Quantitative Analysis of Facial Paralysis Using Local Binary Patterns in Biomedical Videos

Shu He, *Student Member, IEEE*, John J. Soraghan, *Senior Member, IEEE*, Brian F. O'Reilly, and Dongshan Xing

***Abstract*—Facial paralysis is the loss of voluntary muscle movement of 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 a novel framework for objective measurement of facial paralysis. The motion information in the horizontal and vertical directions and the appearance features on the apex frames are extracted based on the local binary patterns (LBPs) on the temporal–spatial domain in each facial region. These features are temporally and spatially enhanced by the application of novel block processing schemes. A multiresolution extension of uniform LBP is proposed to efficiently combine the micropatterns and large-scale patterns into a feature vector. The symmetry of facial movements is measured by the resistor-average distance (RAD) between LBP features extracted from the two sides of the face. Support vector machine is applied to provide quantitative evaluation of facial paralysis based on the House–Brackmann (H-B) scale. The proposed method is validated by experiments with 197 subject videos, which demonstrates its accuracy and efficiency.**

***Index Terms*—Facial image analysis, facial paralysis measurement, local binary patterns (LBPs).**

I. Introduction

FACIAL paralysis is a common clinical entity, which is mainly caused by Bells palsy, trauma, tumors, injury during surgery, middle ear infection, and Ramsey–Hunt syndrome. A reliable assessment system would be an invaluable tool to audit the effectiveness of the treatment and the surgical techniques. House–Brackmann (H-B) grading system [1] is the most popular means for assessing facial paralysis. It 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. The drawbacks are that it relies on a subjective judgment with significant interobserver and intraobserver variation [2]–[4] and is insensitive to regional facial nerve function as it offers a single figure description of facial function.

Manuscript received April 19, 2008; revised September 29, 2008. First published March 27, 2009; current version published June 12, 2009. This work was supported in part by the Institute of Neurological Sciences, Southern General Hospital, Glasgow, U.K. *Asterisk indicates corresponding author.*

S. He is with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XQ, U.K. (e-mail: shu.he@eee.strath.ac.uk).

J. J. Soraghan is with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XQ, U.K. (e-mail: j.soraghan@eee.strath.ac.uk).

B. F. O'Reilly is with the Institute of Neurological Sciences, Southern General Hospital, Glasgow G51 4TF, U.K., and also with Gartnavel General Hospital, Glasgow G12 0YN, U.K. He is also with the University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: brian.o'reilly@northglasgow.scot.nhs.uk).

D. Xing is with the Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: dsxing@dcs.gla.ac.uk).

Digital Object Identifier 10.1109/TBME.2009.2017508

Many approaches have been reported for the objective measurement of facial paralysis. Some involve the use of markers on the face [4]–[6]. This makes the image processing simpler but there are negative implications. Neely *et al.* [8]–[10] and McGrenary *et al.* [7] measured facial paralysis by the differences between the frames of a video. Although their results correlate with the clinical H-B grade, the methods cannot cope with irregular or paradoxical motion in the weak side. Wachtman *et al.* [11], [12] measured facial paralysis by examining the facial asymmetry on static images. This method is sensitive to the extrinsic facial asymmetry caused by orientation, illumination, shadows, and the natural bilateral asymmetry. An optical-flow-based method for facial palsy grading [13] was proposed. The results were encouraging but the optical flow is computational intensive and is hard to estimate accurately for iconic changes.

To overcome the aforementioned difficulties, we present an objective facial paralysis grading system, which extracts the motion features based on multiresolution local binary pattern (MLBP). The paper is organized as follows. Section II outlines the overall design of our approach for facial paralysis grading. Section III presents the motion features extraction by LBP and its extension. Section IV describes the symmetry evaluation of facial movements by the resistor-average distance (RAD) and quantitative assessment of facial paralysis. Experiments results are presented in Section V. Finally, Section VI concludes the paper.

II. Overview of Facial Paralysis Grading System

Facial paralysis is measured by the asymmetry of the facial motion between two sides of the face in our work as simultaneous bilateral facial paralysis is a rare clinical entity [14]. In our work, all patients were videotaped with a front-view face and reasonable light conditions in such a way that each side of the face has roughly similar lighting. The patient is requested to perform five facial movements, which are raise eyebrows, close eyes gently, close eyes tightly, screw-up nose, and smile. The video sequence begins with the patient at rest, followed by the five movements, going back to rest between each movement.

A general overview of the proposed system is presented in Fig. 1. The raw video footage is fed into the system, whereby the first initial resting frame is analyzed as the reference frame. A highly efficient facial features localization method is employed in the reference frame. The face area is normalized by the interpupil distance and rotated so that the face midline is made perpendicular in the image. This affine transformation is employed in the subsequent frames. A face region map is assigned, as shown in Fig. 2. The region of interest (ROI) of the face, including facial features, is defined in the reference frame and

0018-9294/$25.00 © 2009 IEEE

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on March 26,2020 at 11:15:02 UTC from IEEE Xplore. Restrictions apply.

