https://doi.org/10.1109/EMBC.2019.8857743

Motion Signatures for the Analysis of Seizure Evolution in Epilepsy

David Ahmedt-Aristizabal[1,2], M. Saquib Sarfraz[2,3], Simon Denman[1], Kien Nguyen[1],
Clinton Fookes[1], Sasha Dionisio[4], Rainer Stiefelhagen[2]

***Abstract*— In epilepsy, semiology refers to the study of patient behavior and movement, and their temporal evolution during epileptic seizures. Understanding semiology provides clues to the cerebral networks underpinning the epileptic episode and is a vital resource in the pre-surgical evaluation. Recent advances in video analytics have been helpful in capturing and quantifying epileptic seizures. Nevertheless, the automated representation of the evolution of semiology, as examined by neurologists, has not been appropriately investigated. From initial seizure symptoms until seizure termination, motion patterns of isolated clinical manifestations vary over time. Furthermore, epileptic seizures frequently evolve from one clinical manifestation to another, and their understanding cannot be overlooked during a presurgery evaluation. Here, we propose a system capable of computing motion signatures from videos of face and hand semiology to provide quantitative information on the motion, and the correlation between motions. Each signature is derived from a sparse saliency representation established by the magnitude of the optical flow field. The developed computer-aided tool provides a novel approach for physicians to analyze semiology as a flow of signals without interfering in the healthcare environment. We detect and quantify semiology using detectors based on deep learning and via a novel signature scheme, which is independent of the amount of data and seizure differences. The system reinforces the benefits of computer vision for non-obstructive clinical applications to quantify epileptic seizures recorded in real-life healthcare conditions.**

I. INTRODUCTION

During seizures, patients with epilepsy may exhibit stereotypical behavior or motor manifestations. These can include jerking, spasm or posturing, head turning, facial expressions and hand movements. The analysis of such signs is termed *semiology*. Along with the electrophysiological and neuroimaging recordings, seizure semiology constitutes a crucial set of clues, which provides localizing information of the brain networks affected, enabling the progression to successful surgery in patients who are drug-resistant [1], [2]. Nevertheless, the study of video monitoring recordings is, to a certain extent, dependent on the experience and training of the clinician and the interpretation can differ from physician to physician, and between cases. Automated quantification and interpretation of semiology enables more objective information gathering from epileptic patients [3]. However, this is a challenging and largely underdeveloped field due to a lack of datasets and the highly complex natural clinical setting [4].

Recent advances in computer vision and deep learning have shown promising results in quantifying and distinguishing epileptic seizures [4] eliminating the need for feature engineering. However, because of their reliance on supervised learning, when unusual seizures are encountered such systems are of no use and they don't provide clinicians with intuitive tools to support the assessment of seizure evolution. Approaches that use statistical information [5], provide quantitative movement parameters by considering the totality of semiology duration *i.e.* one metric describes the motion for the full length of the seizure; and as such, this method does not capture information on the semiology changes. Deep learning based approaches [6], [7], [8] cannot determine the discriminative features that differentiate semiology and their relationship with specific body movements, or what portion of the body should be observed for diagnosis. These systems provide a single result that reflects the classifier's decision and (to an extent) confidence. Additionally, by analyzing short video sequences rather than a whole video of the seizure [9], [10] the systems are limited when distinguishing frames related to the clinical onset from those which show the propagation of semiology. Overall, the task of developing a computer-based tool that may analyze the stepwise progression of clinical features, which is the scope of this manuscript, has neither been considered nor reported in the literature.

The analysis of seizure evolution, which aims to identify the presence or absence of certain movement features (including the order in which they occurred) and the dynamic changes in movement frequency and amplitude during a seizure, is a major component of epilepsy patient assessment [1]. This is an important step to evaluate electroclinical patterns of a seizure, *i.e.* a close observation of clinical features (semiology) and their relation to the region primarily or secondarily involved in the epileptic discharge, allowing a spatio-temporal profile of the seizure's origin and propagation patterns to be obtained [2], [11]. Seizures characterized by motor manifestations are analyzed and classified on the basis of the type of motor symptomatology. Simple motor seizures are defined by unnatural movements, which can be divided into myoclonic, clonic, tonic, versive, and tonic-clonic seizures; depending on the duration of the muscle contraction, the rhythmicity of movement repetition, and the muscle involved (*e.g.* asymmetric posture, flexion of the

This research was supported by Mater Centre for Neuroscience under Research Governance authorization RG-17-008 and the KIT research exchange sponsorship from the Deutschen Akademischen Auslandsdienst (DAAD).

[1]Image and Video Research Laboratory, SAIVT, Queensland University of Technology, Brisbane, Australia. david.aristizabal@hdr.qut.edu.au

[2]CV-HCI Laboratory, Karlsruhe Institute of Technology, Karlsruhe, Germany. muhammad.sarfraz@kit.edu

[3]Daimler TSS, Germany.

