Non-intrusive Head Movement Analysis of Videotaped Seizures of Epileptic Origin

Bappaditya Mandal[1], How-Lung Eng[1], Haiping Lu[1], Derrick W. S. Chan[2] and Yen-Ling Ng[2]

Abstract— In this work we propose a non-intrusive video analytic system for patient's body parts movement analysis in Epilepsy Monitoring Unit. The system utilizes skin color modeling, head/face pose template matching and face detection to analyze and quantify the head movements. Epileptic patients' heads are analyzed holistically to infer seizure and normal random movements. The patient does not require to wear any special clothing, markers or sensors, hence it is totally non-intrusive. The user initializes the person-specific skin color and selects few face/head poses in the initial few frames. The system then tracks the head/face and extracts spatio-temporal features. Support vector machines are then used on these features to classify seizure-like movements from normal random movements. Experiments are performed on numerous long hour video sequences captured in an Epilepsy Monitoring Unit at a local hospital. The results demonstrate the feasibility of the proposed system in pediatric epilepsy monitoring and seizure detection.

I. INTRODUCTION

Epileptic seizures involve uncoordinated movement in a patient's body. It represents one of the most frequent clinical signs of central nervous system dysfunction in the newborn and has been associated with significant long term adverse sequelae such as neurologic impairment, developmental delay, postnatal epilepsy, etc [1]. The analysis of these uncoordinated (or involuntary) movements is an important clinical factor in seizure identification and subsequent diagnosis followup and treatment. For many years researchers have concentrated on developing quantification algorithms to extract electroencephalogram (EEG) features which are invisible to traditional visual inspection. Very little efforts have been made to quantify body parts movement in video recorded seizures [1], [2]. The recent advances in digital video-EEG technology based on computer multimedia and large storage devices have opened up many such possibilities [3], [4].

The quantification of human motion is a very challenging and complex problem as the human body is not a rigid object and may present a multitude of shapes, postures and self occlusions. In addition, the illumination/subject variation, noise induced by devices and foreign occlusions complicate the movement analysis process [5]. To address this complex problem, some researchers used marker-based methods to extract features from human body motion for clinical diagnosis [6]. Motion of these markers reflect the

movement of underlying body parts and their dynamics. However in practice, it is very difficult to keep them attached to their bodies, especially for children and infants. Miyazaki [7] developed a system to quantify the eyelid movement for clinical purposes. They attached markers to the lower margin of the upper eyelid of a subject. The subject was requested to blink and the marker movements were recorded and evaluated. In [8], the subject's hands are required to be placed steadily on the desk so that the relating positions of the markers are easily identified according to their horizontal positions to enable the motion correspondence. This approach will fail in complex movements such as rotation, crossing, etc.

To avoid the constrains induced by markers, researchers in [3], [9] used an image optical flow analysis approach to quantitatively evaluate body motion and visualize the apparent velocity field of the entire body motion, including both chest respiratory movement and posture change. Some model-based approaches are used for tracking and recognizing daily human motion such as running, waving hands, etc [10], [11]. However, these approaches can handle only few simple motions. It is difficult to use them in our target application as the seizure movement is very complex, subtle, uncoordinated and hence, very difficult to model and quantify [4].

Our present goal is to detect seizures inside the Epilepsy Monitoring Unit (EMU) for a continuous monitoring of suspected Epilepsy patients. To detect seizures we propose to detect patient's head/face using person-specific skin color model, template matching on various face/head poses and face detection. Spatio-temporal features are extracted from the tracked head movement. Finally, support vector machines (SVM) are used on these features to classify the seizures and random normal movement. The system is developed to work continuously in real time.

