Tampere University

Yahya Abdelnasser Fawzy Kamel Barghash

SEIZURE INTENSITY ESTIMATION USING VIDEO RECORDING

Faculty of Information
Technology and Communication
Sciences

Master of Science Thesis

May 2020

ABSTRACT

Yahya Barghash: Seizure intensity estimation using video recording

Master of Science Thesis

Tampere University

Master's Degree Programme in Electrical Engineering

May 2020

Epilepsy is one of the most common neurological diseases worldwide. It affects around 50 million people worldwide. Therefore, it is believed to be one of the most common neurological diseases globally. It is defined as a state of recurrent seizures, which induce, in many cases, involuntary and abnormal movements . These movements may involve a part of the body (partial seizures) or the entire body parts (generalized seizures). This motor manifestation during seizures represents a clinical marker of seizure identification and has a significant role in evaluating epilepy in addition to the treatment impact on the progression of the disease.The current method of epileptic seizures analysis are qualitatively evaluated by professionals with visual inspection. Threrefore it is a highly subjective method and susceptible to human errors and bias.

In this thesis, an automatic tool (i.e. pose estimation) was used to obtain 3D kinematic analysis of epileptic seizures from stereo video-recordings. The data samples and the automatic tool used in this thesis are provided by Neuroeventlabs Oy. Seizure events belong to a patient with Lennox–Gastaut syndrome (LGS). They consist of four nocturnal seizure events recorded during sleep at home. 3D kinematic analysis of epileptic seizure events was conducted for intensity assessment. To provide a qualitative assessment, visual assessment was done on the seizure events as well as a seizure-free event during REM sleep to assess their severity on a simple rating scale having categorical values (i.e. weighted score) from (0-9).

The main goal of the thesis is to conduct a comparative experiment between the human visual assessment (as the gold standard in the field) and the automatic quantitative approach in a control-free environment for to demonstrate whether it is possible to use the current computer vision and deep learning techniques for conducting motion analysis during seizures.

Keywords: epilepsy, nocturnal seizures, kinematics, computer vision, stereo, pose estimation

PREFACE

This master's thesis was carried out in Neuroeventlabs Oy, Tampere, Finland.

Firstly, I would like to thank my family and fiancé for giving me support to complete my degree.

I would also like thank Anna Hakala and Marko Niemelä (Neuroeventlabs Oy) for providing me with the necessary means to get and process the data needed for my thesis.

I would also like to show sincere gratitude to Associate Professor Juha Nousiainen, Associate Professor Joni Kämäräinen and Professor Järi Viik who taught me a lot during my studies, I owe them much.

I dedicate my thesis to every patient with epilepsy across the globe. They and their relatives have been suffering and I hope my work becomes a modest small leap forward towards the relief of their pain.

Tampere, April 2020

Author

Yahya Barghash

CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2D	Two dimensions
3D	Three dimensions.
ACM	Accelerometer
AED	Antiepileptic drug
CNN	Convolutional Neural Network
CNS	Central nervous system
CPN	Cascaded Pyramid Network
DSI	Disparity space image
EEG	Electroencephalogram
EMUs	Epilepsy Monitoring Units
FPS	Frames per Second
EPSP	Excitatory postsynaptic potential
IBE	International Bureau for Epilepsy
ILAE	International League Against Epilepsy
ILP	Integer Linear Problem
IPSP	Inhibitory postsynaptic potential
IR	Infrared
LGS	Lennox–Gastaut Syndrome
NMJ	Neuromascular junction
PAF	Part affinitiy field
PNS	Peripheral Nervous System
PSP	Post synaptic potential
REM	Rapid eye movement sleep
ROI	Region of interest
SMA	Supplementary Motor Area
SEEG	Stereoelectroencephalography
WHO	World Health Organization

c_0, c_1	Camera planes
d	Disparity
f	Focal length
u	A pixel coordinate in x-dimension
v	A pixel coordinate in y-dimension
x,x'	x-coordinates of points in pixels
X	x-coordinate in real world
x_p	x-coordinate of the principal point in pixels
y, y'	y-coordinates of points in pixels
Y	y-coordinate in real world
y_p	y-coordinate of the principal point in pixels
$I[]$	Input signal to a filter
$O[]$	Output signal from a filter
Z	z-coordinate in real world
ρ_h	Height of a pixel
ρ_w	Width of a pixel

1. INTRODUCTION

Epilepsy is one of the most common neurological diseases. It affects 50 million people worldwide [1]. Almost a tenth of the entire population have at least one epileptic seizure during their life span and a third of those develop epilepsy [2]. According to the WHO organization, epilepsy is responsible for 1% of the global burden of disease, putting it head-to-head with lung cancer in men and breast cancer in women [3]. Epilepsy can affect people of all ages and is characterized by a variety of manifestations and causes [2]. It has negative impacts on the patients socially, psychologically and physically. Therefore, precise diagnosis and regular monitoring of differential epilepsy are necessary to improve the living conditions of the patients and their relatives.

Epilepsy and seizures have been observed since the rise of history. People with epilepsy were thought to be possessed or seized by supernatural powers during seizures, hence the term “seizure”. Epilepsy is currently defined as “a state of recurrent spontaneous seizures” [4], since epileptic seizures are considered the main symptoms of epilepsy [5]. However, the difference between epilepsy and seizures can sometimes be a source of confusion, because they are not the same; An individual that has one seizure does not necessarily mean he/she has epilepsy, since the seizure might have been provoked for certain reasons (i.e. non-epileptic seizure) and will not occur again [4]. There are different types of epilepsies and each type has its own symptomatology that manifests as a special characteristic movement during seizures. Many epileptic seizures demonstrate motor phenomena, which are abnormal uncoordinated movements such as muscle twitches, myoclonic jerks, and muscle stiffening [5], [6]. Analyzing those movements can provide significant information that help properly diagnosing and assessing the disease prognosis in addition to the treatment efficacy [7]. Seizure semiology is defined as analysis of the clinical manifestations that present during seizures and are related to the activation of the symptomatogenic zone [8]. Thus, seizure semiology analysis is necessary in the evaluation of epilepsy patients. It can provide important information concerning the localizing and lateralizing on the anatomical areas in the brain that are responsible for seizure initiation. For instance, dystonic posturing of one hand is found to be related to contralateral seizure origin in temporal lobe epilepsy [9], while automatisms that are induced in a single side of the body can be assigned to ipsilateral origin [10]. The current clinical practice of epilepsy is based significantly on

seizure analysis by visual inspection [11]. Movement analyses of seizure-induced movements are conducted by visually checking the video recordings in a qualitative way using rating scales, which can be sometimes prone to bias [12].

Computer vision is a field that develops methods for computers to acquire human-like understanding from images and videos recordings. It involves acquiring, processing and analyzing visual data to help understand and automate tasks that human visual system can do [13]. Recent advancements in deep learning techniques have enabled computer vision to achieve remarkable success and better accuracy [14]. Deep learning is a branch of the machine learning techniques that utilizes multiple layers to progressively extract features from the input data [14]. Computer vision has recently become more dependent on deep learning techniques because they have shown better accuracy, outperforming traditional approaches in many tasks, specially the problem of human pose estimation [15]. They have been used to improve the precision of a single and multi-human pose estimation by enhancing its accuracy.