[4]Department of Mater Advanced Epilepsy Unit, Mater Centre for Neurosciences, Brisbane, Australia.

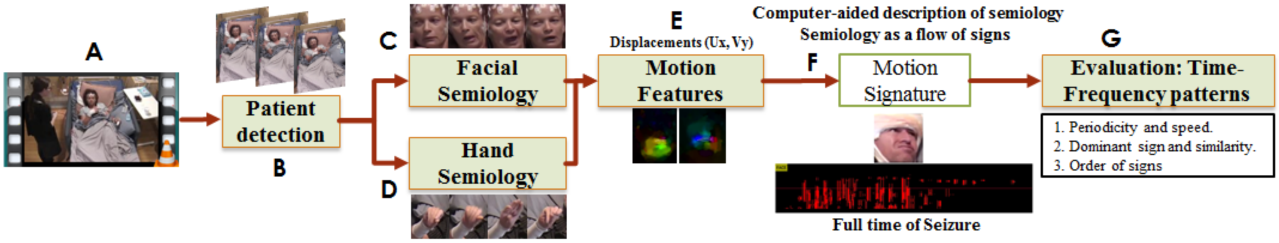

Fig. 1. The proposed system that captures the motion dynamics of semiology and creates a motion signature that represents the evolution of an epileptic seizure. **A.** Video recording during seizures. **B.** A region of interest is defined to improve the detection of isolated clinical manifestations: face and hands. **C.** Sequences of facial semiology are created based on face detection and tracking. **D.** Sequences of hand semiology are produced via pose estimation techniques. **E.** Extraction of motion features is performed using the optical flow from aligned sequences of consecutive images. **F.** Sequences of flow vectors are used to compute the motion signature for the full length of semiology in terms of time and position. **G.** The computer-aided tool visualizes semiology as a flow of signs to provide quantitative information to support the diagnosis of patients.

neck, abduction of both arms, turning of eyes and head to one side). In more complicated motor manifestations, on the other hand, patients may experience movements that appear natural and involve different body segments (*e.g.* manual and oral automatisms such as chewing, swallowing, smacking the lips and flumbing) [1], [12]. Seizure manifestations may vary from repetitive rhythmic movement of trunks, limbs or hands such as whole body rocking or manipulation of an object to a more extreme form of presentation with excessive amounts of amplitude, speed, and acceleration [2], [13].

Motivated by the significance of analyzing the evolution of semiology, in this contribution we provide a system that visualizes the dynamic changes in semiology over an entire seizure, which we term a *motion signature*. The system has the potential to provide an overview of the motion patterns observed and will support the assessment of patients independent of the motion rate and range, and the amount of data available. We adopt a framework that aims to capture semiology from video recordings and provides interpretable signals of the motion as presented in Fig. 1. We exploit the discriminative power of deep learning architectures to detect body regions for isolated clinical manifestations (face and hand semiology), and to extract representations of motion between consecutive frames. Then, we compute the *motion signature* from a sparse saliency representation established by the magnitude of the motion. The signature (the semiology as a signal) highlights the motion history of the seizure and correlates types of semiology in the case that a patient experiences multiple semiologies simultaneously. The motion profile uses sliding windows or sequences to simultaneously capture movement locations and temporal relations of the motion based entirely on the flow information.

The contributions of our work are summarized as follows:

1) We present a first of its kind computer-aided system that captures the dynamics of semiology as a flow of signals enabling the visualization and diagnosis of the seizure, *e.g.* semiology evolution and the correlation between body parts in a real-life clinical setting.
2) The system forms the basis for further research to assess electroclinical features [4] for monitoring tools, based on multisensory feature techniques that can fuse visual and time-series signals (electroencephalography) [14].

The remainder of this paper is organized as follows: Section II explains the video recordings used from patients with epilepsy and discusses our proposed approach to develop the motion signature; Section III presents the experimental setup to validate the capability of the system and provides examples of how time-frequency properties can be used for semiology diagnosis. Finally, Section IV concludes the paper.

II. MATERIALS AND METHODS

In this paper, we propose a computer vision approach that computes motion signatures that describe the evolution of semiology from video recordings of patients with epilepsy. A block diagram of the proposed system is displayed in Fig. 1. Each recorded seizure has a duration of roughly 1-2 minutes, and is captured at 25 frames per second. We capture clinical manifestations from face and hand motions by implementing deep learning and image processing techniques. Then, we estimate the patterns of apparent motion in a defined sequence at pixel-level by computing the optical flow between consecutive frames (*i.e.* a displacement vector assigned to each pixel position). This motion representation is used to estimate the motion signature, which captures the spatial location of semiology and the temporal relation between frames. The proposed system enables the diagnosis of seizures from video feeds as a flow of signals and we show how the motion signature can be used for diagnosis of epilepsy using quantitative information from signal processing techniques. Details of our proposed system are described in the following subsections.

