Epilepsy & Behavior 82 (2018) 17–24

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Deep facial analysis: A new phase I epilepsy evaluation using computer vision

David Ahmedt-Aristizabal [a,*], Clinton Fookes [a], Kien Nguyen [a], Simon Denman [a], Sridha Sridharan [a], Sasha Dionisio [b]

[a] Speech, Audio, Image and Video Technologies (SAIVT) Research Program, School of Electrical Engineering & Computer Science, Queensland University of Technology, Brisbane, Australia
[b] Department of Mater Advanced Epilepsy Unit, Mater Centre for Neurosciences, Brisbane, Australia

ARTICLE INFO

Article history:
Received 13 November 2017
Revised 7 February 2018
Accepted 14 February 2018
Available online xxxx

Keywords:
Epilepsy evaluation
Facial semiology
Deep learning
Convolutional neural network (CNN)
Long short-term memory (LSTM)
Neuroethology

ABSTRACT

Semiology observation and characterization play a major role in the presurgical evaluation of epilepsy. However, the interpretation of patient movements has subjective and intrinsic challenges. In this paper, we develop approaches to attempt to automatically extract and classify semiological patterns from facial expressions. We address limitations of existing computer-based analytical approaches of epilepsy monitoring, where facial movements have largely been ignored. This is an area that has seen limited advances in the literature. Inspired by recent advances in deep learning, we propose two deep learning models, landmark-based and region-based, to quantitatively identify changes in facial semiology in patients with mesial temporal lobe epilepsy (MTLE) from spontaneous expressions during phase I monitoring. A dataset has been collected from the Mater Advanced Epilepsy Unit (Brisbane, Australia) and is used to evaluate our proposed approach. Our experiments show that a landmark-based approach achieves promising results in analyzing facial semiology, where movements can be effectively marked and tracked when there is a frontal face on visualization. However, the region-based counterpart with spatiotemporal features achieves more accurate results when confronted with extreme head positions. A multifold cross-validation of the region-based approach exhibited an average test accuracy of 95.19% and an average AUC of 0.98 of the ROC curve. Conversely, a leave-one-subject-out cross-validation scheme for the same approach reveals a reduction in accuracy for the model as it is affected by data limitations and achieves an average test accuracy of 50.85%. Overall, the proposed deep learning models have shown promise in quantifying ictal facial movements in patients with MTLE. In turn, this may serve to enhance the automated presurgical epilepsy evaluation by allowing for standardization, mitigating bias, and assessing key features. The computer-aided diagnosis may help to support clinical decision-making and prevent erroneous localization and surgery.

1. Introduction

Epilepsy is among the most common of the neurological conditions. Mesial temporal lobe epilepsy (MTLE), often with hippocampal sclerosis, is one of the most common causes of drug-resistant epilepsy [1]. Epilepsy surgery has been accepted as an effective treatment for patients with medically refractory epilepsy or whose seizures are nonresponsive to medication [2]. The complete resection of the epileptogenic zone (*i.e.*, the region of the brain that generates epileptic seizures) is the primary goal. Patients with epilepsy exhibit different clinical manifestations based on the underlying networks activated. Semiology has played a pivotal role to provide localizing and lateralizing information in order to allow for successful surgery in addition to neurophysiological and imaging data [3,4]. While semiology is important, a single sign in isolation is not helpful but rather the progression of events that underlie the integration of various neuronal networks. In MTLE for example, certain facial modifications are more commonly exhibited (although not exclusive), including unilateral blinking, eye deviation, chewing automatisms, fear expression, disgust, unilateral mouth deviation, and postictal nose wiping [5–8].

Epilepsy monitoring relies on video analysis to assist with the diagnosis of seizures. However, this evaluation is subjective and dependent on observer experience and may lead to misdiagnosis [3]. Automated analysis of semiological patterns, *i.e.*, detection, quantification, and recognition of body movement patterns, could help increase diagnostic precision [9] by standardizing the assessment evaluation among evaluators and identifying features that are unambiguous. However, the automated analysis of semiology has made little progress over recent years [10,11]. The majority of existing automated systems are limited to the ictal analysis of limb and head movements [10,12]. While some attempts at automating the semiology of facial expressions

* Corresponding author at: Science and Engineering Faculty, School of Electrical Engineering and Computer Science, Queensland University of Technology, S Block, 612, GP, GPO Box 2434, Brisbane, QLD 4001, Australia.
E-mail address: david.aristizabal@hdr.qut.edu.au (D. Ahmedt-Aristizabal).

https://doi.org/10.1016/j.yebeh.2018.02.010

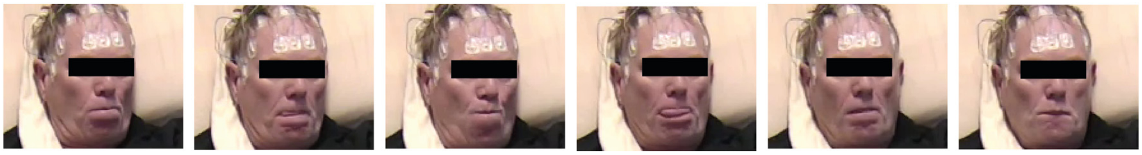

Fig. 1. Selected example of facial semiology from mouth motion in patients with MTLE.

have been made [13–16], the field is still largely unexplored. One reason for this includes the immense complexity in detecting and tracking key facial regions, especially in the clinical environment, where the face may often be obscured from view with electrodes, bedding, inadequate camera capture and positioning, poor illumination, and movements during seizures [10].

