are used for these symmetry measurements. The symmetry of the facial motion is quantified by Eq.11 and 12.



(a) normal subject I (b) normal subject II

Fig.13. Results of optical flow estimation on forehead for two normal subjects

$$Sym_y = 1 - \frac{\left| v_{left} - v_{right} \right|}{\left| v_{left} \right| + \left| v_{right} \right|} \tag{11}$$

$$Sym_r = 1 - \frac{\left\| \left\| \vec{v}_{left} \right\| - \left\| \vec{v}_{right} \right\| \right\|}{\left\| \vec{v}_{left} \right\| + \left\| \vec{v}_{right} \right\|} \tag{12}$$

where the $v_{left}$ and $v_{right}$ are the overall vertical displacements for left side and right side, $\vec{v}_{left}$ and $\vec{v}_{right}$ are the overall flow vector for left side and right side. $Sym_y$ and $Sym_r$ will be within the range 0-1. The motions on each side of face are symmetrical when both approximate 1. When both approximate 0 the dysfunctional side has no movement at all. While when $Sym_y$ = 0 and $Sym_r$ = 1 indicates the motion on each side are the same amplitude but opposite direction, i.e. one eye closed, the other eye can not close but the iris moves upwards in the presence of severe paralysis.

The muscle around the mouth will move to the side of face when normal people smile. The horizontal displacements should be negative, i.e. move towards the left on the left side of mouth and should be positive, i.e. move towards the right on right side. This is used as a constraint to calculate the overall flow vector of left region and right region, formulated as Eq 13.

$$u_{left} = \sum_{u_i<0} u_i * w_i \qquad v_{left} = \sum_{u_i<0} v_i * w_i \qquad u_{right} = \sum_{u_i>0} u_i * w_i \qquad v_{right} = \sum_{u_i>0} v_i * w_i \tag{13}$$

In the left mouth region, each motion vector with a negative horizontal displacement is taken into account. Only those with the positive horizontal displacement are taken into account for the right side. This allows elimination of the apparent muscle movement on the weak side produced by the muscles on the normal side as in Fig.11.(e) and Fig 12.(e).

This method was test in 197 videos. $Sym_y$ and $Sym_r$ are correlated with H-B grade around 0.83 in the forehead and around 0.7 in the rest of region. Details are shown in Table 1.

**TABLE 1 CORRELATION ANALYSIS BETWEEN $Sym_y$ , $Sym_r$ AND H-B GRADE**

| | CORR WITH $Sym_y$ | CORR WITH $Sym_r$ |
|---|---|---|
| Forehead | 0.8303 | 0.8381 |
| Eye gentle | 0.7351 | 0.7603 |
| Eye tight | 0.6968 | 0.7071 |
| Nose | 0.6981 | 0.7199 |
| Mouth | 0.7158 | 0.7206 |

## VI. QUANTITATIVE ASSESSMENT AND EXPERIMENTS

### *A. Quantitative Assessment*

To map the motion magnitude and optical flow information into a H-B grade is a classification problem. There are a number of classification methods. k-Nearest-Neighbor (k-NN), Artifical neural network (ANN) and Surport Vector Machine (SVM) are the most widely used classifiers. They can be used successfully for pattern recognition and classification on data sets with realistic sizes. These three classification methods were employed for the quantitative assessment of regional paralysis and the over-all facial paralysis.

The H-B grades the overall facial nerve function and it is insensitive to small changes in each facial region.

The regional facial function is measured by examining the key movements in the relevant region and classified to six grades from 1(normal) to 6 (total paralysis). Five classifiers are trained for the five movements respectively. Each has four inputs as follows:

1. $\arg\min(mag_{left}, mag_{right}) / \arg\max(mag_{left}, mag_{right})$

   Here $mag_{left}, mag_{right}$ denotes the total relative pixel change in the region from the resting face to the peak of the movement, which can be calculated using Eq.9. The input value computed here gives the ratio of the total pixel change between the dysfunctional side and the normal side.
2. The illumination compensation factor, calculated by Eq.10, which is the ratio of mean intensities for each region between the dysfunctional side and the normal side. Although the illumination compensation factors can be used to correct the magnitude if it is between 0.56 and 1.8, the performance of this linear compensation is not ideal. As shown in Fig.9.(d), the two highlighted regions have the similar intensity but are not identical. In order to further compensate for the illumination, the illumination factor is included as an input to the classifier.
3. $Sym_y$, defined by Eq.11, represents the symmetry relative to the vertical component of the total amount of displacements from the resting face to the peak of movement.
4. $Sym_r$, defined by Eq.12, represents the symmetry relative to the strength of the total amount of displacements from the resting face to the peak of movement.

Outputs are graded from 1 to 6, with 6 representing severe palsy and 1 being normal. These regional results are then used as the inputs for the overall classifier to analyze the H-B overall palsy grade.