A. Video monitoring of epileptic seizures

We acquired data to compute the motion signatures from epileptic patients undergoing the routine Video-EEG (scalp electroencephalography) and Video-SEEG (stereo electroencephalography) monitoring protocol at the Mater Hospital in Brisbane. Participants were diagnosed with mesial temporal (MTLE) and extra-temporal (ETLE) lobe epilepsy. Cameras are positioned on a wall, opposite the end of the bed (Fig. 1A) and have an infrared capability, enabling night monitoring. A total of 25 seizures from 7 patients were analyzed. In this paper, we illustrate the potential of motion signatures to evaluate the representation of semiology by discussing three selected patients who show diverse face and hand semiology.

Fig. 2. Selected sequences of facial semiology captured with the face detection and tracking architecture.

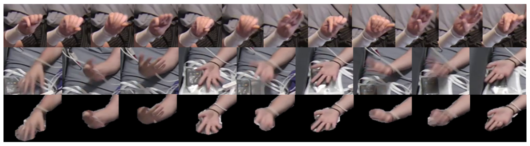

Fig. 3. Selected sequences of hand semiology captured using the hand detection strategy. Top: subtle finger motions; Middle: fast hand motions (waving); Bottom: copped images with background removed.

B. *The motion signature for analysis of seizure evolution*

1) Region of interest definition: We define the region of interest (RoI) as the location of the bed that contains the patient. This ensures that motions from the face and hands come from the patient, and not family members or physicians also visible in the videos. This also helps overcome changes in camera-bed viewing angles, in the inclination angle of the bed, and camera resolution. The detection of this region is inspired by the approach of [10], which uses the bounding box coordinates of the detected patient and bed to define the RoI in the x-axis of the frame, and retain the original height of the video. To estimate the RoI, we use Mask-RCNN [15] trained on the COCO dataset [16]. We expand the detected RoI with an offset of 20% of the total width on each side to avoid the extremities of the patient being located outside of this boundary due to movements during a seizure. We crop all images with the RoI as shown in Fig. 1B.

2) Face detection: From each seizure we extract sequences of images that capture facial modifications commonly exhibited such as unilateral blinking, chewing automatisms, ictal pouting, and smacking [1]. We adopt the implementation of [17] to detect the patient's face during epileptic seizures. This approach can perform face detection in challenging scenarios, and better deals with scale variation in the benchmark face datasets FDDB [18] and Wider-Face [19], outperforming the previous results documented with seizures [7], [10]. The end-to-end trainable model, which uses a backbone architecture based on ResNetXt with a depth of 101 layers [20], first creates a coarse image pyramid with the input image and $2X$ interpolation. Then, shared CNNs predict template responses (for both detection and regression) at every resolution. Finally, the model uses non-maximum suppression (NMS) at the original resolution to get the final detection results.

To consistently localize the face bounding box in terms of size and position with minimal jitter between frames, we fuse the face detector with a tracking algorithm for videos sequences based on the open source SORT tracker [21] and it's extensions proposed in [22], which integrates appearance information using a deep descriptor. Fig. 2 illustrates sequences of facial semiology detected and cropped using this approach.

3) Hand detection: We capture hand and finger semiology such as waving, snapping fingers, tapping or grabbing, thumb adduction, and fumbling [1], [23] by detecting both hands automatically.

Based on the results in [10], we detect each hand using region-based methods that detect hand-bounding boxes in challenging healthcare conditions. Although hand detection performance has progressed significantly with the use of CNNs [24], we adopt a hand detection strategy based on the body pose location, where the position of the wrist and elbow are used to approximate the hand location as discussed in [25]. This approach, which is heavily constrained by the predicted pose, allows us to capture fast motions involving upper limb translations (*e.g.* waving), and helps ensure that the fingers of the patient are not located outside the bounding box due to fast movements during a seizure.

For human pose estimation in videos, we adopt the architecture of [26]. This lightweight, yet highly effective two-stage approach first uses a 3D Mask R-CNN [15] to predict the human pose, then implements a lightweight optimization that links the predictions in time. We use a model trained on the PoseTrack [27] dataset to detect the patient wrist and elbow in each frame. Once the pose estimation is performed, we estimate the location of each hand according to [25] and crop the images to a fixed bounding-box size of 120×120 pixels.

The extracted hand bounding box captures all motions related to the hand and fingers, but also includes information pertaining to background motions such as movements in the bedding, cables and monitoring equipment. To suppress background motion, we adopt a simple strategy of skin segmentation using thresholds adapted for the illumination conditions of our dataset. This algorithm is implemented in OpenCV [28] based on the HSV color space. Samples of detected hands and the background removal are depicted in Fig. 3.