Deep learning (DL) has entered the mainstream in computer vision and machine learning in the last several years, achieving near-human and superhuman performance in many tasks such as object detection and sequence learning [17]. Using DL, researchers have also demonstrated state-of-the-art performance in analyzing videos, outperforming traditional techniques in emotion recognition and facial expression analysis [18–20]. Deep learning is now becoming widely employed in biological and medical applications; however, despite its advantages, there has not been an application of this technology for the purpose of monitoring facial changes in seizures, particularly in presurgical evaluation. While automated detection of electroencephalography (EEG) signals based on DL exists to help identify seizures [21,22], apart from video recordings, there are no devices which detect ictal changes in semiology [10]. Deep learning is a promising field for analyzing video data because of its advantages in automatically learning key features extracted from raw data. Deep learning is a technique that can be adapted to new problems because of its ability to perform transfer learning, *i.e.*, learning on one dataset and applying the trained model to another. Thus, this results in a richer representation and greater learning capability [17,23].

Given that epilepsy surgery requires considerable accuracy to help the patient [24], this research has been developed to improve diagnostic precision using quantitative methods from clinical data. In this paper, we endeavor to develop quantitative methods that characterize motion semiology using facial expression. We have concentrated our techniques to distinguish between ictal and nonictal/random facial expressions in patients with MTLE. Localization of MTLE was confirmed in these patients with a combination of stereo-EEG assessment or seizure freedom for over two years in the setting of a lesion, *i.e.*, hippocampal sclerosis. The techniques and equipment employed in this study are a combination of video detection systems and advanced computer vision techniques. Deep learning architectures characterize the various semiological patterns from quantitative motion detection and training. The learned patterns from ictal facial modification are extracted to identify between ictal and nonictal patterns. We conducted experiments using our own dataset jointly developed by the Queensland University of Technology, Australia (QUT) and the Mater Advanced Epilepsy Unit, Brisbane, Australia. The remainder of this paper is organized as follows: Section 2 describes our dataset, the methodology, and experimental plan; Section 3 presents the results; and Section 4 discusses the main findings and the significance of the results. Finally, Section 5 draws the paper's concluding remarks.

2. Materials and methods

2.1. Video monitoring dataset

The video recordings were captured as a part of the routine long-term video-EEG monitoring protocol at the Mater Hospital in Brisbane, Australia with patients with epilepsy who were undergoing phase I workup for their drug-resistant epilepsy. The patients with epilepsy were monitored over a time period ranging from 2 to 7 days. The patients with epilepsy were selected if clinical evaluation suggested a possible surgical procedure may be suitable or if invasive studies were necessary to analyze the onset. A random sample of 16 patients with MTLE was retrospectively selected from the overall dataset. Localization of MTLE was confirmed either from a stereo-EEG evaluation or if a temporal lobectomy had been performed in the setting of hippocampal sclerosis, with seizure freedom of no less than 2 years.

All seizures recorded from patients with MTLE were assessed and categorized according to gestural motor behaviors including chewing, blinking, fear or wide-open eyes, eye gazing, and motions in the mouth area as illustrated by a selected example in Fig. 1. The observation of semiology was the essence of the first step of this study, where it was crucial to choose well-defined terms to describe different signs. Instances of seizures in the video recorded for the dataset were selected from the first epileptic discharge until the full expression of semiology prior to version and convulsion if it was experienced. All digitized recorded images from each video clip, recorded at a frame rate of 25 frames/s, were in the PNG format with an image dimension of 1280×720 pixels.

Following this preliminary study, we developed our dataset to perform quantitative identification between facial expressions during ictal (Class 1) and nonictal (Class 2) events. Table 1 illustrates the demographic statistics of the 8 patients nominated as Class 1 with the most common ictal pattern, while for Class 2, we randomly selected 8 patients with epilepsy with video clips recording with nonictal/random facial expressions such as answering questions from the doctors, eating, watching television, and speaking with family members. To avoid incorrect instances of natural behavior in Class 2, video recordings of interictal periods were not considered. A total of 55 videos clips from the day and night monitoring were recorded, representing 24 videos for Class 1 and 31 videos for Class 2.

Seizure events are uncommon, and public data of semiology are absent. Inadequate training data hinder the validation of algorithms that could quantify semiology. For this reason, in order to exploit the ability of transfer learning of DL architectures in the epilepsy task, public datasets traditionally used in the facial analysis under unconstrained conditions were considered to train models used in the proposed approach and will be described in Sections 2.2.1 and 2.2.2.

2.2. The proposed system

Given a sequence of color images of a patient, our method estimates whether a facial expression sequence has an ictal pattern of MTLE. A facial expression can be observed as a dynamic variation of key parts, which are fused to form the variation of the whole face. The aim of our

Table 1
Patients with mesial temporal lobe epilepsy (MTLE) demographics during ictal activity.

Test subject	Number of seizures	Number of frames	Main semiology
1	1	1700	Mouth and tongue movement
2	2	3750	Mouth movement and left eye blinking
3	3	1700	Fear expressions
4	2	1025	High blinking frequency
5	2	2225	Mouth and tongue movement
6	4	3750	Fear expressions and blinking
7	4	3600	Mouth movement and swallowing
8	6	6675	Mouth and tongue movement
Total	**24**	**24,425**	