II. SYSTEM FRAMEWORK

Normally patient at EMU stays for 3-5 days for monitoring. During these days our cameras constantly watch this patient. Fig. 1 shows our proposed system framework. It performs three tasks, (i) segments the patient's skin portion, (ii) matches the stored templates of the patient's face/head and (iii) it also tries to find faces in images of the incoming video streams. Template matching gets the first priority and if successful, proceeds to head tracking. Secondly it tries to find the face in the segmented skin portions and lastly, if both fails then it tries to find the face in the images. If all three modules fail then the system locates the face same

[1]B. Mandal, [1]H-L Eng and [1]H. Lu are with the Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632, {bmandal, hleng, hlu} at i2r.a-star.edu.sg

[2]D. Chan and [2]Y-L Ng are with the KK Women and Children Hospital, Singapore, {Derrick.Chan.WS, Ng.Yen.Ling} at kkh.com.sg

as that in the previous frame. The head tracking module keeps the history of the tracked head obtained either from the preceding three modules. After tracking the head, the system extracts the spatio-temporal features. These features takes into account both the spacial coordinates of the tracked head and also the distance moved along with the time taken to make such a move. These features are then passed through SVM to perform the movement analysis and classify the video stream (of definite period) to be as seizures or normal random movements. We have trailed our system at a local hospital. In the following sections we shall explain each of these modules in detail.
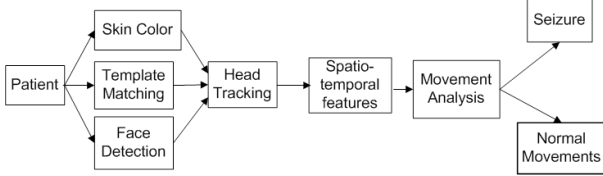

Fig. 1. System framework.

A. Skin Segmentation

Our skin is reflective in nature. In RGB color space representation, the skin color cluster is very large and scattered and hence, very difficult to model. To overcome this problem Hsu *et al.* [12] proposed skin cluster modeling in YUV (YC_bC_r) color space. They noticed that the skin color cluster can be determined by two central curves (the centers) $\bar{C}_b(Y)$ and $\bar{C}_r(Y)$ and by deviation curves (spread of the cluster) [12]. All these are dependent on the Y component, which represents the luminance. They proposed the central curves based on a large number of fixed parameters that are obtained from very large training databases. However, these parameters are not suitable for patients with large variations in their skin colors, hence, not suitable for our purpose.

It is evident from [13] that the normalized RGB color space representation (NRGB) is very popular among the researchers. It can be obtained from the RGB values by normalization procedure. Each color pixel $c(x, y)$ at location (x, y) in NRGB space is represented by the corresponding color vector $v_c = (r, g, base)^T$, where

$$r = \frac{R}{base}, g = \frac{G}{base}, b = \frac{B}{base}, base = R + G + B. \quad (1)$$

For matte surfaces, NRGB is invariant (under certain assumptions, like ambient light) to changes of surface orientation relatively to the light source [13]. Under this representation the skin color forms a compact cluster whose position, size and shape vary from person to person and context specific side conditions. For our application, we propose to define an ellipsoid within person-specific NRGB space. Let μ be the mean and S be the covariance matrix of all the color vectors of all skin color pixels within the image. If P is the total number of skin color pixels within an image then

$$\mu = \frac{1}{P} \sum_{i=1}^{P} v_i = \begin{bmatrix} \mu_r \\ \mu_g \\ \mu_{base} \end{bmatrix}, \quad (2)$$

and

$$\mathbf{S} = \frac{1}{P-1} \sum_{i=1}^{P} (v_i - \mu)(v_i - \mu)^T. \quad (3)$$

Our system requires an user initialization of the patient's skin only in the first frame. The system then uses these selected person-specific skin color pixels to compute the matrices μ and $\mathbf{S}$ using (2) and (3) respectively. Covariance matrix $\mathbf{S}$ computes the variations of the skin color pixels among r, g and $base$ planes. It can be seen as

$$\mathbf{S} = \begin{bmatrix} \sigma_r & \rho_{r,g} & \rho_{r,base} \\ \rho_{g,r} & \sigma_g & \rho_{g,base} \\ \rho_{base,r} & \rho_{base,g} & \sigma_{base} \end{bmatrix}, \quad (4)$$

where σ represents the variations among the skin pixels in the same plane and ρ represents the variations among the skin pixels in the different planes. We use this variance information to form an ellipsoidal cluster in NRGB space. We use Mahalanobis distance measure to define the ellipsoidal bases, which is given by