4) Extraction of motion features in sequences: We adopt an optical flow based strategy to capture important information from each type of semiology, including the spatial arrangement of body parts and the rate of change of the arrangement [29].

Prior to computing the optical flow, successive frames are geometrically aligned by warping the images relative to each other and comparing the pixel intensity values using the enhanced correlation coefficient (ECC) [30]. We use the Euclidean transformation model where the aligned image is a rotated and shifted version of the first image. This method gives good results under various changes in brightness and contrast. Once all successive pairs of images are aligned, we resize all images to a resolution of height $H = 224$ and width $W = 224$ pixels and compute the optical flow.

The optical flow is computed between adjacent frames using FlowNet v2 [31], which is a coarse-to-fine approach that uses stacking CNNs for optical flow refinement allowing the robust analysis of small displacements. We use one threshold on the flow to ensure that there is motion in the frame, *i.e.* more than 10% pixels have optical flow values

Fig. 4. Selected samples of a motion signature of showing facial semiology. Upper: patient DG with a fast mouth and eye movement (ETLE-opercular-upper bank). Lower: patient PU with subtle or slow lower mouth movements (MTLE-lower areas). At the right, representation of the landmarks used to estimate the location of the nose.

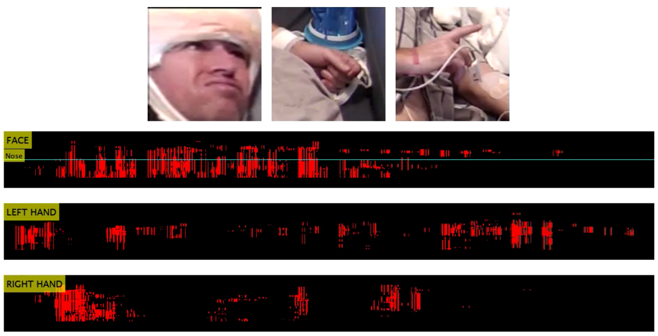

Fig. 5. Selected samples visualizing the motion signature of face and hand semiology. Patient MB with fast cheek motion is diagnosed with MTLE.

above zero. Finally, we obtain $N-1$ optical flow maps for N frames, where each flow map has horizontal and vertical (u and v) components.

5) Constructing the proposed motion signature: To analyze the evolution of a seizure as a flow of signs, we develop a compact image representation of semiology, using the optical flow information, which illustrates the location variance and periodicity of motion. To have an intuitive understanding of our proposal, consider two flow maps of size $H \times W$ (the same size as the original image/frame). We can measure the change in the motion patterns between them by computing an absolute difference. This should only be high at spatial locations where there was strong movement. We can summarize this change along a given direction in this difference flow map. We sum the obtained difference values along the horizontal direction W, thereby getting a $H \times 1$ motion profile between two flow maps. This motion profile represents spatial motion from top to bottom of the image along the W direction. If we keep computing such motion profiles between successive optical flow maps and stack them together, we will have a *motion signature* that represents the temporal change of motion along the horizontal x-axis over time, and captures the spatial motion of body parts (*e.g.*, eyes, mouth, etc.) along the vertical y-axis. Together, one can see the motion from top to bottom of the image and how it is progressing over time.

To obtain a stable estimate of temporal motion change, we define a temporal window of length L over successive optical flow maps. The sequence of L flow maps in this window is used to capture the motion change in the corresponding image frames. In the sequence, the motion profile (as described above) is computed between all combinations of optical flow maps. To understand it better, if our sequence window length $L=4$, we will compute the motion profile between flow map 1 and 2, 3 and 4, similarly flow map 2 with 1, 3, 4 and so on. Therefore, for a sequence length L we will have a total of $L \times (L-1)$ motion profiles. In our case for $L=4$ we generate 12 motion profiles. Using such all combinations provides a more stable motion progression and proves to be more robust with respect to any noise such as misalignment between frames. Since we have a defined sequence window to compute such motion profiles we can use the average of the values in the obtained motion profiles (a scalar) as a threshold to only keep real motion. Such a binarization of motion profiles helps to determine if a motion segment is related to a clinical manifestation or noise from the optical flow estimation. This strategy allows us to use it as an online monitoring tool of the evolution of motion *e.g.* by using a buffer of $L+1$ frames.

Fig. 4 illustrates the motion signature computed from two patients that experienced facial semiology and Fig. 5 displays the motion signature of a patient that exhibited face and hand semiology simultaneously. The motion of each displacement vector is summarized and represented by a red vertical bar distributed in the x-axis from the start to the full expression of the clinical manifestation and the y-axis corresponds to the spatial location of the motion in the input frame. Black means no seizure related motion.