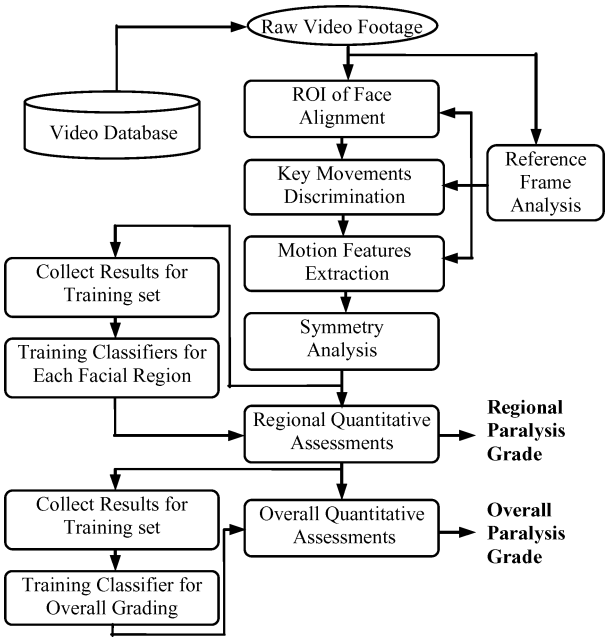

Fig. 1. Framework of the proposed facial paralysis grading system.

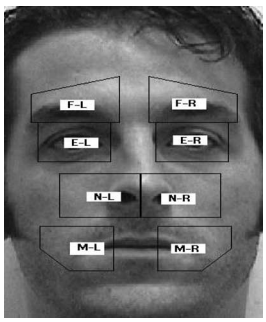

Fig. 2. Illustration of facial regions. F—forehead region; E—eye region; N—nasal region; M—mouth region; L—left; R—right.

is aligned in the subsequent frames to remove the rigid head motion by using block matching techniques, so that only the nonrigid facial expressions are kept in the image sequences for the further analysis. Image subtraction is then employed to identify the period of each key facial movement and to detect the five apex frames, on which the subject performs the maximum volume of facial motion. Further details on facial feature localization, stabilization, the discrimination of the key movements, and the detection of the apex frames can be found in [13]. This paper will focus on the motion features extraction and symmetry analysis.

The regional image sequence is cropped only in the frames corresponding to its relevant movements, e.g., the left and right forehead regions are cropped only in the frames presenting raising eyebrows. The facial regional image sequence is divided into several blocks from which LBP histograms on three orthogonal planes are computed and concatenated into three spatially and temporally enhanced features. The symmetry of facial movements is measured by the RAD between LBP histograms extracted from the two sides of the face. Five classifiers are trained for five facial movements to quantify the regional facial nerve function. The regional results are fed into another classifier to provide an overall quantitative evaluation of facial paralysis. Three classification methods, *k*-nearest-neighbor (k-NN), artificial neural network (ANN), and support vector machine (SVM), were employed as they have been used successfully for pattern recognition and classification on datasets with realistic size.

III. Motion Analysis With LBPs

LBP has proven to be highly discriminative descriptor for texture analysis. The most important property of the LBP operator in real-world applications is that it provides better tolerance against illumination changes than most of the other texture methods. Another equally important property is its computational simplicity, which makes it possible to analyze images in challenging real-time settings [15].

A. Local Binary Patterns

LBP operator was originally designed for texture analysis by Ojala *et al.* [16]. It can be seen as a unifying approach to structural and statistical texture analysis, defined as (1) and (2) below

$$\mathrm{LBP}_{P,R}(x,y) = \sum_{p=0}^{P} S\left(I\left(x + R\cos\frac{2\pi p}{P}, y + R\sin\frac{2\pi p}{P}\right) - I(x,y)\right) 2^p \tag{1}$$

where P circular neighbors $(x + R\cos(2\pi p/P), y + R\sin(2\pi p/P))$ with the radius R are first thresholded by the center pixel (x, y). A sign function S is then used to transform the differences in the neighborhood into a P-bit binary code. A binomial weight 2^p is assigned to convert the binary code into a unique LBP code, which depicts the local texture feature around the center, such as spots, flat area, and edges. Finally, the patterns of an image I can be approximately described with the distribution of LBP codes

$$H_i = \sum_{(x,y)\in I} \delta\left(\mathrm{LBP}_{P,R}(x,y), i\right) \tag{2}$$

where $i = 1, \ldots, n$, and n is the number of histogram bins, δ is the delta function where $\delta(j, i) = 1$, if $j = i$, and 0 otherwise.

A number of extensions to LBP are presented in [17]. A LBP is called a uniform pattern, $\mathrm{LBP}^{u2}_{P,R}$, if there are at most two one-to-zero or zero-to-one transitions in the circular binary code. All nonuniform patterns are labeled in a single bin in the histogram computation. Currently, most LBP formulations involve the use of "uniform" patterns, which have been experimentally observed to correlate well with real-world structures [17]. Our experiments prove that $\mathrm{LBP}^{u2}_{P,R}$ provides more discriminative motion features than the others.

B. Multiresolution LBPs

The most prominent limitation of the LBP operator is its small spatial support area. Each LBP code is calculated in a

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on March 26,2020 at 11:15:02 UTC from IEEE Xplore. Restrictions apply.