In this thesis, an automatic tool (i.e. pose estimation) was used to obtain 3D kinematic analysis of epileptic seizures from stereo video-recordings. The data samples and the tool are provided by Neuroeventlabs Oy. Seizure events belong to a patient with Lennox–Gastaut syndrome (LGS). They consist of four nocturnal seizure events recorded during sleep at home. Quantitative analysis of epileptic seizures was done for seizure intensity assessment. Visual assessment was done on the seizure events as well as a seizure-free event during REM sleep to assess their severity on a simple rating scale having categorical values (i.e. weighted score) from (0-9). The main goal of the thesis is to conduct a comparative experiment between the human visual assessment (as the gold standard in the field) and the automatic quantitative approach in a control-free environment for to demonstrate whether it is possible to use the current computer vision and deep learning techniques for conducting motion analyses during seizures.

This thesis consists of the following: Chapter 2 consists of three main subchapters: previous relevant work, biological background and technical background. The first part presents a literature review of the methods used for analyzing seizures' severity. Elaborating some of basic principles in neurobiology could give good insight about epilepsy and seizures. In order to understand basic mechanisms about how seizures can be initiated and the development of the state of epilepsy. The second part describes some of the fundamental aspects of neurons, action potentials and synaptic transmission. Additionally, essential terminology about epilepsy and seizures and basic mechanisms underlying them are reviewed. Finally, the third part discusses the technical background

of computer vision techniques and pose estimation that are used in the thesis. It elaborates details in image formation and stereo vision as well as background of the pose estimation tool, used in the work. Chapter 3 entails a description of the used approach. Chapter 4 presents the results and discussion. Chapter 5 presents the conclusions, challenges, and future work.

2. THEORITICAL BACKGROUND

2.1 Relevant work

The traditional evaluation of the impact of surgical and medical treatments on the progression of epilepsy has used seizure counts (frequency) as index of efficacy. It has been noticed that relying solely on numbers of seizures is not sufficient when measuring the effect of treatment and quantifying the seizure severity. Furthermore, antiepileptic drugs (AED) may cause seizures to become less severe and shorter, which is considered a marked improvement in treating the disease, even though the number of seizures is still the same, which is an important detail that could be lost if seizure severity is not considered [16].

Early attempts have been made to develop clinical rating scales to help clinicians assess seizure severity qualitatively [7], [16]–[24]. The traditional method followed in these protocols is a completion of multiple questionnaires. Part of them are completed by the epileptic patient or by his/her family, while others are completed by the medical staff [11]. For example, the Liverpool Seizure Severity Scale was used to measure patients' subjective impression of changes in the severity of their seizures. It originally comprised 16 items, describing two main factors: the first factor is the patient's ability to control his/her seizures, which is influenced by their ability to predict the seizure onset. The second factor is the subjective perception of the severity of seizures (including disturbance in consciousness, duration and severity of postictal confusion, falls as results of seizures, headache after seizures, sleepiness after seizures, tongue biting, incontinence, etc.) [11]. Each of the 16 items was scored on a (1 to 4) point scale. Later, additional items were added to make the scale more comprehensive such as: seizures in sleep only, timing during the day, and seizure clustering vs random seizures [25]. Another example is the Chalfont Seizure Severity Scale, which is observer and patient-based scale, was designed to measure the intensity of seizures, for different seizure types. The questionnaire is completed by patients with epilepsy and their close relatives who observe their seizures. Scoring considers impaired consciousness, automatism, dropping off objects, falling to the ground, incontinence, having an injury, time to return to normal, convulsions and duration of the seizure [16]. Although rating scales are valid measures, this method of evaluation is subjective to the expert's opinion. It is difficult to make comparative studies between different measurements made by different specialized observers or made by the same observer but in different times.

Systems combining video-audio-monitoring with electroencephalography (EEG) are considered the most accurate tools for the diagnosis of seizures and differential epilepsy [26], [27]. They become reliable tools due to their ability to provide simultaneous behavioral and EEG monitoring [28]. In other sense, the success of diagnosing epilepsy depends especially upon evaluating the relevant EEG data , which led to the use of Stereoelectroencephalography (SEEG) (Figure 1).

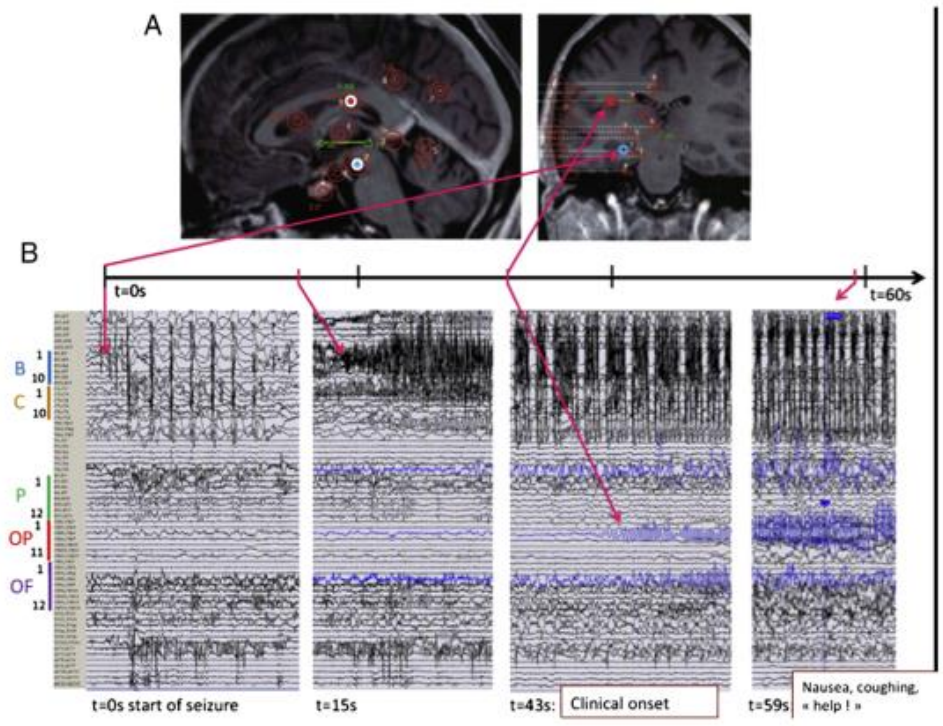

Figure 1. *A recorded temporal seizure using SEEG. The seizure shows as a series of spikes. The electrical discharge is synchronized in 3 areas: (A) amygdala, (B) anterior hippocampus and (C) posterior hippocampus* [29].

Since then, seizure semiology acquired from video-based monitoring beside its relevant electrical events on EEG have become a cornerstone in epilepsy evaluation and seizure understanding [5]. This modern approach allows a detailed look at epileptic seizures to investigate new and deeper features for further analyses, which is particularly useful in the presurgical assessment [1]. Such advancements have dramatically improved our knowledge of epilepsies and epileptic seizure semiology. However, factors such as: short recording times, artificial environments (all V-EEG measurements are conducted in Epilepsy Monitoring Units (EMUs)), uncomfortable conditions due to wearing electrodes for long times and high cost effect have limited its use [6]. Moreover, seizure are analyzed qualitatively by specialists' subjective opinions which makes the inter-observer reliability poor for most semiological characteristics [5].

Most seizures present abnormal, but characteristic, movements. Due to its importance, researchers have made several attempts to assess seizures that can be characterized by motor phenomena. These methods are mainly classified to: video-based methods and noncamera approaches [1].