$$D_c = (v_c - \mu)^T S^{-1} (v_c - \mu). \quad (5)$$

All distance measures (D_c) less than a threshold D_t are treated as skin pixels. Different values of distance measures D_t will lead to different number of classified skin pixels. To avoid exhaustive search for the best parameter value, D_t is fixed to be 1 in all our experiments of this paper and for smooth running of our system. A sample of the skin segmentation of the Patient's body is shown in Fig. 2 (patients' faces are masked in all images of this paper). The skin segmented regions then undergo a series of morphological operations to fill in the holes and filter by larger blobs areas.
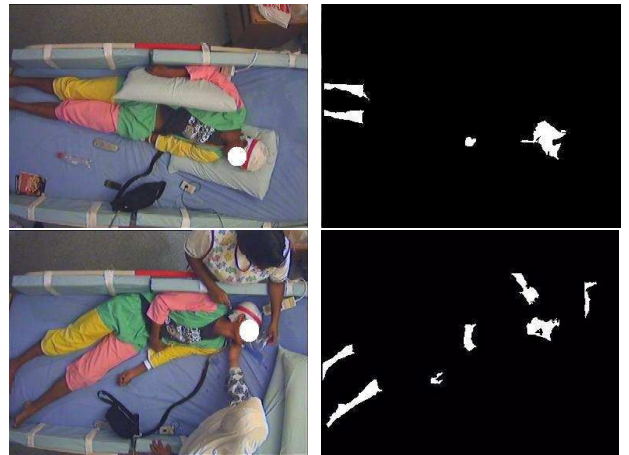

Fig. 2. Top at start and bottom, after 2 hours. Sample original images and their skin segmentations.

B. Template Matching

In the beginning of the patient monitoring, our system involves a registration process, where the nurse or care giver captures 7 face/head poses (frontal, left profile, right profile, left side profile, right side profile, look up profile and look down profile) of the patient. Fig. 3 shows the sample pose

templates used in our experiments and test bedding. The patient inside EMU wears white cap for capturing the EEG data (as shown in Fig. 3). Using these registered templates and modified version of the OpenCV library [14] head/face detection is performed. Both the input and template are down sampled to reduce the computation time. If the matching score is above the confidence score (in our case we set this to 70%) the system takes this as head/face. Some sample detected face/head images are shown in Fig. 4.

Fig. 3. Sample patient's face/head templates.

C. Face Detection

We use OpenCV version for face detection (FD) originally developed by Paul Viola and improved by Rainer Lienhart [14]. It uses Haar-like features that encode the existence of oriented contrasts between regions in the image. A set of these features is used to encode the contrasts exhibited by a human face and their spatial relationships and it works for upright images. Since our application is focused on patients lying in the bed and sometimes sitting up we rotate the image in a 90 degree counter-clockwise direction to make it upright. Some sample images are shown in Fig. 4.

Fig. 4. Detected head/face using face/head templates, skin segmentation and face detection.

Our proposed system utilizes the face template matching and uses this as the first preference in finding the head/face. If the confidence level of this module falls below the threshold, the system then uses the history of the skin segmentation and associated skin blobs to find the head/face. If both these modules fail the system then uses the face detection (FD) module. In our patient monitoring application, this FD module often fails because most of the times patient's face is not clearly visible due to self occlusion, articulation and EEG cap. Hence, the face detection module gets the least priority in the proposed system.

D. Head/face Model Fitting

At first our system performs the template matching, if it finds the face/head it proceeds to tracking the head and records its history. On unsuccessful finding of head/face in the template matching process, the system computes an overlapping probability density value of the skin regions (in form of blobs) in the current frame with that of the blobs in the previously frame. This is achieved by maximizing the overlap between pixels belonging to the rectangular shape

and detected pixels of the skin regions. The overlapping function is given by:

$$\Phi_i = \frac{|A_i \bigcap A_{Pixel_{Head}}|}{\sqrt{|A_i||A_{Pixel_{Head}}|}}, \qquad (6)$$

where A_i denotes the pixel area of the ith blob in the skin region and $A_{Pixel_{Head}}$ denotes the pixel area of the rectangle. The blob with the highest probability or most overlapped is taken as the head/face. Fig. 5 shows the fitted model of the patient's head/face at EMU. If the overlap values Φ_i, $i \in$ {all skin blobs} is below some threshold then the system proceeds to face detection. If the face detection also fails then the system keeps the previously found head/face locations.