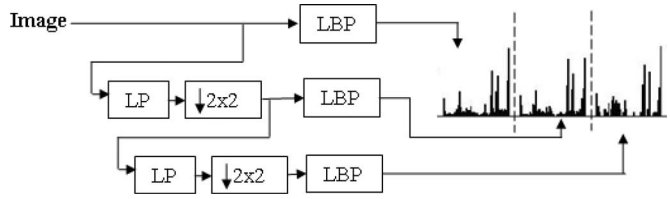

Fig. 3. Multiresolution LBPs.

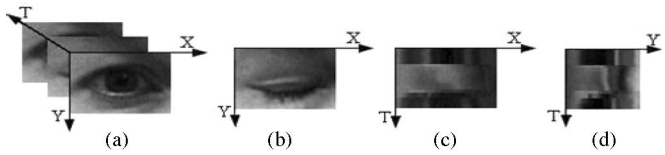

Fig. 4. (a) Image sequence for closing eye gently (89 images of 120 × 90 pixels). (b) *XY* in $t = 46$ (apex frame). (c) *XT* in $y = 45$ (middle row in the frame). (d) *YT* in $x = 60$ (middle column in the frame).

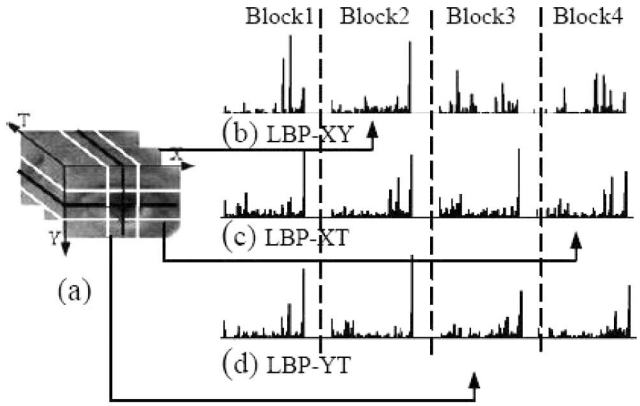

Fig. 5. Block-based LBP histogram on temporal–spatial domain.

small local neighborhood and characterizes the micropatterns. It is hard to capture large-scale structures that may be the dominant features of some textures. A straightforward way to address this issue is to explore the property in large spatial area by increasing the neighborhood radius and the number of neighbors of the LBP operators. Further, the LBP operators with varying spatial resolutions can be combined for multiscale analysis, for example, $\mathrm{LBP}^{u2}_{8,1}$, $\mathrm{LBP}^{u2}_{16,3}$, and $\mathrm{LBP}^{u2}_{24,5}$. While a large P produces a long histogram that is computationally expensive both in terms of computing speed and memory consumption. Another approach would be only increasing R and keeping P consistent. While the sparse sampling exploited by LBP operators with large neighborhood radius involves losing information of texture. It is noise sensitive and it decouples the statistics between scales.

To overcome these limitations, we incorporate a Gaussian pyramid into the LBP operator. As shown in Fig. 3, LBP operators with fixed R and P are employed on different scales of the image. $\mathrm{LBP}^{u2}_{8,2}$ is used in our work since it is a good tradeoff between recognition performance and feature vector length. It is first applied to the original image to extract the micropatterns of the image. The image is passed to a Gaussian low-pass filter and downsampled with a factor of 2, from which $\mathrm{LBP}^{u2}_{8,2}$ is calculated to extract more global patterns. These operations can be iterated to extract patterns over the different scales according to the size of the image. Finally, The LBP histograms from different resolutions are concatenated, which increases the algorithmic robustness, reduces noise effects, and keeps computation simple.

C. Block-Based LBP on Temporal–Spatial Domain

LBP has been verified to effectively represent the distribution of the texture patterns in the spatial domain. It also can be applied to an image sequence to extract the patterns of the pixel displacements in the temporal domain. LBP is extended to temporal–spatial domain for dynamic texture recognition over image sequences [18]. As shown in Fig. 4, the image sequence is considered as a stack of *XY* planes in axis *T*, a stack of *XT* planes in axis *Y*, and a stack of *YT* planes in axis *X*. LBP histogram on *XY* plane describes the appearance features. LBP histogram on *XT* and *YT* planes provide the motion information in the horizontal and vertical directions. These LBP histograms describe only the occurrences of the local patterns without any indication about their location when they are computed over the whole image sequence. For efficient representation, these LBP features should retain the spatial information and therefore a block-based approach is employed. The images are divided into several small blocks from which LBP histogram on three planes are computed and concatenated into three spatially enhanced features. Similarly, the images on *XT* and *YT* planes in each block can be divided into subblocks along the *T*-axis from which LBP histograms are extracted and concatenated into a temporally enhanced histogram. Finally, the temporally enhanced histogram in each spatial block is concatenated into a temporally and spatially enhanced feature. In this way, LBP-*XY* provides an appearance feature retaining spatial information. The motion features in specific location and time are characterized by LBP-*XT* and LBP-*YT*. To optimize the description of LBP, the number of neighbors P and the radius R of LBP operators should be set according to the image resolution, frame rate, and texture characters. For example, LBP-*XY* contains a lot of redundant information if a P greater than eight conjunct with $R = 1$ is used in a high-resolution image with flat texture. LBP-*XT* and LBP-*YT* with a large radius and $P = 8$ involve losing motion information when conducting on the low frame-rate video with the rigid motion.