Noncamera approaches involve using accelerometers (ACMs), gyroscopes [30], magnetometers and other sensors to quantify and classify motion among different epilepsy types [1], [11], [26], [31]–[34]. For example, Accelerometers can be attached to arms or legs to detect changes in velocity and direction during seizures.

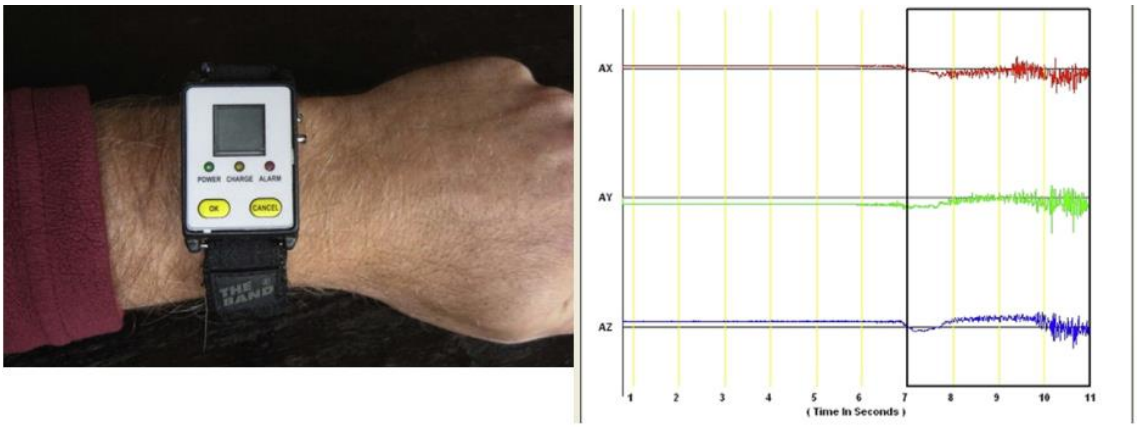

Figure 2. *SmartWatch (left). Motion detection and quantification in 3D during tonic–clonic seizure using SmartWatch (right) [33].*

The system used for recording signals consists mainly of a tri-axial motion/accelerometer sensor, a processor unit, and a battery. One of these contributions is a commercial product called SmartWatch (Smart-Monitor.com) (Figure 2), which is used as a shake detector. When a seizure occurs, the watch sends a wireless signal to electronic device such as: a computer, or smart phone. The signal contains the duration of the movements, time, date of the event, and the motion intensity data in 3D, which can be visualised graphically (Figure 2).

Another example, the work in [34] present visual analysis of signals coming from 3D ACMs attached to patients arms to detect epileptic seizures (Figure 3).

Those 3D accelerometers measure inertial acceleration induced by movements as well as the acceleration by gravity (dynamic and static acceleration). The main challenge for ACM characterization system is to detect seizures from non-seizure movements [26].

Some researchers used magnetometers for detection and quantification of motor seizures. The magnetometer is a device that generally measures magnetism (i.e. strength, direction, or the relative change of a magnetic field at a certain location).

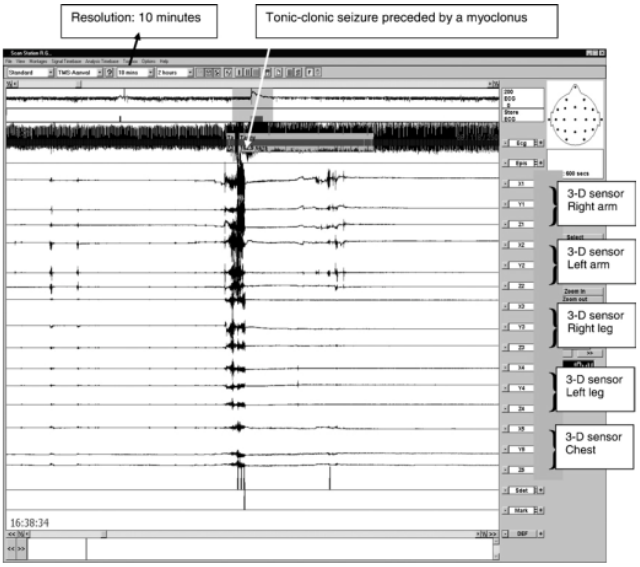

Figure 3. *Examples of a myoclonic seizure that turns to a tonic–clonic seizure. The signal is detected using 3D ACM sensors attached to four body extremities [34].*

It measures the inclination of a body by using a 3D magnetometer to track 3D changes in orientation [26], [35]. Some attempts go for the multimodal approach by combining magnetic sensors to accelerometers in a measurement system referred to as inertial and magnetic sensors (IMS) [36]. In this work [37], three modules, each of them is containing a 3D magnetometer and a 3D ACM and were attached in both wrists and in the head and of patients. The system was able to recognize tonic, clonic, hypermotor seizures, and also non-movement events [37]. Although these systems are promising, inexpensive and their low power consumption enables them to be used in ambulatory monitoring, they need consistent maintenance such as calibration and batteries. In addition, they are susceptible to dislocations during measurements and always the attached devices lead to discomfort to the patients for long term monitoring.

The problem of human motion analysis during seizures using video-based monitoring has been active in the research field of epilepsy for almost 15 years. Several works have attempted to prove that vision-based human motion recognition systems could provide promising possibilities in the field of epilepsy, where seizures often demonstrate abnormal motor behavior. Automated and semi-automated approaches can help

providing objective, quantitative information about of a patient's kinematics during seizures, which can help specialists to diagnose epilepsy accurately. Hence, customizing appropriate treatment. All contributions in this specific research domain (i.e. vision-based systems for movement quantification and analyses of epileptic seizures) can be categorized into two main groups: 2D and 3D approaches. Each approach used marker-based and marker-free systems [38]. Marker-based methods use sensors/markers such as infrared reflective materials attached to the keypoints in the human body to track motion trajectories and velocity. Conversely, marker-free approaches use computer vision techniques to calculate the key points positions over time [38].

Li at his work [39] aimed at conducting analysis of motion trajectories for the human body parts using 2D video recordings by a CCD camera. A system of 22 landmark positions defining the whole body parts was used. Reflective infrared material markers are attached to these landmarks and motion tracking is estimated [40]. The feature extracted is the motion trajectory in two dimensions (Figure 4). Motion trajectory is defined as the path taken by a moving object as a function of time.

Following this work, several studies have attempted to interpret signs from limbs during seizures such as [6], [10]. For instance, this work [6] used a 2D color-based video system to extract features such as displacement and oscillation.

Figure 4. *The estimated motion trajectories of 22 markers attached on the human body. The upper graph (a) shows their relevant motion trajectories versus time in x dimension. (b) the lower graph shows their relevant motion trajectories versus time in y dimension [40].*

Although these studies show ambitious results, they depend on manual intervention by the user to define the region of interest (ROI). Additionally, the used reflectors/markers are sources of discomfort to the patient, especially in the long term monitoring, and

susceptible to instabilities due to the dislocation of the them during severe motor seizures. In addition, the features are extracted in 2D and the depth information is lost.

Latest technological developments in 3D computer vision tracking have made it possible to transition towards 3D systems. To the best of our knowledge, there are only two reported works: [27], [41]. The first work is a marker-based system. It uses a 3D motion capture system for seizure quantification, consisting of 4 infrared motion-tracking SVCams and a set of markers is attached to the keypoints of a patient [41]. The second work is a semi-automatic markerless system [42].