For the diagnosis of facial semiology, we consider that the location of the nose divides the face into two regions: upper and lower areas. The upper area corresponds to motions in upper facial regions (eyes and eyebrows), while the lower area is related to mouth and chin motions as depicted in Fig. 4. The continuous line in the facial semiology signature represents the average location of the nose during the full expression of semiology, in order to appreciate the difference between motions from upper and lower facial regions. We compute the location of the nose by adopting the state-of-the-

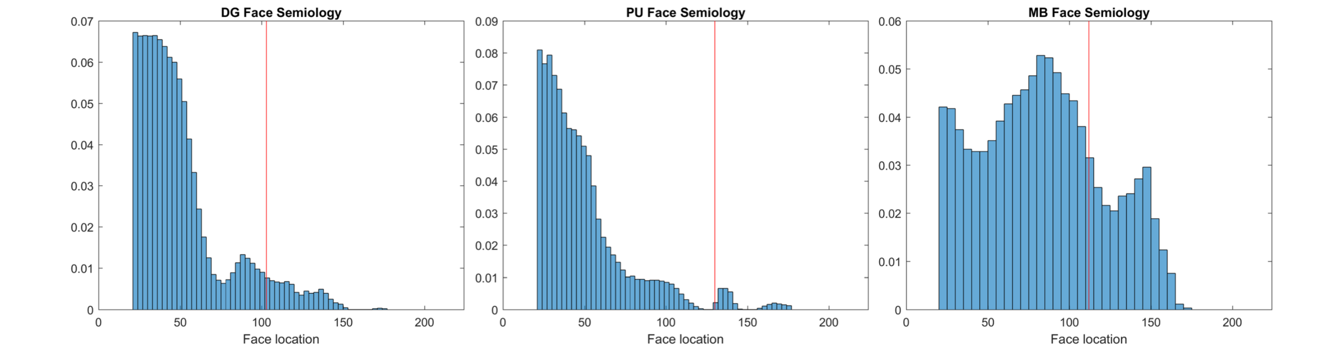

Fig. 6. Visualization of dominant motions in the face through normalized histograms. Patients exhibit semiology in the eyes, mouth, chin and cheeks. The vertical line represents the average nose location.

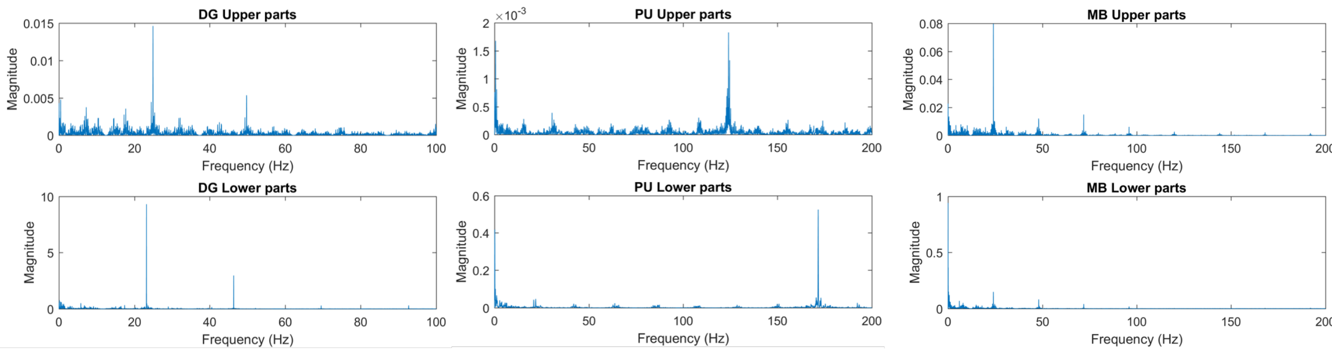

Fig. 7. Visualization of the periodogram for the upper and lower facial regions for each patient.

art facial 2D & 3D landmark estimation system [32]. This location is considered as the average estimation of the six facial landmarks that represent the lower nose as shown at the right in Fig. 4. To provide and display the information in an appealing way, motion signatures are saved in a video format that allows the user to visualize the motion and the visible image for the entire seizure simultaneously.

III. EXPERIMENTS

A. Experimental setup

To demonstrate the capability of the system to quantify semiology based on motion signatures, we show how this representation can be used in a clinical environment by providing quantitative information from the representation of semiology as a flow of signs. We compute a flow that highlights the most common events and the changes within the recorded seizure, and we show how time-frequency properties computed from the motion signatures can support the assessment by:

- Analyzing the motion signature itself, *e.g.* which is the dominant and most frequent sign, blinking or mouth motion.
- Applying frequency-based analysis to identify periodic motions, and showing how we can also use autocorrelation to support this process, to illustrate if each semiology is periodic or a single episode, and it's speed.
- Using power spectrum analysis to quantify the strength of periodic components to determine the dominant semiology (face or hand).
- Displaying the order of signs as a stepwise progression, which is very important as it allows the analysis of underlying seizure spread.