In our application, the LBP features are extracted for each facial movement in its relevant regions. As an example, consider eye closing wherein the left eye region and the right eye region are cropped, respectively, from the frames presenting eye closing at rest. Fig. 5(a) shows the cropped image sequence for closing eye gently from the left eye. It is divided into 2×2 blocks. In each block, the LBP histogram on *XY*, *XT*, and *YT* planes are computed. LBP-*XY* characterizes the appearance feature. The contribution of the natural bilateral asymmetry to the spatial asymmetry of facial motion decreases as the intensity of facial motion increases, while the contribution of intrinsic facial asymmetry increases and reaches the maximum at the apex frame. The apex frame refers to the frame on that the subject

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on March 26,2020 at 11:15:02 UTC from IEEE Xplore. Restrictions apply.

performs the maximum volume of facial motion. For those subjects without significant natural bilateral asymmetry (e.g., big scar on one side of the face), the contribution of intrinsic facial asymmetry overwhelms that of the natural bilateral asymmetry at the apex frame. Therefore, the LBP histogram on *XY* plane is produced from the apex frame to measure the spatial asymmetry of facial motion caused by intrinsic facial asymmetry. Fig. 5(b) shows four concatenated histograms of $\mathrm{LBP}^{u2}_{8,2}$ on *XY* plane. Fig. 5(c) and (d) shows the concatenated histograms of $\mathrm{LBP}^{u2}_{8,3}$ on *XT* plane and on *YT* plane without temporal subblock.

In order to compare with left region, the cropped image sequences from the right side are flipped in the left–right direction before the LBP operator is conducted.

IV. Facial Paralysis Grading

A. Symmetry Measurement by RAD

A number of dissimilarity functions exist to compare distributions of variables. The performance of these dissimilarity measures strongly depends on the feature space. Kullback–Leibler (KL) divergence is a widely used information–theoretic approach for measuring the difference between two probability distributions. Let $p(x)$ and $q(x)$ represent two probability distributions of a discrete random variable *x*, the KL divergence is defined to be

$$D_{\mathrm{KL}}(p\|q) = \sum_x p(x)\log\left(\frac{p(x)}{q(x)}\right) \tag{3}$$

where $D_{\mathrm{KL}}(p\|q) = 0$, if $p(x)$ and $q(x)$ are identical distributions. $p(x)\log(p(x)/q(x))$ produces a significant contribution to $D_{\mathrm{KL}}(p\|q)$ when $p(x) \gg q(x)$ and $p(x)$ is large. $D_{KL}(p\|q)$ quantifies how well $p(x)$ is explained by $q(x)$.

Recent work shows that log-likelihood statistic, i.e., $-\sum_i p(x)\log(q(x))$, has been successfully used with LBP features for the classification problem [17], [18], [20]. For a given sample's distribution $p(x)$ and a set of distribution of the classes $\{q^i(x), i = 1, \ldots, C\}$, the entropy of $p(x)$ is constant and redundant for measuring the likelihood that the sample $p(x)$ is from a class $q^i(x)$. This is the reason that log-likelihood statistic is more commonly used for face recognition and expression recognition than KL divergence. In our work, we wish to quantify the distance between the two LBP histogram features extracted from two sides of the face, which is not a simple recognition problem. Either log-likelihood statistic or KL divergence is not a distance metric because they are, in general, asymmetric, e.g., $D_{\mathrm{KL}}(p\|q) \neq D_{\mathrm{KL}}(q\|p)$. In contrast to log-likelihood statistic, KL divergence has more properties of distance metric, e.g., $D_{\mathrm{KL}}(p\|q) = 0$, if $p(x)$ and $q(x)$ are identical distributions. Despite the KL divergence's computational and theoretical advantages, what becomes a nuisance in applications is its lack of symmetry. To address this problem, RAD [19] is given by

$$D_{RA}(p,q) = \left[D_{\mathrm{KL}}(p\|q)^{-1} + D_{\mathrm{KL}}(q\|p)^{-1}\right]^{-1}. \tag{4}$$

Here, $D_{RA}(p,q)$ increases as the dissimilarity between $p(x)$ and $q(x)$ increases. It keeps the interpretative attributes of KL divergence as it is directly computed from KL divergence. RAD between two distributions is approximately equal to the minimum divergence if the KL divergence in one direction is much greater than the other. RAD approximates the average of KL divergences in both directions, as the KL divergences in both directions approximately equal. It can be seen that RAD very much behaves like a smooth minimum of the KL divergences on both direction. In contrast to the KL divergence, RAD is symmetric and numerically stable. It accurately reflects the average error rate of an optimal classifier between two classes [19].