It presents a new low-cost system called NeuroKinect. It adopts the popular Microsoft Kinect system, which is used initially for gaming applications, to quantify seizures in 3D space. This system incorporates RGB-D cameras (color and depth) with the aid of a 3D tracking algorithm based on Horn-Schunk optical flow method [27].

Figure 5. *Motion tracking in 3D using NeuroKinect system. The skeleton estimation has not been completely accurate due to misdetection of the left arm [27].*

Different metrics are computed in 3D such as motion displacement, velocity, acceleration, jerk, and covered distance. Although this seems promising, the ROI tracking can be sometimes inaccurate due to many reasons such as limitations in the used cameras and difficult body poses during seizures with respect to the camera (Figure 5) [1]. Additionally, such systems are still conducted in a full control environment because they are based on Video-EEG monitoring systems that are used in hospitals. Moreover, the system is semi-automatic, meaning that the observer should intervene to assign the region of interests (ROI) that need to be tracked [27].

2.2 Biological background

2.2.1 Neurons

To understand the neurobiology of seizures, it is important to look closely at the controls that are exploited by the nervous system to keep balance between inhibition and excitation mechanisms. In fact, they entail intricate network of different levels ranging from: ions and ion channels, to neurons and synapses, and finally the large neural networks.

The nervous tissue is composed of neurons and glia cells (neuroglia). Neurons play the fundamental role that makes the nervous system able to perform the computations and communications, while neuroglia mostly serve as supporting cells. Neurons or nerve cells are the basis of the nervous system; carrying electrical signals (impulses) that hold communication information about sensations of stimuli, motor responses to these stimuli and inducing cognition in the brain. Though neurons differ from each other structurally, they have common features such as: soma (cell body) and processes (See Figure 7). The cell body contains the nucleus besides other organelles like the other cells in the body. The major difference is it lacks centrioles. What makes neurons special is that they have many extensions rising from the soma called processes. Processes that transmit impulses towards the cell body are called dendrites, while those which carry electrical messages from the cell body to other cells are called axons. A neuron can have multiple number of dendrites but only one axon [42]. The axon, emerging from the soma through a special region called axon hillock, is covered along its length with white and fatty materials referred to as myelin sheaths. Myelin sheaths are separated by gaps called nodes of Ranvier (See Figure 7). Myelin is formed by a special type of glia cells called Oligodendrocytes. This Myelin helps protect and insulate the neurons besides increasing the transmission rate of the nerve signals. The axon forms at its end hundreds to thousands axon terminals. An axon terminal contains hundreds of tiny vesicles containing neurotransmitters. Each of the axon terminals are separated from the other neuron’s dendrite by a gap called a synaptic cleft. The functional junctions that permit the neuron to transmit nerve signals to other neurons are called synapses [43].

Neurons can be structurally classified based on number of processes, which includes both dendrites and axons, into three categories: unipolar, bipolar and multipolar neurons. Unipolar neurons have only one process including the axon on one end and the

dendrite in the other end. They always act as sensory neurons where the dendrite is receiving the sensory information from the stimulus and the axon is transmitting this sensory information to the CNS.

Second type is the bipolar neurons, and this type has two processes emerging from the cell body: the dendrite and an axon. This type is found in the olfactory epithelium and the retina [43]. The third type is the multipolar neurons and those have one axon and two or more dendrites. They are the most common type of neurons (See Figure 6).

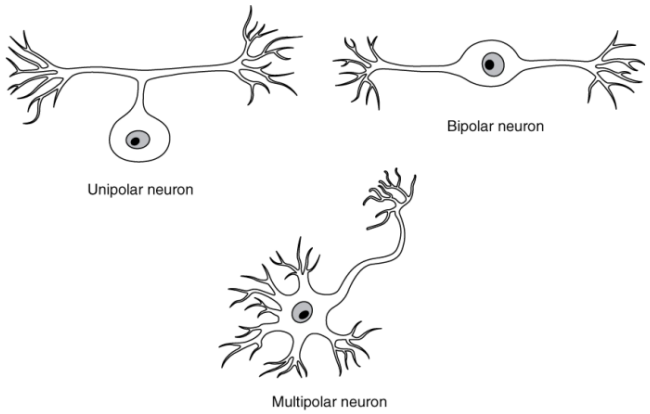

Figure 6. *Neuron classification by structure: Unipolar neuron has one process including the dendrite and the axon, bipolar neuron has two processes emerging from the soma and multipolar neuron has more than two processes [43] P.514*

2.2.2 Actions potentials

The functions of the nervous system depend substantially on the functions of the neurons. And to understand how the nervous system can sense, integrate and respond to stimuli, it is necessary to understand how neurons can generate and communicate nerve signals. It is believed that the action potential is the basis of these processes.

A neuron is inactive in the resting state. In this state, the neuron is polarized, meaning that there are slightly a smaller number of positive ions inside the cell than outside. During this state, potassium (K+) ions are found in bigger numbers inside the cell, whereas sodium (Na+) ions are found in bigger numbers outside the cell.

When a stimulus excites the neuron, it changes the permeability of local patches in the plasma membrane, which in turn, opens the way to sodium ions to diffuse inside the cell changing its polarity until inside becomes more positive than outside, an event called depolarization, and this electrical state is called graded potential.

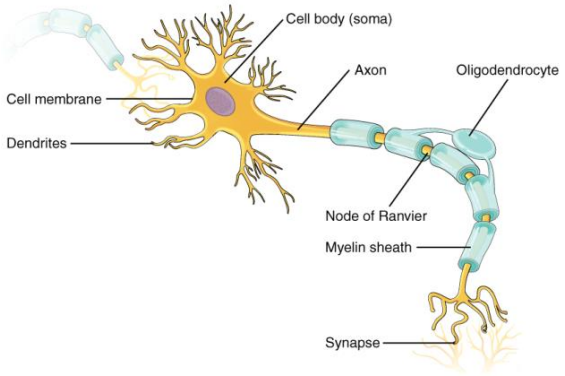

Figure 7. *Major parts of a multipolar neuron from the central nervous system [43]*

If the stimulus is strong enough to surpass the threshold, the neuron is completely activated, and an electrical signal is produced and transmitted to the next neuron.

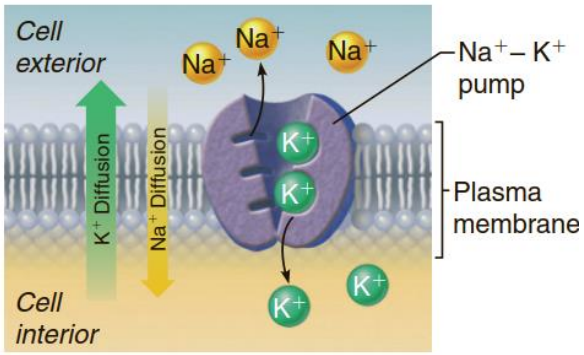

Figure 8. *The sodium-potassium pumps* [42]

This signal is called action potential (nerve impulse) or spike [43]. After that, the membrane permeability to sodium ions changes and becomes impermeable, leading to blocking sodium ions from entering the cell and becoming permeable the potassium ions which diffuse rapidly outside the cell into the interstitial fluid until the polarity is restored again and this even is called repolarization. The repolarization phase goes beyond the resting potential reaching the hyperpolarization phase.