B. Experimental results

Identifying dominant and frequent signs in facial semiology: The motion signatures of three seizures from three selected patients are represented as images in Fig. 4 and Fig. 5. To calculate time-frequency properties of each signature, we represent the image as a one-dimensional signal, which contains information of the motion location. Using the one dimensional signal of the facial motion, we compute histograms to quantify the number of events recorded in each face location. These histograms are shown in Fig. 6. The *x*-axis represents the face location (from lower face to upper face in a scale of 0-224 pixels) and shows the location of the nose. From the histograms, we can see that for the three patients the dominant sign is mouth semiology. Patient PU has motions in the lower mouth area, patient DG has motion in the lower and upper mouth and patient MB has more motions in the upper parts of the mouth (cheeks).

Identifying periodic motions in facial semiology: We analyze the periodicity and speed of a signal via spectral analysis based on the power spectral density (PSD) [33] by computing the periodogram, which is given as the discrete-time Fourier transformation of the auto-correlation function. Fig. 7 illustrates the periodogram for the upper and lower facial regions of each patient. Analyzing the lower facial parts for patients DG and PU, the spectral analysis shows statistically significant periods and harmonics, or cycles in the data that stand out from the background noise. However, for patient MB, there are no clear dominant oscillations or periodicity in the motion of the lower face. For the upper face for the three patients, there is considerable noise that affects the identification of cyclic behaviour, and they also

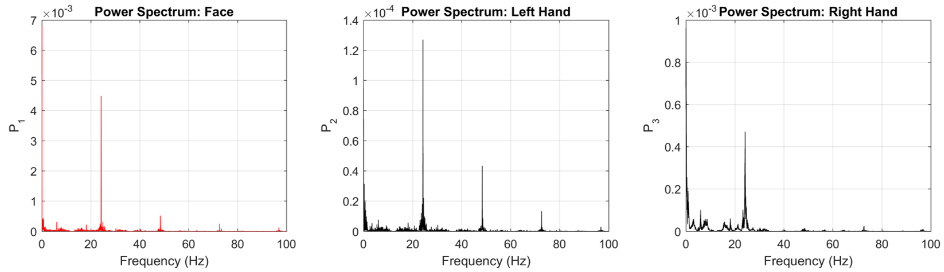

Fig. 8. Visualization of the power spectrum for face and hands semiology for patient MB.

TABLE I
AUTOCORRELATION IN THE TIME-DOMAIN.

Patient	Upper Face	Lower Face	Patient MB	
DG	0.3938	**3.2935** ^{P}	Face	0.8684
PU	0.3745	**2.2653** ^{P}	Left Hand	**1.3860** ^{P}
MB	0.6182	0.6204	Right Hand	0.6666

A value greater than one means the signal has high correlation once the lag time matches the period and can be considered periodic. *P*: Periodic.

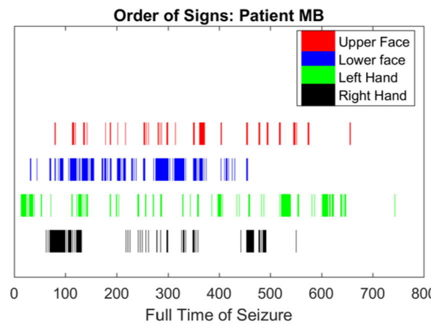

Fig. 9. Order of signs as a stepwise progression for patient MB .

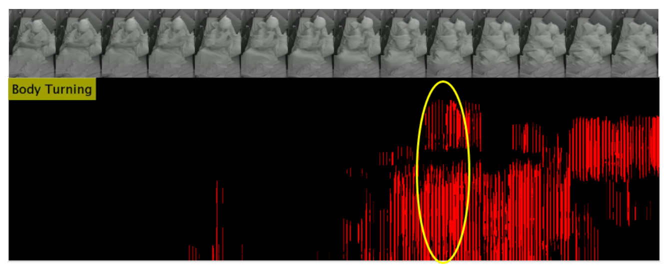

Fig. 10. Motion signature of the isolated semiology known as body turning.

show several spurious peaks that are likely caused by noise.

Overall, patient DG exhibits cyclic behaviour in the lower face with a frequency of approximately 25 Hz while patient PU shows cyclic behaviour in the lower face with a frequency of approximately 170 Hz. These dominant oscillations allow us to confirm that the speed of the motion for patient PU is higher than patient DG.