RAD is employed to measure the similarity between two sets of MLBP features extracted from two sides of the facial regions. Let $D_{RA} - XY, D_{RA} - XT,$ and $D_{RA} - YT$ denote the RAD of the LBP histograms on *XY* plane, *XT* plane, and *YT* plane, respectively. $D_{RA} - XY$ describes the spatial asymmetry of facial motion, and $D_{RA} - XT$ and $D_{RA} - YT$ indicate the temporal asymmetry of the facial motion in horizontal and vertical directions, respectively.

B. Quantitative Assessment

The paralysis in the each region is assessed using the regional H-B scale, which goes from grade I (normal) to grade VI (complete paralysis). To map the three symmetry attributes, i.e., $D_{RA} - XY, D_{RA} - XT,$ and $D_{RA} - YT$, into a regional H-B grade is a classification problem. ANN, k-NN, and SVM were considered as appropriate classifiers. They can be used successfully for pattern recognition and classification on datasets with realistic sizes. These three classification methods were employed in our work for the quantitative assessment of the regional paralysis and overall facial paralysis. Five classifiers are trained for the five movements respectively. Each one has three inputs, namely $D_{RA} - XY, D_{RA} - XT,$ and $D_{RA} - YT$. The regional facial function is classified to six grades from 1(normal) to 6 (total paralysis). These five regional H-B grades are then used as the inputs for another classifier to analyze the overall H-B grade.

V. Experiments

In our experiments, 197 videos taken from subjects with Bell's palsy, trauma to the nerve from skull fracture, surgical damage, etiologies as well as normal subjects. Each of them presents five facial movements in the 500–700 frames of 720 × 576 pixels. Their regional and overall H-B gradings were evaluated by a clinician.

A. Symmetry Measurements by LBP Operators

Two kinds of LBP operators, i.e., original LBP and MLBP, were used to extract the appearance and motion features. $\mathrm{LBP}^{u2}_{P,R}$,$\mathrm{LBP}^{ri}_{P,R}$, and $\mathrm{LBP}^{riu2}_{P,R}$ with $P = 4, 8, R = 1, 2, \ldots, 6,$ and different blocks dividing modes: 1 × 1, 1 × 2, …, 3 × 3, were used in spatial domain, and 1 × 1, 1 × 2, and 1 × 3 in temporal domain. Experiments show $\mathrm{LBP}^{u2}_{8,2} - XY, \mathrm{LBP}^{u2}_{8,5} - XT,$ and $\mathrm{LBP}^{u2}_{8,5} - YT$ with 2 × 2 blocks in spatial domain and 1 × 2 temporal domain provides higher correlation between the clinician assessments and D_{RA}. Details are shown in Table I.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on March 26,2020 at 11:15:02 UTC from IEEE Xplore. Restrictions apply.

TABLE I
CORRELATION BETWEEN THE SUBJECTIVE GRADING AND D_{RA} OBTAINED FROM $\mathrm{LBP}^{u2}_{8,2} - XY$, $\mathrm{LBP}^{u2}_{8,5} - XT$, AND $\mathrm{LBP}^{u2}_{8,5} - YT$ WITH 2×2 BLOCKS IN SPATIAL DOMAIN AND 1×2 BLOCKS IN TEMPORAL DOMAIN

	$D_{RA} - XY$	$D_{RA} - XT$	$D_{RA} - YT$
Forehead	0.9061	0.8198	0.7743
Eye gentle	0.6875	0.7475	0.8244
Eye tight	0.6041	0.6919	0.7389
Nose	0.6364	0.7803	0.8486
Mouth	0.8803	0.8016	0.7585

TABLE II
CORRELATION BETWEEN THE SUBJECTIVE GRADING AND D_{RA} OBTAINED FROM MULTIRESOLUTION $\mathrm{LBP}^{u2}_{8,2}$ WITH RESOLUTION LEVEL 3, 2×2 BLOCKS IN SPATIAL DOMAIN AND 1×2 IN TEMPORAL DOMAIN

	$D_{RA} - XY$	$D_{RA} - XT$	$D_{RA} - YT$
Forehead	0.9107	0.8214	0.7760
Eye gentle	0.7191	0.7502	0.8339
Eye tight	0.6535	0.7066	0.7495
Nose	0.6830	0.7879	0.8602
Mouth	0.8944	0.8197	0.7631

Multiresolution $\mathrm{LBP}^{u2}_{P,R}$ with resolution level $=$ 2 or 3, $P = 8, R = 1, 2$, and different blocks dividing modes: 1×1, $1 \times 2, \ldots, 3 \times 3$, were used in spatial domain, and 1×1, 1×2, and 1×3 in temporal domain. Experiments show multiresolution $\mathrm{LBP}^{u2}_{8,2}$ with resolution level 3, 2×2 blocks in spatial domain, and 1×2 in temporal domain provides higher correlation between the clinician assessments and D_{RA}. As shown in Table II, D_{RA} on the MLBP is more correlated with the subjective grading than D_{RA} on the original LBP in the Table I. It indicates that the proposed MLBP provides more discriminative power than the original LBP.