Eventually, after hyperpolarization, sodium-potassium pumps are activated to restore the initial concentrations, by ejecting sodium ions and bringing in potassium ions. These pumps use Adenosine triphosphate (ATP) in this process [43] (See Figure 8, Figure 9).

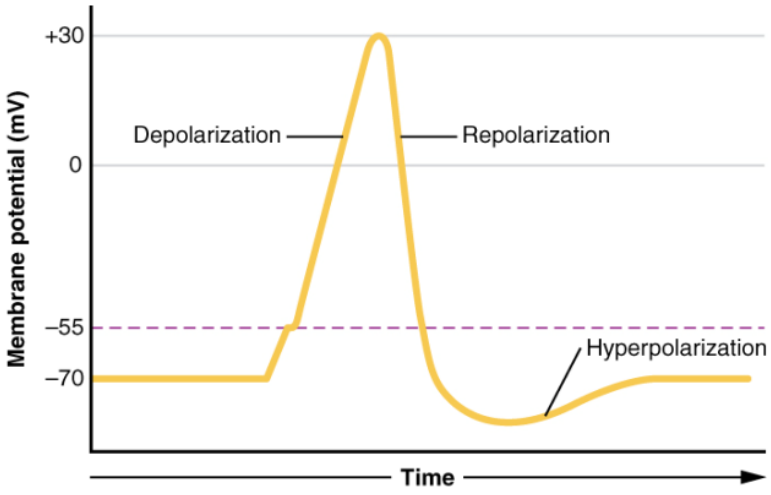

Figure 9. *A diagram of the voltage measured across the cell membrane as versus time, an action potential starts with the depolarization phase, followed by the repolarization phase which decreases below the resting potential (i.e. the hyperpolarization phase), and eventually the transmembrane potential returns to its resting state* [43].

2.2.3 Synaptic transmission

There are many types of synapses in the nervous system. However, they all have common characteristics as follows [43]:

- presynaptic element
- postsynaptic element
- neurotransmitter
- synaptic cleft
- receptor proteins
- neurotransmitter re-uptake or elimination

When an action potential reaches the axon terminal ends, it stimulates voltage-gated calcium channels (Ca^{+2}) there to open. The (Ca^{+2}) ion concentration increases which helps the tiny vesicles containing neurotransmitters merge with the presynaptic membrane. This in turn releases the neurotransmitters into the gap between the two neurons called a synaptic cleft. Those neurotransmitters bind to specific receptors in the

membrane of the target neuron (postsynaptic membrane). If the postsynaptic potential is strong enough, it will cause a graded potential in this neuron. If the summation of the total postsynaptic potentials at the dendrites of this neuron surpasses a certain threshold, an initiated action potential will be produced in the receiving neuron and the whole process repeats again.

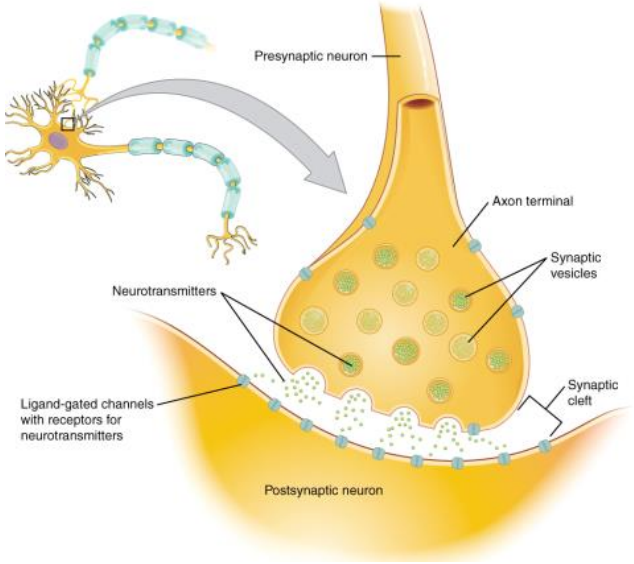

Figure 10. *The synapse is a junction that connects between two neurons. When action potential reaches the axon terminal, it stimulates Ca2+ ions to enter the bulb which causes the fusion of vesicles and the release of neurotransmitters through the synaptic cleft, then they bind to their type-specific receptors in the postsynaptic membrane causing the postsynaptic potentials. Finally the neurotransmitters are removed either by enzymatic degradation, glial reuptake or neuronal reuptake* [43]*, P.535.*

The released neurotransmitters are soon removed from the synaptic clef either enzymatic degradation, glial reuptake or neuronal reuptake [43]. A postsynaptic potential (PSP) is defined as the graded voltage potential in the dendrites of the target neuron which is receiving action potentials from other neurons. Postsynaptic potentials can be either excitatory or inhibitory potentials. Excitatory potentials cause depolarization in a postsynaptic potential (EPSP). While inhibitory potentials cause hyperpolarization in a postsynaptic potential (IPSP) [43]. These PSP potentials will be summated (temporally or spatially or both) in the membrane and the net result will determine whether an action potential is initiated if the membrane potential reaches the threshold.

2.2.4 Neurotransmitter systems

Neurotransmitters, as previously mentioned, are the substances that are released by the presynaptic neuron's end terminal at a synapse and subsequently bind to their specific postsynaptic receptors in the target cell (or cells) to induce specific response. This target can be another neural cell or organs such as: glands and muscles. There are also groups that refer to the neurotransmitters, and inside the groups are specific systems of neurotransmitters.

The first group is the cholinergic system. This system is based on acetylcholine. It is involved in the regulation of attention and higher cognitive processing [44] as well as memory [44]. it includes, for instance, the Neuromuscular junction (NMJ). They are found in the autonomic nervous system, and throughout the brain [43]. It has two types of receptors: nicotinic and muscarinic receptors.

Another example is amino acids. It includes glutamate (Glu), GABA (gamma-aminobutyric acid), and glycine (Gly). Their chemical structure has an amino group and a carboxyl group. Each amino acid neurotransmitter is part of its own system: Glutamate is part of the glutamatergic system, GABA is part of the GABAergic system, and glycine (Gly) is part in the glycinergic system. These amino acids are removed from the synapse by re-uptake by a pump in the cell plasma membrane of the presynaptic neuron, which removes the amino acid from the synaptic cleft to be recycled and repackaged in their vesicles, for future use [43].

Another example is the biogenic amine. These neurotransmitters are made of amino acids. The difference between them and the previous amino acids neurotransmitters is that they do not have carboxyl groups such as Serotonin. Other biogenic amines, such as dopamine, epinephrine, and norepinephrine, are made from tyrosine. For instance, the dopaminergic system has dopamine and dopamine receptors.

A note worth considering is that the receptor proteins determine the effect of the neurotransmitter totally. If there is no receptor protein in the postsynaptic element, the neurotransmitter will have no effect. Protein receptors also decide the depolarization and repolarization effect. For instance, several variants of the muscarinic receptor when they bind to acetylcholine, they can have depolarization or hyperpolarization effects accordingly [43]. Glutamate is considered the main excitatory neurotransmitter in the brain because the Glu receptors have depolarization effect on the postsynaptic cells. Consequently, it plays an important role in the initiation and spread of seizure activity in addition to epileptogenesis [45]. Glycine and GABA are considered inhibitory amino acids,

because their receptors play inhibitory role in the hyperpolarization of postsynaptic cells. This is important to consider when designing drugs for epilepsy.