We confirm the analysis of periodicity of the fundamental spikes in the frequency domain with the autocorrelation of the signal in the time domain. The autocorrelation of a periodic signal has the same cyclic characteristics as the signal itself. Thus, autocorrelation can help verify the presence of periodic behaviour and determine the period [34]. If the data is periodic, it should have high correlation once the lag time matches the period. As shown in Table I, we can confirm that the mouth motions of patients DG and PU are periodic.

Dominant signs and periodicity considering face and hand semiology: To evaluate dominant signs, we also can estimate the power of each frequency by computing the power spectrum (PS). The PS of a time-domain signal is the distribution of power contained within the signal over frequency, based on a finite set of data. Considering the total average power as the sum of the power of all the frequency components of the signal, it can be estimated that the mouth motions are more powerful in patient DG (2.1063) than patient PU (1.0581). For patient MB, who experiences face and hands semiology (Fig. 5), we compute the total average power of the PS as shown in Fig. 8. It is possible to confirm that the facial motion is the dominant sign in the semiology according to the average power: Face (1.0271), right hand (0.0300) and left hand (0.0021). Only in the left-hand signal are dominant oscillations clear, with spikes in the other signals the result of noise. This periodicity of the signals is confirmed with the autocorrelation results (Table I), where the left hand can be considered to have a cycle of subtle motion. Considering all motions from the face compared with each hand it is possible to find matching frequencies. It can be seen in Fig. 5 that the signals have a similar component at 24 Hz. The order of signs of in patient MB is illustrated in Fig. 9, which is important as it allows the analysis of underlying seizure spread.

Analyzing the entire body simultaneously: The motion signature can be also implemented to analyze the whole body simultaneously, and to evaluate isolated semiology such as the complex motor behaviour of body turning [13]. Fig. 10 illustrates the motion signature for this type of semiology where it is possible to appreciate when the rotation happens and how long it lasts for.

C. Discussion

In this manuscript, we present a novel and intuitive computer-aided tool to support the expertise of clinical practitioners in the complex area of seizure semiology. The motion signatures are flexible and provide diagnostic assistance when analyzing videos in real-life healthcare conditions, presenting semiology as a flow of signs. This strategy enables the use of simple and robust time-frequency techniques to evaluate seizure recordings and isolate repeating patterns. The approach for assistive medical diagnosis in assessing video recordings of seizures, quantifying the dominance, correlation, and motion evolution of semiology from different body parts, has not been previously documented.

One drawback of our system is the reliance on the accurate detection of the regions we monitor for semiology (face and hand), and their alignment and the extraction of flow information; thereby triggering the need for further investigation. However the system is flexible in that performance of the motion detection and quantification can be easily improved (Fig. 1C,D,E) by incorporating new computer vision approaches. For example, it is worth evaluating the computational cost of considering image registration using deep convolutional techniques [35] which have comparable or better accuracy than feature-based or direct methods.

IV. CONCLUSIONS

In this work, we have presented an efficient, in both computation and architecture, computer vision approach to capture motion signatures of face and hand semiology, to provide a diagnostic tool to clinicians to evaluate the evolution of clinical manifestations in patients with epilepsy. The motion signatures of epileptic seizures provide relevant features to the physician and a way to intuitively assess the patient's movement, which is helpful for proper disease management. We expect that a computer-aided tool visualizing semiology as a signal could support the electroclinical analysis that neurologists perform, to aid the progression to successful surgery in patients who are drug-resistant to epilepsy. Finally, the simplicity of our method may enable the diagnosis of patients based on online and real-time monitoring of patients' behavior.

Ethics statement: The experimental procedures involving human subjects described in this paper were approved by the Mater Health Services Human Research Ethics Committee.

REFERENCES

[1] S. Noachtar and A. S. Peters, "Semiology of epileptic seizures: a critical review," *Epilepsy & Behavior*, vol. 15, no. 1, pp. 2–9, 2009.

[2] F. Bonini, A. McGonigal, A. Trébuchon, M. Gavaret, F. Bartolomei, B. Giusiano, and P. Chauvel, "Frontal lobe seizures: from clinical semiology to localization," *Epilepsia*, vol. 55, no. 2, pp. 264–277, 2014.

[3] S. Noachtar, "Video analysis for defining the symptomatogenic zone," *Handbook of Clinical Neurophysiology*, vol. 3, pp. 187–200, 2003.

[4] D. Ahmedt-Aristizabal, C. Fookes, S. Dionisio, K. Nguyen, J. P. S. Cunha, and S. Sridharan, "Automated analysis of seizure semiology and brain electrical activity in presurgery evaluation of epilepsy: A focused survey," *Epilepsia*, vol. 58, no. 11, pp. 1817–1831, 2017.

[5] J. P. S. Cunha, H. M. P. Choupina, A. P. Rocha, J. M. Fernandes, F. Achilles, A. M. Loesch, C. Vollmar, E. Hartl, and S. Noachtar, "Neurokinect: a novel low-cost 3dvideo-eeg system for epileptic seizure motion quantification," *PloS one*, vol. 11, no. 1, p. e0145669, 2016.