B. Quantitative Assessment of Regional and Overall Paralysis

SVM, k-NN, and radial basis function neural network (RBFNN) were applied as classifiers to quantitatively analyze the paralysis of 197 subjects. In Dr. Brian O'Reilly's previous research, 36 videos from 23 patients were assessed by two groups of experts, who were all ear, nose, and throat (ENT) consultants. The overall H-B agreement percentage between two groups was only 47.2%. In the studies of Kanerva [21], 28 clinicians grades 8 facial palsy patients. The overall agreement percentage between doctors was 48%. The disagreement within one grade between doctors (no difference, or one grade difference between their assessments) was 89.29%. Many similar studies of subjective assessment have shown that the best that clinical assessment alone can achieve is usually an interobserver or intraobserver variation of at least one grade. In our experiments, the outputs of the classifiers with the disagreement no more than one grade of the clinician assessment are accepted; otherwise, the classifiers provide the wrong grading.

As the dataset was not large, the leave-*k*-out cross-validation test scheme instead of *k*-fold was adopted. The classifiers are evaluated by the average performance for 20 repetitions of leave-*k*-out cross validation, with $k = 20$. The pairwise SVMs with linear, polynomial and RBF kernels, followed by vote schemes were tested and experiments demonstrate the pairwise SVM with RBF kernels achieves best performance. Fig. 6 provides a graphic comparison of the average performance of SVM, RBFNN, and k-NN based on the MLBP approach. The accuracy of grading by pairwise SVM with RBF kernels is at least 3% higher than the RBFNN and the k-NN. Experiments show that SVM also provides more stable results.

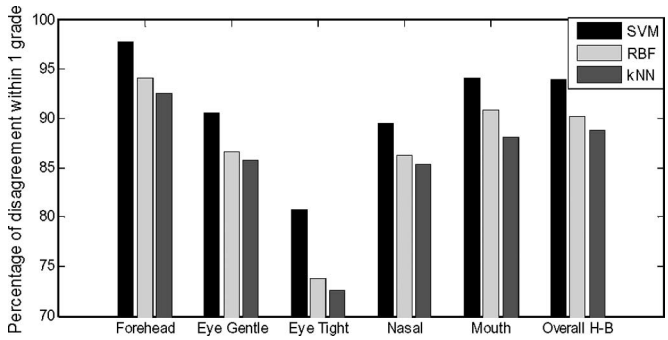

Fig. 6. Comparison of the performance of SVM, RBFNN, and k-NN.

TABLE III
AVERAGE PERFORMANCE OF SVM ON THE ORIGINAL LBP FEATURES

Disagreement	0	1	2	3	4	5	<=1
Forehead	67.3	30.1	2.6	0	0	0	**97.4**
Eye gentle	45.9	40.2	8.5	5.4	0	0	**86.1**
Eye tight	34.6	39.3	17.2	6.4	2.5	0	**73.9**
Nose	56.1	27.5	9.3	7.1	0	0	**83.6**
Mouth	60.3	31.8	7.9	0	0	0	**92.1**
H-B	60.7	30.6	8.7	0	0	0	**91.3**

Tables III and IV present the average classification performance of the pairwise SVM with RBF kernels, in percentages, for the 20 repetitions of the leave-*k*-out cross validation, with $k = 20$. Table III presents the classification results of the original LBP approach and Table IV presents the results of the MLBP approach. The numbers in the first columns give the percentages, which are the same as the clinician assessments. Columns 2–6 show the percentages where the disagreement between the outputs of classifiers and the clinician assessments is from 1 to 5 grades, respectively. The last columns show that the percentage of the disagreement within one grade. The results demonstrate that the better agreement is in the forehead region and the mouth region. The grading for eye closing tightly has the worst agreement. The approach based on the MLBP provides the better agreement than the original LBP, especially in the eye region.

A comparison of the performance of the proposed method based on original LBP, MLBP, and the optical-flow-based method [13] is provided in Table V. The MLBP-based approach outperforms the original LBP-based approach, especially in the regions of eyes and nose with the improvement of 4.5%–6.8%. The optical-flow-based method provides the better agreement than both of the LBP-based approaches in the eye region. The MLBP approach provides about 3.2%–6.1% higher accuracy in

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on March 26,2020 at 11:15:02 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
AVERAGE PERFORMANCE OF SVM ON THE ORIGINAL MLBP FEATURES

Disagreement	0	1	2	3	4	5	<=1
Forehead	71.3	26.6	1.0	1.2	0	0	**97.8**
Eye gentle	52.6	38.0	8.2	1.2	0	0	**90.6**
Eye tight	41.1	39.6	12.8	4.3	2.2	0	**80.7**
Nose	62.7	26.8	7.4	3.1	0	0	**89.5**
Mouth	63.4	30.7	5.9	0	0	0	**94.1**
H-B	69.3	24.6	6.1	0	0	0	**93.9**

TABLE V
COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD BASED ON ORIGINAL LBP, MLBP, AND THE OPTICAL-FLOW-BASED METHOD

	LBP	MLBP	Optical Flow
Forehead	97.4%	97.8%	93.2%
Eye gentle	86.1%	90.6%	92.2%
Eye tight	73.9%	80.7%	93.1%
Nose	83.6%	89.5%	86.3%
Mouth	92.1%	94.1%	88.0%
Overall HB	91.3%	93.9%	94.1%

the regions of forehead, nose and mouth than the optical-flow-based method. Both methods achieve an accuracy of around 94% for the overall H-B grading.