Fig. 5. Left at start, Right after 2 hours. Sample patient's detected face/head. The box shows the detected face/head with various poses and confidence levels (%).

III. FEATURE EXTRACTION AND MOVEMENT ANALYSIS

A. Spatio-temporal Features

In general, epileptic patients often move their head with very rapid forces just at the seizure onset and also during the course of the seizure (as shown in Fig. 4). On successful finding the head/face, the system computes the spatio-temporal features using the current and previous frames. Our proposed system captures how much the head moved in one second or the velocity of the head movements. This movement uses the displacement of the centroid of the tracked region between the current and previous frames. The velocity of head movement is given by

$$\vartheta_k = \eta * \sqrt{(x_{k-1} - x_k)^2 + (y_{k-1} - y_k)^2}, \qquad (7)$$

where k is the current frame and η is the frames per second (fps).

B. Movement Analysis

The features which encode the rapid, jerking, subtle and normal movements are extracted from the head movements. One entire seizure event is captured and SVM model is trained with 500 frames (≈ 42 seconds) interval along with overlap of 200 frames. This trained model are either person specific or generic. Currently, we are exploring with person-specific seizure modeling. All incoming video streams are passed through the above steps (as shown in Fig. 1) and classified as seizure or normal. The seizure detection would have a latency period (≈ 25 seconds).

IV. EXPERIMENTS AND PRELIMINARY TRIALS

Video data of various epileptic and suspected epileptic patients are collected in the EMU at the KK Women and Children Hospital, Singapore. The video sequences are captured at 12 frames per second and the resolution is 384×288. All head movements are quantified and analyzed. Similar to EEG tracings, we follow [6] to display the quantified movement information as movement trajectories for the clinicians easy interpretation. In this paper we present results for head movements and seizures detected in the continuous monitoring of a patient, where two seizure events are automatically detected by our system for demonstration.

Figs. 6 and 7 show the head/face movements. Magnitude of the head movements are plotted in the y-axis and time in the x-axis. All the movements are presented in blue dots and detected seizures are highlighted in red circles. From the plots it is evident that during seizure events the patient head moved very rapidly and repeatedly. In other normal random movements, although they are of similar magnitude but are not any kind of sustained movement.
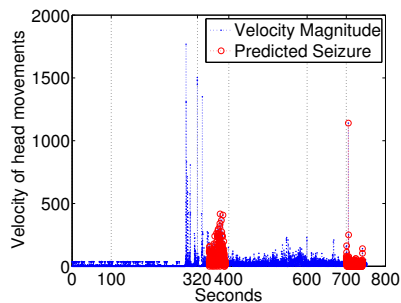

Fig. 6. Seizure prediction is shown in red circles and velocity magnitude is shown in blue dots. Actual seizure is between 320th second to 420th second.

In Fig. 6, the actual seizure event occurred between 320th second to 420th second (duration 100 seconds), and our system detected two events: 350th second to 400th second (duration 50 seconds) and 700th second to 750th second (duration 50 seconds). In Fig. 7, the actual seizure event occurred between 330th second to 420th second (duration 90 seconds), and our system detected one event: 370th second to 420th second (duration 50 seconds).
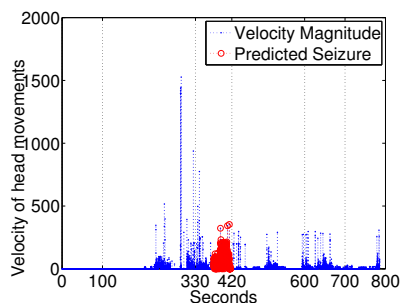

Fig. 7. Seizure prediction is shown in red circles and velocity magnitude is shown in blue dots. Actual seizure is between 330th second to 420th second.