2.2.5 Epilepsy Neurobiology of seizures

Epilepsy is a very complex neurological disorder [4]. It is not a specific disease, nor even a single certain syndrome, but rather a wide set of complexes that arise from diverse pathologic processes of any number of disordered brain functions that themselves might be secondary of another wide variety of pathologic processes [46]. It is defined by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) as: “a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition. The definition of epilepsy requires the occurrence of at least one epileptic seizure.” [47].

Some epilepsies show mixed seizure types such as Lennox-Gastaut Syndrome (LGS) in which tonic seizure and myoclonic jerks are the major components [48], [49]. Patients with this type of epilepsy typically have extremely frequent seizures. Those seizures can take the form of tonic, tonic-clonic, myoclonic, atypical absences [48].

There are diverse mechanisms that can induce seizures. The fact that they are multiple is logical given that the different ways the nervous system normally use to control the firing mechanisms. The nervous system has controls that prevent neurons from excessive discharging, and also it has controls that facilitate neuronal action potential discharges so that the nervous system could function flawlessly. Normally, there is balance between both types of mechanisms. The main principle of seizures is the disruption in the balance between excitation and inhibition mechanisms. Disturbing inhibition mechanisms and stimulating excitation mechanisms could lead to seizures and in contrast, promoting inhibition mechanisms and disturbing excitation mechanisms can control seizures.

On the cellular and the subcellular levels, there are many factors that can disturb this balance. As previously discussed in section 2.2.2, one of the ways to control the excitability is maintaining certain concentrations of the ionic ingredients. For example, in normal conditions, potassium (K+) ions are found in bigger numbers inside the cell, whereas sodium (Na+) ions are found in bigger numbers outside the cell. This leads to a net transmembrane potential = -70 mV (Figure 9). If this balance is perturbed (i.e. Potassium ions concentration is increased in the extracellular space), this can lead to a vicious cycle of excessive discharges which induce seizures [50]. Sodium-potassium

pumps have the major responsibility to restore ion concentrations to initial state after the action potential. Abnormalities in theses pumps could raise the possibility to induce seizures [51]. Glia cells have also important role in controlling ion concentrations and maintaining normal transmembrane potentials. They play important role in regulating the synaptic transmission. Any disruption to those cells could lead to seizures [52]. Sodium channels and potassium channels play a significant role in regulating action potentials (see section 2.2.2). Defects in sodium channels for instance, can decrease the threshold at which a neuron produces action potentials. Therefore, malfunctioning in these channels can suggest mechanisms of seizures. It has been shown that genetic mutations in some genes which are responsible for regulating the subunits of voltage-dependent sodium channel have impact on seizure initiation [53]. Synaptic transmission has been shown to play important role in maintaining the balance between excitation and inhibition and thus, defects in every component in the synaptic transmission system could lead to seizures [4].

It is important to notice that excessive discharge in neurons does not alone cause seizures. Synchronization in the networks which involve those neurons is pivotal in inducing seizures [4]. There are multiple factors that can cause synchronization which is not found (i.e. synchronization between neurons) in normal neuronal activity. For instance, research has shown that glutamatergic interconnections mechanisms can lead to seizures. Since gap junctions provide low resistance current between the coupled neurons, defects in gap functions can also facilitate synchronicity between the discharged neurons [54]. Since the brain with epilepsy is structurally different from a normal brain, it has also been shown that the altered brain of epileptic patients is susceptible to synchronizations more than a normal healthy brain [4].

2.2.6 Epileptic seizures and epileptic seizure types

Epileptic seizures represent the major symptomatology of epileptic patients. Hence, the aim of any treatment for epilepsy is to control them. They can be characterized by several symptoms. The proper identification of these symptoms serves well the improvement of seizure syndromes classification. In addition, seizure semiology (i.e. types of seizures) can help distinction between epileptic seizures and non-epileptic seizures [5].

It is useful to consider some definitions that relate to seizures before diving in epileptic seizure types. The word “seizure” comes from the Greek language meaning to take hold [47]. A seizure is defined as a period of abnormal excitations in nerve cells that happen synchronously. It typically takes seconds to several minutes and can be prolonged in special cases [4].

The word “semiology” refers to the branch of linguistics that is concerned with signs and symptoms [55]. Seizure semiology is defined as “the manifestation of the activation of the symptomatogenic zone that might indicate that it is a result of ictal spread from a more distant epileptogenic zone”. In other sense, seizures initiated from different epileptogenic zones can activate the same zone. Additionally, seizure arising from the same epileptogenic zone can activate different symptomatogenic zones, which in turn, produce different seizure semiologies [8].

An epileptogenic lesion is defined as an area of damaged tissue in the brain that induces epileptic seizures [56]. The lesion itself might not contain active neural tissues [46]. Epilepsy may arise from diverse brain lesions—for example, head injuries and brain tumors [56].

The epileptogenic zone is defined as the area located adjacent to the lesion and substantial for seizure genesis and removing it is generally considered essential for improving seizure control [57].

The time period during the seizure is called the ictus or ictal period. The time period which occurs immediately after a seizure is the postictal period. The intervals between seizures are the interictal periods [58]. Focal seizures (also called partial seizures and localized seizures) begin in one area of the brain, they can be partial or complex seizures [58]. The crucial difference between simple and complex focal seizures is that consciousness is disturbed in the complex seizures but not in the simple seizures [58].

Four main categories are involved in the ictal phenomena of the seizure in general. They are: motor sphere, consciousness sphere, sensorial sphere and autonomic sphere [5]. These spheres are involved in most of the seizures simultaneously [5]. Seizure semiology can be categorized into auras, autonomic seizures, dialeptic seizures, motor seizures, special seizures and paroxysmal events [5]. In this thesis we focus on the seizures that manifest motor behavior following the classification mentioned in this work [5]. Motor seizures are divided into two main categories: simplex and complex motor seizures.

Simple motor seizures are simple movements. They can be divided into: tonic seizures, tonic–clonic seizures, myoclonic seizures, clonic seizures, versive seizures, epileptic spasms.

Myoclonic seizures consist of irregular shock-like movements. They occur in a non-rhythmical way. They often are brief jerks (small duration). There can show in both partial and generalized seizures. In partial seizures, jerks affect one side. In generalized seizures, jerk movements affect shoulders, and arms [5]. Lennox–Gastaut Syndrome

(LGS) also shows more often generalized myoclonic seizures [59]. The primary motor cortex is the area involved in inducing these seizures [5].

Tonic seizures are mainly characterized by muscle stiffening. In normal humans, the muscle tone is the muscle normal tension at rest. In tonic seizures, the tone increases greatly, and it takes 3 seconds in average. Tonic seizures can affect the whole body, arms and legs. They often show contralaterally producing asymmetric posture [5]. Generalized tonic seizures involving disturbance in awareness are commonly involved in Lennox–Gastaut syndrome [59]. Motor cortical areas in the brain are most involved in inducing these type of seizures [5].

Clonic seizures are short, rhythmical and repeated jerks of different muscle groups. They affect the whole body or parts such as hands, legs and the face. The primary motor cortex epileptic activation is the main driver of these seizures. Clonic seizures can be unilateral seizures, which are present in the focal epilepsy. They are also present in frontal lobe epilepsy in the early stage of the seizure evolution [5].