There is no public database of facial paralysis (either video, or image). The existing research reported the results using their own clinical data. Most of them tested their methods on less then 30 patients. For a small size of dataset, the results always lack the statistical stability since a small number of samples cannot represent the substantial sampling variability. The proposed method is tested on a dataset with 197 videos taken from variant palsy cases and achieves an accuracy of around 94% for the overall H-B grading. Neely *et al.* [8]–[10] and McGrenary *et al.* [7] used a subtraction technique that measures the differences of pixels intensity between the frames of a video. The features extracted by the methods [7]–[10] were input to the classifiers for quantitative assessment in our dataset and achieves 86.2% of the overall H-B disagreement within one grade. Wachtman *et al.* [11], [12] studied the palsy based on static images by measuring the intensity difference and edge difference between the two sides of the face. The features extracted by this method were used for quantitative assessment and achieves an accuracy of 65.3% for the overall H-B grading.

C. Discussion

The eye region is full of wrinkles, especially closing tightly. These small ridge or crease on the eyelids vary irregularly over the whole movement, sometimes asymmetrically even for a normal subject. LBP features are not robust to model these subtle motions. However, MLBP-based approach provides the better discrimination of moderate weakness and severe palsy in the eye region. It reveals the structural changes of the appearance, i.e., the significant difference between the normal side and the severe palsy side for eye closing. It can model the iconic changes in the eye region, which optical flow usually fails to track. Closer examination of the performance in the eye region, as shown in Table VI, reveals interesting statistics in terms of the specific abilities of the two methods. The paralysis in the eye region is assessed using the regional H-B scale, which goes from grade I (normal) to grade VI (complete paralysis). Column A* shows the distribution of the clinician assessment in the dataset. For example, 51 subjects are normal during closing eye gently. Column B* and Column C* give the number of the correct classification, i.e., the disagreement within one grade, by the optical-flow-based method [13] and the MLBP-based approach. It demonstrates that the optical-flow-based method provides the better assessment of normal to moderate weakness in the eye region. The MLBP-based approach provides the better discrimination of severe palsy in the eye region.

TABLE VI
DISTRIBUTION OF PARALYTIC STATES AND CLASSIFICATION RESULTS IN THE EYE REGION

Regional H-B	Eye Gentle			Eye Tight		
	A*	B*	C*	A*	B*	C*
1	51	51	48	56	56	50
2	59	57	55	47	44	35
3	31	29	24	47	45	37
4	43	36	38	34	29	25
5	9	6	8	11	8	10
6	4	2	4	2	0	2

The most encouraging aspect of these results is that the MLBP approach outperforms the optical-flow-based method in the regions of forehead, nose, and mouth with the improvement of 3.2%–6.1%. Memory consumption and computational burden in the MLBP-based approach are the clear advantages over the optical-flow-based method. Once the classifiers are trained, the difference of the computation time mainly depends on the computational complexity of the motion feature extraction, i.e., MLBP calculation, optical flow estimation. The motion features using MLBP for the five facial movements were computed in 0.0625 s on a 1.73 GHz laptop, while the optical flow motion features for the five facial movements implemented in [13] were extracted in 62.75 s on the same computer. The MLBP-based approach is more suitable for real-time operation.

VI. CONCLUSION

In this paper, we present a novel approach to objectively and quantitatively measure facial paralysis in videos. MLBP are employed on temporal–spatial domain to extract the motion features. The symmetry of facial motion is measured by the RAD between MLBP features. SVM provides the quantitative assessment of the facial paralysis. The limitation of the proposed method is that it is sensitive to out-plane facial movements, with significant natural bilateral asymmetry. However, MLBP-based method has many merits that are valuable for our application.

1) MLBP features describe the structural changes of appearance. They reveal the significant difference between the normal side and the severe palsy side for eye closing, and

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on March 26,2020 at 11:15:02 UTC from IEEE Xplore. Restrictions apply.

can model the iconic changes in the eye region, which optical flow usually fails to track.

2) MLBP is insensitive to the illumination changes since LBP is independent of monotonic transformation of gray scale.
3) Its computational simplicity makes MLBP-based approach more suitable for real-time operation.

REFERENCES

[1] J. W. House, "Facial nerve grading systems," *Laryngoscope*, vol. 93, pp. 1056–1069, 1983.

[2] C. H. Beurskens and P. G. Heymans, "Positive effects of mime therapy on sequelae of facial paralysis: Stiffness, lip mobility, and social and physical aspects of facial disability," *Otol. Neurotol.*, vol. 24, no. 4, pp. 677–681, Jul. 2003.

[3] J. B. Kahn and R. E. Gliklich, "Validation of a patient-graded instrument for facial nerve paralysis: The FaCE scale," *Laryngoscope*, vol. 111, no. 3, pp. 387–398, Mar. 2001.

[4] C. Linstrom, "Objective facial motion analysis in patients with facial nerve dysfunction," *Laryngoscope*, vol. 112, no. 7, pp. 1129–1147, Jul. 2002.