V. CONCLUSIONS

This paper presents a non-intrusive video analytic system for patient's body parts movement analysis in EMU. The system utilizes skin color modeling, head/face pose template matching and face detection to track and quantify head movements. Spatio-temporal features are extracted from the head movements and SVM is used to classify the seizure and normal random movements. Experiments are performed on numerous long video sequences captured in an EMU at a local hospital. The experimental studies show the feasibility of the proposed system in pediatric epilepsy monitoring and seizure detection. Furthermore, the proposed system is promising in a wider application field of fine semiology and other head-movement behavioral change investigations.

ACKNOWLEDGMENT

REFERENCES

[1] N. B. Karayiannis, B. Varughese, G. Tao, J. Frost, J. M. Wise, and E. Mizrahi, "Quantifying Motion in Video Recordings of Neonatal Seizures by Robust Motion Trackers Based on Block Motion Models," *IEEE Trans. Biomed. Engg.*, vol. 56, no. 2, pp. 1065–1077, June 2005.

[2] N. B. Karayiannis, S. Srinivasan, R. Bhattacharya, J. M. Wise, J. Frost, and E. Mizrahi, "Extraction of Motion Strength and Motor ActivitySignals From Video Recordings of Neonatal Seizures," *IEEE Trans. Medical Imaging*, vol. 20, no. 9, pp. 965–980, Sept 2001.

[3] N. B. Karayiannis, Y. Xiong, G. Tao, J. Frost, J. M. Wise, R. Hrachovy, and E. Mizrahi, "Automated Detection of Videotaped Neonatal Seizures of Epileptic Origin," *Epilepsia*, vol. 47, no. 6, pp. 966–980, 2006.

[4] H. Lu, H. Eng, B. Mandal, D. Chan, and Y. Ng, "Markerless Video Analysis for Movement Quantification in Pediatric Epilepsy Monitoring," in 33rd *IEEE Engg. in Medicine and Biology Society (EMBC'11)*, Boston, MA, USA, Dec 2011, pp. 8275–8278.

[5] B. Mandal and H.-L. Eng, "Regularized Discriminant Analysis for Holistic Human Activity Recognition," *IEEE Intelligent Systems*, vol. 27, no. 1, pp. 21–31, Jan-Feb 2012.

[6] Z. Li, A. Silva, and J. Cunha, "Movement Quantification in Epileptic Seizures: A New Approach to Video-EEG Analysis," *IEEE Trans. Biomedical Engineering*, vol. 49, no. 6, pp. 565–573, June 2002.

[7] S. Miyazaki, A. Ishida, and A. Komatsuzaki, "A clinically oriented video-based system for quantification of eyelid movement," *IEEE Yrans. Biomed Eng.*, vol. 47, pp. 1088–1096, Aug 2000.

[8] A. Jobbagy, E. H. Furnee, P. Harcos, M. Tarczy, I. Krekule, and L. Komjathi, "Analysis of movement patterns aids the early detection of Parkinsons disease," vol. 2, Chicago, IL, Nov 1997, pp. 1760–1763.

[9] T. M. K. Nakajima, A. Osa and H. Miike, "Evaluation of body motion by optical flow analysis," *Jpn. J. Appl. Physics*, vol. 36, no. 5A, pp. 2929–2937, 1997.

[10] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780–785, July 1997.

[11] B. Mandal and H.-L. Eng, "3-parameter Based Eigenfeature Regularization for Human Activity Recognition," in *IEEE 35th International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010)*, Dallas, Texas, USA, Mar 2010, pp. 954–957.

[12] R. Hsu, M. Mottaleb, and A. Jain, "Face Detection in Color Images," *IEEE Trans. on Pattern Analysis and Machine Intelligent*, vol. 24, no. 5, pp. 696–706, 2002.

[13] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques," in *In Proc. GRAPHICON-2003*, 2003, pp. 85–92.

[14] (25 May 2012) Wiki-opencv. [Online]. Available: http://opencv.willowgarage.com/