[6] F. Achilles, F. Tombari, V. Belagiannis, A. M. Loesch, S. Noachtar, and N. Navab, "Convolutional neural networks for real-time epileptic seizure detection," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–6, 2016.

[7] D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen, S. Denman, S. Sridharan, and S. Dionisio, "Deep facial analysis: A new phase i epilepsy evaluation using computer vision," *Epilepsy & Behavior*, vol. 82, pp. 17–24, 2018.

[8] Z. Fang, H. Leung, and C. S. Choy, "Spatial temporal gru convnets for vision-based real time epileptic seizure detection," in *ISBI*. IEEE, 2018, pp. 1026–1029.

[9] D. Ahmedt-Aristizabal, K. Nguyen, S. Denman, S. Sridharan, S. Dionisio, and C. Fookes, "Deep motion analysis for epileptic seizure classification," in *EMBC*, 2018.

[10] D. Ahmedt-Aristizabal, C. Fookes, S. Denman, K. Nguyen, S. Sridharan, and S. Dionisio, "A hierarchical multi-modal system for motion analysis in epileptic patients," *Epilepsy & Behavior*, vol. 87, pp. 46–58, 2018.

[11] P. Chauvel and A. McGonigal, "Emergence of semiology in epileptic seizures," *Epilepsy & Behavior*, vol. 38, pp. 94–103, 2014.

[12] W. T. Blume, H. O. Lüders, E. Mizrahi, C. Tassinari, W. van Emde Boas, and J. Engel Jr, Ex-officio, "Glossary of descriptive terminology for ictal semiology: report of the ilae task force on classification and terminology," *Epilepsia*, vol. 42, no. 9, pp. 1212–1218, 2001.

[13] H. Leung, K. Schindler, H. Clusmann, C. G. Bien, A. Pöpel, J. Schramm, P. Kwan, L. K. Wong, and C. E. Elger, "Mesial frontal epilepsy and ictal body turning along the horizontal body axis," *Archives of neurology*, vol. 65, no. 1, pp. 71–77, 2008.

[14] A. Owens and A. A. Efros, "Audio-visual scene analysis with self-supervised multisensory features," in *ECCV*, 2018.

[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *ICCV*. IEEE, 2017, pp. 2980–2988.

[16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*. Springer, 2014, pp. 740–755.

[17] P. Hu and D. Ramanan, "Finding tiny faces," in *CVPR*, 2017, pp. 1522–1530.

[18] V. Jain and E. G. Learned-Miller, "FDDB: A benchmark for face detection in unconstrained settings," *UMass Amherst Technical Report*, 2010.

[19] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *CVPR*, 2016, pp. 5525–5533.

[20] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *CVPR*. IEEE, 2017, pp. 5987–5995.

[21] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *ICIP*. IEEE, 2016, pp. 3464–3468.

[22] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *ICIP*. IEEE, 2017, pp. 3645–3649.

[23] B. C. Jobst, A. M. Siegel, V. M. Thadani, D. W. Roberts, H. C. Rhodes, and P. D. Williamson, "Intractable seizures of frontal lobe origin: clinical characteristics, localizing signs, and results of surgery," *Epilepsia*, vol. 41, no. 9, pp. 1139–1152, 2000.

[24] S. Yan, Y. Xia, J. S. Smith, W. Lu, and B. Zhang, "Multiscale convolutional neural networks for hand detection," *Applied Computational Intelligence and Soft Computing*, vol. 2017, 2017.

[25] T. Simon, H. Joo, I. A. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping." in *CVPR*, vol. 1, 2017, p. 2.

[26] R. Girdhar, G. Gkioxari, L. Torresani, M. Paluri, and D. Tran, "Detect-and-track: Efficient pose estimation in videos," in *CVPR*, 2018, pp. 350–359.

[27] M. Andriluka, U. Iqbal, A. Milan, E. Insafutdinov, L. Pishchulin, J. Gall, and B. Schiele, "Posetrack: A benchmark for human pose estimation and tracking," in *CVPR*, 2017.

[28] G. Bradski *et al.*, "The opencv library," *Doctor Dobbs Journal*, vol. 25, no. 11, pp. 120–126, 2000.

[29] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[30] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.

[31] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *CVPR*, vol. 2, 2017.

[32] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)," in *ICCV*, 2017.

[33] S. A. Fulop and K. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 360–371, 2006.

[34] M. Vlachos, P. Yu, and V. Castelli, "On periodicity detection and structural periodic similarity," in *Proceedings of the SIAM international conference on data mining*. SIAM, 2005, pp. 449–460.

[35] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Deep image homography estimation," *arXiv preprint arXiv:1606.03798*, 2016.