### *B. Experiments*

There are 197 subject videos in our database taken from subjects with Bell's palsy, Ramsey Hunt Syndrome, trauma and other aetiologies as well as normal subjects. Their H-B and regional gradings were evaluated by a clinician. As the dataset was not large the Leave-k-out-cross-validation test scheme instead of k-fold was adopted.
Multilayer Perceptron (MLP) network and Radial Basis Function (RBF) network are the most popular neural network architectures [24,25]. Experiments show RBF networks provide consistently better performance than MLP networks for facial palsy grading. The centers of each RBF NN were initialized using the k-means clustering algorithm before starting training.

Table 2, Table 3 and Table 4 present the average classification performance, in percentages, for the 20 repetitions of the leave-k-out cross-validation, with k = 20. The numbers in the first columns give the percentage of the results which are the same as the clinician's assessments. Columns 2-6 show the percentages where the disagreement is from 1 to 5 grades respectively. The last columns show the percentage of the disagreement within 1 grade. The comparison of the performance is graphically illustrated in Fig.14. The results show the RBF NN outperforms the k-NN and SVM. The disagreement within one grade between the results of the RBF NN and the clinical assessment is 94.18% for the H-B overall grading, which is 5.38% higher than SVM and 10.71% higher than k-NN. The variation of the

performance for RBF NN is similar to that of SVM. Both RBF NN and SVM provide more stable results than the k-NN. The variation of the results of disagreement within 1 grade is shown in Table 5.

The RBF network has similar structure as SVM with Gaussian kernel. RBF networks are typically trained in a Maximum Likelihood framework by minimizing the error. SVM takes a different approach to avoiding overfitting by maximizing the margin. Although SVM outperforms RBF networks from the theoretical view, they can be competitive when the dimensionality of the input space is small. There are only 4 or 5 inputs in our work. The centers of each RBF NN were initialized using the k-means clustering algorithm before starting training. Experiments show RBF networks can discover the nonlinear associations better than SVM and k-NN in our application.

**TABLE 2 TEST DATA PERFORMANCE OF RBF NN**

| Disagreement | 0 | 1 | 2 | 3 | 4 | 5 | <=1 |
|---|---|---|---|---|---|---|---|
| Forehead | 68.71 | 24.52 | 4.02 | 2.75 | 0 | 0 | 93.23 |
| Eye gentle | 44.36 | 47.90 | 3.23 | 4.51 | 0 | 0 | 92.26 |
| Eye tight | 41.27 | 51.84 | 3.69 | 3.20 | 0 | 0 | 93.11 |
| Nose | 61.78 | 24.53 | 10.18 | 3.51 | 0 | 0 | 86.31 |
| Mouth | 49.80 | 38.22 | 8.43 | 3.55 | 0 | 0 | 88.02 |
| H-B | 63.92 | 30.26 | 5.82 | 0 | 0 | 0 | 94.18 |

**TABLE3 TEST DATA PERFORMANCE OF K-NN**

| Disagreement | 0 | 1 | 2 | 3 | 4 | 5 | <=1 |
|---|---|---|---|---|---|---|---|
| Forehead | 65.33 | 21.12 | 5.79 | 4.61 | 3.15 | 0 | 86.45 |
| Eye gentle | 39.91 | 44.57 | 7.73 | 5.90 | 1.89 | 0 | 84.48 |
| Eye tight | 36.73 | 45.62 | 6.65 | 5.28 | 5.72 | 0 | 82.35 |
| Nose | 55.68 | 21.83 | 13.47 | 7.11 | 1.91 | 0 | 77.51 |
| Mouth | 44.01 | 35.19 | 11.62 | 8.35 | 0.83 | 0 | 79.20 |
| H-B | 58.13 | 25.26 | 12.88 | 3.73 | 0 | 0 | 83.39 |

**TABLE4 TEST DATA PERFORMANCE OF SVM WITH GAUSSIAN RADIAL BASIS FUNCTION KERNEL**

| Disagreement | 0 | 1 | 2 | 3 | 4 | 5 | <=1 |
|---|---|---|---|---|---|---|---|
| Forehead | 65.98 | 25.14 | 7.22 | 1.66 | 0 | 0 | 91.12 |
| Eye gentle | 41.05 | 44.32 | 9.81 | 4.82 | 0 | 0 | 85.37 |
| Eye tight | 38.79 | 48.33 | 10.88 | 2.10 | 0 | 0 | 87.12 |
| Nose | 59.28 | 23.45 | 14.32 | 2.95 | 0 | 0 | 82.73 |
| Mouth | 43.37 | 37.11 | 17.58 | 1.94 | 0 | 0 | 80.48 |
| H-B | 59.73 | 29.07 | 11.2 | 0 | 0 | 0 | 88.80 |