[5] H. Scriba, S. J. Stoeckli, and D. Veraguth, "Objective evaluation of normal facial function," *Ann. Otol. Rhinol. Laryngol.*, vol. 108, pp. 641–644, 1999.

[6] P. Dulguerov, F. Marchal, and D. Wang, "Review of objective topographic facial nerve evaluation methods," *Amer. J. Otol.*, vol. 20, no. 5, pp. 672–678, 1999.

[7] S. McGrenary, B. F. O'Reilly, and J. J. Soraghan, "Objective grading of facial paralysis using artificial intelligence analysis of video data," in *Proc. 18th IEEE Symp. Comput.-Based Med. Syst.*, 2005, pp. 587–592.

[8] J. G. Neely, A. H. Joaquin, L. A. Kohn, and J. Y. Cheung, "Quantitative assessment of the variation within grades of facial paralysis," *Laryngoscope*, vol. 106, pp. 438–442, 1996.

[9] T. D. Helling and J. G. Neely, "Validation of objective measures for facial. Paralysis," *Laryngoscope*, vol. 107, no. 10, pp. 1345–1349, 1997.

[10] J. G. Neely, "Advancement in the evaluation of facial function," *Adv. Otolaryngol. Head Neck Surg.*, vol. 15, pp. 109–134, Jan. 2002.

[11] 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," presented at the Combined Annu. Conf. Robert H. Ivy Ohio Valley Soc. Plastic Reconstr. Surg., Pittsburgh, PA, June, 2002.

[12] Y. Liu, K. Schmidt, J. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," *Comput. Vis. Image Understanding J.*, vol. 91, pp. 138–159, Jul. 2003.

[13] S. He, J. J. Soraghan, and B. F. O'Reilly, "Biomedical image sequence analysis with application to automatic quantitative assessment of facial paralysis," *EURASIP J. Image Video Process.*, vol. 2007, no. 4, pp. 81282-1–81282-11, 2007.

[14] N. Stahl and T. Ferit, "Recurrent bilateral peripheral facial palsy," *J. Laryngol. Otol.*, vol. 103, pp. 117–119, 1989.

[15] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.

[16] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognit.*, vol. 29, pp. 51–59, 1996.

[17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray scale and rotation invariant texture analysis with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[18] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[19] D. H. Johnson and S. Sinanovic, "Symmetrizing the Kullback–Leibler distance," Tech. Rep., Rice Univ., Houston, TX, 2001.

[20] T. Mäenpää and M. Pietikäinen, "Multi-scale binary patterns for texture analysis," *Proc. SCIA 2003* (Lecture Notes in Computer Science 2749). Berlin, Germany: Springer-Verlag, 2003, pp. 885–892.

[21] M. Kanerva, "Peripheral facial palsy grading, etiology, and Melkersson–Rosenthal syndrome," Ph.D. thesis, Univ. Helsinki, Helsinki, Finland, 2008.

Shu He (S'07) received B.Eng. degree from Xidian University, Xi'an, China, in 1996, and M.Eng. degree from Chalmers University of Technology, Göteborg, Sweden, in 2005. She is currently working toward the Ph.D. degree at the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K.

She was a Software Engineer with the Institute of Telecommunication Technique, Ministry of Information Industry, and a leading Telecom Equipment Vendor in China. Her current research interests include computer vision, pattern recognition, statistical data analysis, and artificial intelligence.

John J. Soraghan (M'83–S'84–SM'96) received the Ph.D. degree in electronic engineering from the University of Southampton, Southampton, U.K., in 1989.

In 1986, he joined the University of Strathclyde as a Lecturer and became a Professor in 2003. From 2005 to 2007, he was the Head of the Institute for Communications and Signal Processing. He is currently the Texas Instruments Chair at the University of Strathclyde, Glasgow, U.K. His current research interests include advanced linear and nonlinear signal processing theory and algorithms with applications to telecommunications, biomedical, multimedia systems, remote sensing, and defense. He has supervised 28 Ph.D. students to graduation, holds three patents, and has authored or coauthored over 230 technical papers.

Brian F. O'Reilly received the M.B.Ch.B. degree in medicine from the University of Glasgow, Glasgow, U.K., in 1972, and the F.R.C.S. degree in surgery from Royal College of Physicians and Surgeons, Glasgow, Scotland, U.K., in 1976.

He is currently a Consultant ENT Surgeon and Otologist at Gartnavel General Hospital, Glasgow, U.K., and a Consultant Neuro-Otologist at the Institute of Neurological Sciences, Southern General Hospital. He is also an Honorary Senior Lecturer at University of Glasgow, Glasgow. His current research interests include acoustic neuroma growth and facial palsy.

Mr. O'Reilly is the member of the British Association of Otolaryngologists, Scottish Otolaryngology Society, and the British Skull Base Society, and the President of the Scottish Otolaryngological Society.

Dongshan Xing received the Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China.

He is currently a Postdoctoral Research Assistant in the Department of Computing Science, University of Glasgow, Glasgow, U.K. His current research interests include pattern recognition, data mining, and statistical data analysis.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on March 26,2020 at 11:15:02 UTC from IEEE Xplore. Restrictions apply.