Epileptic spasms are relatively age specific epileptic seizures that may occur in the period between 3 to 12 months age. They involve sudden flexions, extensions or mixed flexion-extension of truncal and proximal muscles and last between 1-2 seconds (longer than myoclonic jerks) [60].

Versive seizures are mainly characterized by the movement of the eyes and the head to one direction. They can involve clonic jerks as well [5].

Tonic-clonic seizures are mainly characterized by generalized tonic phase followed by clonic phase. These seizures evolve as follows: they start occurring with a typical tonic pose. This phase takes several seconds and then evolves into a the clonic phase [5]. The clonic phase is myoclonic jerks in the joints. The whole seizure might take 1 to 2 minutes [5]. These types of seizures involve always impairment in consciousness in the onset of the tonic phase. Generalized tonic-clonic seizures are also followed, in most of cases by a prolonged postictal coma [5].

Hypermotor seizures Hypermotor seizures are complex movements such as the involvement of the body axis or proximal limb segments (such as bicycling or kicking) and result in large amplitudes that appear violent when they are performed rapidly [5], [61]. They last for less than a minute and found frequently in patients have epilepsies affecting the supplementary sensorimotor area and the mesial frontal such as the frontal lobe epilepsy [5], [61].

Automotor seizures are complex motor seizures in which oral and manual automatisms are manifested. These automatisms include the hands and feet while the oral automatisms include the mouth and tongue. Typical examples are: chewing, lips smacking, swallowing and hand automatisms. They are most related to temporal lobe epilepsy.

Gelastic seizures are seizures mainly characterized with “laughing’’ [62], hence the name. They are commonly found in patients with hypothalamic hamartoma [62].

In summary, motor seizures consist of simple seizures (involve simple movements) and complex seizures (involve complex patterns of movements). Simple seizures are different types such as: tonic seizures, tonic–clonic seizures, myoclonic seizures, clonic seizures, versive seizures and epileptic spasms. While complex motor seizures are: hypermotor seizures, automotor seizures and gelastic seizures.

2.3 Technical background

2.3.1 Image formation and perspective transform

The process of image formation in a camera consists of a projection of an object in the 3-dimensional world on a 2-dimensional plane. Therefore, the depth information is lost, and there is no means to tell if this object is big and distant or small and close. This transformation from 3D to 2D dimensions is called perspective projection. In computer vision, it is common to use the central projection model shown in Figure 11. In this model, the rays are coming out of the origin of the camera frame {C}, the object's image is projected onto the camera plane in front of the camera frame (non-inverted image). For a point P= (X, Y, Z) and its projection p= (x, y) from the (Figure 11) we can deduce the following [63]:

$$x = f\frac{X}{Z}\ , \qquad y = f\frac{Y}{Z} \tag{1}$$

Where f is the focal length of the camera.

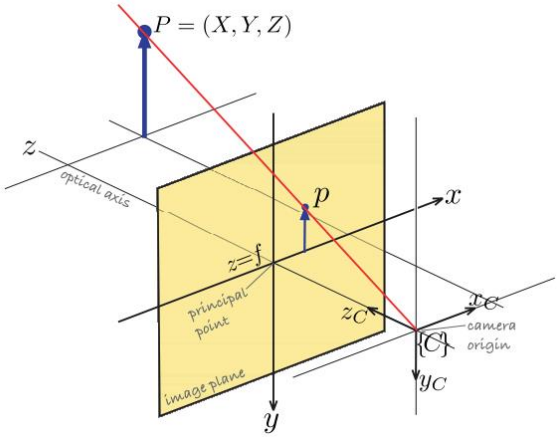

Figure 11. *A diagram of the central projection model* [63] *P. 252.*

In a digital camera has the image plane described by a grid with dimensions Width × Height photosites that directly correspond to pixels. A pixel coordinate can be represented as a 2-element vector (u, v)

$$u = \frac{x}{\rho_w} + x_p\ , \qquad v = \frac{y}{\rho_h} + y_p \tag{2}$$

Where ρ_w, ρ_h are width and height of each pixel respectively. (x_p, y_p) is the coordinates of the principal point (the principal point is the point where the optical axis of the camera intersects with the image plane) [63].

2.3.2 Stereo vision

Stereo vision is the acquisition of 3D information from 2D digital images taken by a pair of cameras. By comparing information about some scene from two known points, the 3D information can be estimated by investigating the corresponding positions of objects in the two perspectives. This approach is similar to the biological process of the human eyes (stereopsis) [64].

A simple model of stereo imaging is shown in (Figure 12) is called epipolar geometry. At any point in the scene, visible to the pair of cameras c_0 and c_1 , will be projected to a pair of image points on the two image planes, called a **conjugate pair**. The displacement between the positions of the two corresponding points (i.e. conjugate pair) is called the **disparity** d. The whole plane comprising the point p and the two camera centers c_0 and c_1 is called the **epipolar plane** [64]. The line e_1x_1 , which is the intersection between the image plane of the camera c_1 and the epipolar plane, is called epipolar line. The distance between the two camera centers is called **baseline** b.

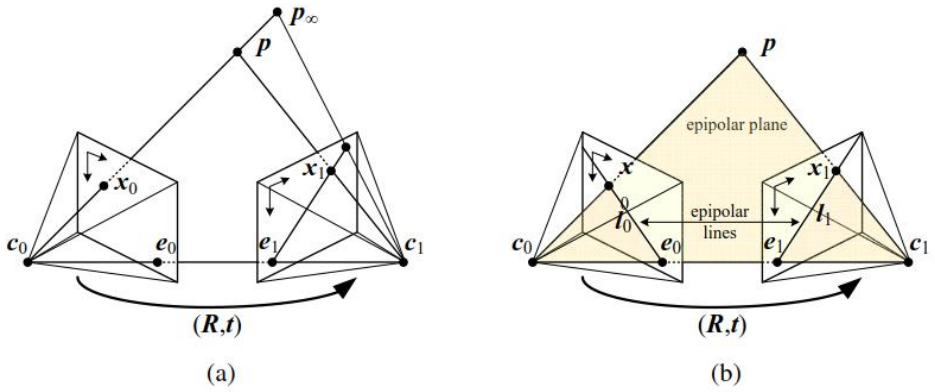

Figure 12. *A simple model of Epipolar geometry: (a) a projection of point P on the left camera plane as a point x_0 and the corresponding epipolar line on the right camera plane. (b) Epipolar lines and epipolar plane* [64] *P.471*

In stereo matching, the relative position of the two cameras and the calibration data for the cameras can be known. Therefore, a projected point on one of the image planes must have its correspondent in the relative epipolar line of the other image plane, and this constraint is called **epipolar constraint**. Although this dramatically reduces the possibility of matching the correspondent points from a whole two-dimensional image plane to a one-dimensional epipolar line, there is still large number of potential correspondences.

One way of making the matching process easier, is to rectify the images taken by the pair imagers. **Rectification** means ”applying a pair of two-dimensional projective transforms, to a pair of images of a known epipolar geometry so that epipolar lines in the original images map to horizontally aligned lines in the transformed images” [65].