**TABLE5 THE VARIATION OF THE PERFORMANCE ( DISAGREEMENT<=1)**

| | RBF | k-NN | SVM |
|---|---|---|---|
| Forehead | 5.6% | 8.1% | 7.1% |
| Eye gentle | 11.3% | 14.6% | 10.7% |
| Eye tight | 11.0% | 15.2% | 10.2% |
| Nose | 9.1% | 13.5% | 8.0% |
| Mouth | 10.2% | 14.2% | 9.3% |
| H-B | 7.1% | 13.3% | 7.8% |

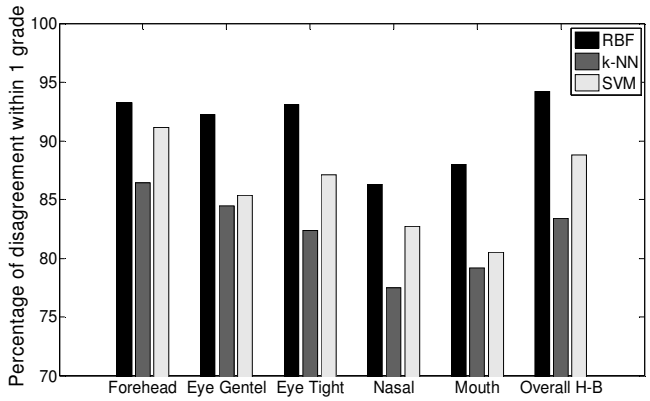



Fig. 14. Comparison of the performance of three RBF, k-NN, SVM

### *C. Discussion*

The most encouraging aspect of these results is that the disagreement within one grade between the results of the RBF NN and the clinical assessment was around 90% for regional grading and 94% 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 as 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 subjects who could not finish the prescribed movements correctly failed to be correctly classified. The patients were asked to practice the prescribed facial movements before being videotaped. These practice runs help minimize the non-correspondence error.

The results show the best agreement is in the forehead region as in this region the optical flow can be estimated with a high degree of accuracy. The estimation of the optical flow in the eye region has poor performance, especially for those faces with makeup or very irregular wrinkles on the eyelids. The structure of the eyebrows does not change significantly during raising of the eyebrows but the structure of eyes changed significantly when performing eye closure. The error of optical flow estimation in the other regions is the major reason for their disagreement being greater than 1 grade. More effective algorithms for the optical flow estimation should be investigated to offer more reliable results and for better performance of the networks for regional measurement. 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 can be processed in 3 minutes on a 1.73 GHz laptop. This overall processing time should satisfy the requirement of the practicing physician.

## VII. CONCLUSION

We have proposed an automatic system that combines facial feature detection, face motion extraction and facial nerve function assessment by RBF networks. The total pixel change was used to measure the magnitude of motion. The optical flow is computed and analyzed to identify the symmetry relative to strength and direction on each side of the face. RBF neural networks are applied to offer regional palsy grades and H-B overall palsy grade. The results of regional evaluation in forehead and the overall H-B grade are the 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 encouraging in that they indicate that it should be possible to produce a reliable and

objective method of measuring the degree of a facial palsy in a clinical setting.

## REFERENCES


[1] I. Frew, "The Facial Nerve," ISBN 0-192-61128-3.
[2] JW. House, "Facial nerve grading systems," *Laryngoscope*, 93: pp. 1056-1069, 1983.
[3] 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.
[4] J. B.Kahn, R. E.Gliklich, "Validation of a Patient-Graded Instrumnet for Facial Nerve Paralysis: The FaCE Scale," *Laryngoscope,* 111(3):pp387-398, March 2001.
[5] C. Linstrom, "Objective Facial Motion Analysis in Patients With Facial Nerve Dysfunction," *Laryngoscope,* 112(7):pp1129-1147, July 2002
[6] H. Scriba, S.J. Stoeckli, and D. Veraguth, "Objective Evaluation of Normal Facial Function," *Ann Otol Rhinol Laryngol*. 108(7 Pt 1):pp.641-644, 1999.
[7] P. Dulguerov, F. Marchal and D. Wang, "Review of Objective Topographic Facial Nerve Evaluation Methods," *American Journal of Otology*, 20(5):pp.672-8. 1999.
[8] 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.
[9] JG Neely, AH Joaquin Kohn LA and JY Cheung, "Quantitative assessment of the variation within grades of facial paralysis," Laryngoscope,106:438- 442 1996
[10] T. D.Helling, JG Neely, "Validation of objective measures for facial. Paralysis",. Laryngoscope, 107(10):1345-9, 1997
[11] JG Neely, "Advancement in the Evaluation of Facial Function," *Advances in Otolaryngology- Head and Neck Surgery,* vol 15 pp 109-134 , Jan 2002.
[12] G.S. Wachtman, Y. Liu, T. Zhao, J. Cohn, K. Schmidt, T.C. Henkelmann, J.M. VanSwearingen, and E.K. Manders, "Measurement of Asymmetry in Persons with Facial Paralysis" Combined Annual Conference of the Robert H. Ivy and Ohio Valley societies of Plastic and Reconstructive Surgeons, June, 2002.
[13] Y. Liu, K. Schmidt, J. Cohn, and S. Mitra, "Facial Asymmetry Quantification for Expression Invariant Human Identification" Computer Vision and Image Understanding Journal, July, 2003.
[14] M.H. Yang, David Kriegman, and Narendra Ahuja, "Detecting Faces in Images: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 24, no. 1, pp. 34-58, 2002
[15] M.H. Yang, "Recent Advances in Face Detection", IEEE ICPR 2004 Tutorial, Cambridge, United Kingdom, August 22, 2004.
[16] G. C. Feng, P. C Yuen, "Multi Cues Eye Detection on Grey Intensity Image," *Pattern Recognition*, 34(5), pp 1033-1046, 2001.
[17] 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.
[18] L. G. Farkas, *Anthropometry of the Head and Face,* Raven Press, 1995.
[19] 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.
[20] *M. Hess and G. Martinez, "Facial Feature Extraction Based on the Smallest Univalue Segment Assimilating Nucleus (SUSAN) Algorithm", Picture Coding Symposium 2004 (PCS-2004), San Francisco, California, USA, December 2004*
[21] C. Guestrin, F. Cozman, " Image Stabilisation for Feature Tracking and Generation of Stable Video Overlays," tech. report CMU-RI-TR-97-42, Robotics Institute, Carnegie Mellon University, November, 1997.
[22] M. Elad , A. Feuer, "Recursive Optical Flow Estimation - Adaptive Filtering Approach," *19th IEEE Conference on Computer Vision,* Nov. 1999.
[23] J.L.Barron, D.J.Fleet, "Performance of optical flow techniques," *International Journal of Computer Vision,* (IJCV1994), 12(1):43-77.
[24] S. Baker, I. Matthews. "Lucas-Kanade 20 years on:A unifying framework," *International Journal of Computer Vision,* 56(3):221–255, 2004.
[25] W. Duch, N. Jankowski, "Transfer function : hidden possibilities for better neural networks" *ESANN'2001 proceedings-European Symposium on Artificcial Neural networks Bruges (Belgium),* pp.81-94, April 2001.
[26] Fausett LV, "Fundamentals of Neural Networks: Architectures, Algorithms and Applications," *Pearson Higher Educatio,*1993



**Shu He** was born in Xi'an, China, in 1975. She received B. Eng. degree in Electrical and Electronic Engineering from Xidian University in 1996 and M. Eng. degree from Chamers Univeristy of Technology in 2005. and a PhD in Electronic & Electrical Engineering from the University of Strathclyde in 2007. In between her degrees she worked as software engineer in the Institute of telecommunication technique, Ministry of Information Industry and a leading telecom equipment vendor in China. Her research interests include computer vision, image processing, visual information retrieval, pattern recognition.

**John J. Soraghan** (S'84–M'86–SM'96) received the B.Eng. (first class honors) and the M.Eng.Sc. degrees in 1978 and 1982, respectively, both from University College Dublin, Dublin, Ireland, and the Ph.D. degree in electronic engineering from the University of Southampton, Southampton, U.K., in 1989.From 1979 to 1980, he was with Westinghouse Electric Corporation. In 1986, he joined the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K., as a Lecturer in the Signal Processing Division. He became a Senior Lecturer in 1999, a Reader in 2001, and a Professor in 2003. From 1989 to 1991, he was Manager of the Scottish Transputer Centre, and from 1991 to 1995, he was Manager of the DTI Centre for Parallel Signal Processing. Since 1996, he has been Manager of the Texas Instruments' DSP Elite Centre in the University. He currently holds the Texas Instruments Chair in Signal Processing in the Institute of Communications and Signal Processing, Department of Electronic and Electrical Engineering, University of Strathclyde. His main research interests include advanced linear and nonlinear signal, image, and

video processing algorithms; wavelets and fuzzy systems with applications to telecommunications; biomedical; and remote sensing. He has supervised 30 Ph.D. students toward graduation, holds three patents, and has published over 240 technical papers.

**Brian F. O'Reilly** is a Consultant ENT Surgeon and Otologist in Gartnavel General Hospital and Consultant Neuro-Otologist in the Institute of Neurological Sciences, Southern General Hospital, Glasgow. He is Honorary Senior Lecturer, Glasgow University and President of the Scottish Otolaryngological Society. He is the member of the British Association of Otolaryngologists, Scottish Otolaryngology Society and the British Skull Base Society. His current research interests are Acoustic Neuroma growth and Facial Palsy.