The resulting rectified geometry helps hugely simplifying the problem of stereo matching, and leads to a simplified inverse relationship between depths Z and disparities d as follows:

$$d = f\frac{b}{Z} \tag{3}$$

$$x' = x + d\,, y' = y \tag{4}$$

$$X = (x + x_p)\frac{Z}{f}, \qquad Y = (y + y_p)\frac{Z}{f} \tag{5}$$

Where d is the disparity, f is the focal length of the camera (measured in pixels), b is the baseline between the pair imagers, the point (x, y) is the original projected point in the first image plane (in pixels), the point (x', y') is the estimated correspondent point in the second image plane (in pixels) [64], (x_p, y_p) is the principal point and (X, Y, Z) is the coordinates of the point in the real word . The method of determining the depth based on the disparity is called **triangulation**. After rectification, the similarities of pixels at corresponding locations can be computed and stored as disparity space image (DSI) for further processing.

The main problem now is how to detect the mutual pairs in stereo images. This is called the correspondence problem. To solve the correspondence problem, two main approaches are used: sparse and dense stereo. Sparse stereo algorithms are feature-based algorithms, they recover the 3D world coordinates (X, Y, Z) for each corresponding point pair. On the other hand, dense stereo recovers the 3D world coordinates (X, Y, Z) for every pixel in the image [63]. Discussing each approach in detail is out of this thesis scope. However, basics underlying the stereo matching will be present.

Although there are a large set of stereo algorithms, they all follow common steps. These steps are mentioned as follows [64]:

- Matching cost computation
- Cost aggregation
- Disparity computation and optimization

- Disparity refinement

All stereo matching algorithms compute a cost criterion to quantitively measure the degree of matching between two pixels. In the matching cost computation stage, it is determined whether the values of two pixels in the two images are related to the same point in the scene. The matching cost is computed at each pixel for all considered disparities [66]. In case rectification has been done, the matching cost can be calculated horizontally in one dimension.

Cost aggregation: in this stage the matching costs over local regions are aggregated [67]. The disparity map is initialized using one of the local or global algorithms. Then the map is refined in the disparity refinement stage.

Stereo matching algorithms consists of global and local algorithms. The local approach applies restrictions over a small number of pixels around the pixel under study. Conversely, the global approach makes smoothness assumptions about the scene's depth of field, and then solve a global optimization problem [64].

2.3.3 Pose Estimation

In the field of Computer Vision, human pose estimation is the task of estimating the location of human skeleton, typically consisting of joints, arms, hands knees, shoulders, elbows and ankles, in an image or a video.

This topic has become of a great importance due to its enormous applications in various fields. Many works have been done to tackle the problem of human pose estimation. They can be divided into three main categories: generative, discriminative and hybrid models.

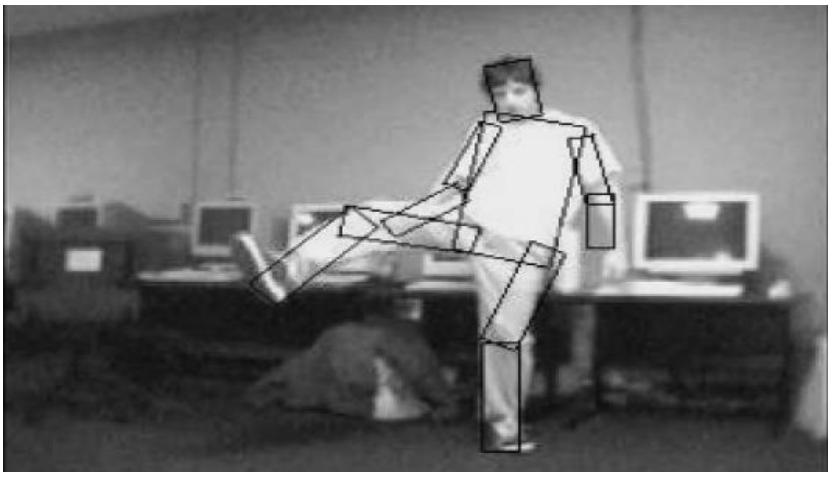

Figure 13. *An example of how body parts are partitioned into smaller parts using the pictorial structure method* [68].

Generative approaches (also referred to as model-based and top-down approaches) are mainly based on building a model of human body using annotated images. This model inherits certain degrees of freedom (parameters) which can model different configurations of the human body the way it is viewed by the camera. Then, the generated model is compared to the given image. The best match is inferred through continuous update of minimizing the error between the instance model and the image.

Generative methods can be divided into holistic models and part-based methods. Holistic methods model the human body as a whole while part-based methods model the human body through its body parts. One of the most important contributions that is considered part-based methods is the models based on the pictorial structures [68]. It is worth considering that pictorial structures were first introduced in 1973 as a way to represent objects in images taken by cameras by identifying the parts of the object in the image and their connections with their neighbors [69]. Pictorial structures' main idea is to describe the human body as a set of rigid templates, each of them describes certain

part in the human body taking the form of tree structures and undirected graphical models (Figure 13).

Discriminative models (model-free methods) are contrary to the generative methods in a sense that they start defining the human pose from the image and learn a direct mapping function between the image features and the human pose. The learning procedure is done in a supervised manner using annotate data and the mapping phase can be either by a classification or regression problem. Examples of regression-based methods can be found in this work [70] and [70]. Classification-based methods learn mapping function by maximizing the score of the classification [71].

Hybrid models combine between both approaches. For example, the likelihood which some object is present at a given location and appearing in a specific appearance can be acquired using a generative approach, and then refined by a discriminative approach using a learning process. An example of this approach is the work proposed by Salzmann and Urtasun [72]. In this work, a combination of ideas from the two approaches are utilized to introduce distance constraints to the human pose estimated by the discriminative methods.

Recent advancements in deep learning have revolutionized the field of computer vision. The incorporation of the latest and efficient state-of- art computing techniques such as Convolutional Neural Networks (CNNs) in solving computer vision problems was natural, human pose was certainly one of these problems. Human pose estimation using deep learning algorithms can be categorized under discriminative approaches, since they learn a feature representation to directly infer the body pose joints. Although, researchers have attempted to distinguish between discriminative methods and deep learning approaches due to the significant impact of the latter in the field.

The pose estimation tool used in this work is based on **Cascaded Pyramid Network** (CPN) [15]. It was developed to estimate multiple human poses. Generally, there are two main approaches for multiple human pose estimation: top-down and bottom-up approaches. **Bottom-up** models aim at directly estimating the keypoints in the image, and then attribute them to their full poses. One example that follows this approach is DeepCut [73]. It deals with the problem of distinguishing different human poses in an image as an Integer Linear Program (ILP) problem and partition part detection into person clusters. The final estimation of poses is obtained by combining person clusters with labeled body parts. DeeperCut [74] is the improved version of DeepCut and it utilizes deeper ResNet [75] which consists of tens or even hundreds of stacked Residual Modules. The OpenPose [76] uses part affinity fields (PAFs) to map the relationship

between different keypoints and then combine them into full human poses. Another example is [77], where the problem is solved by producing score maps and associative embedding to assign different keypoints to their relevant human bodies in the image. In contrast, **top-down** approaches aim at locating and cropping all the persons in the image first in bounding boxes, then applying pose estimation techniques to solve the single pose estimation problem on each of the cropped persons [78] [79] [80]. This approach (i.e. top-down approach) consists of two parts: a human detector and a single pose estimator. The human detector is used to locate different humans in an image while a single person pose estimation is used to locate keypoints of each human body. We review in more detail some of the previous attempts to solve the two problems individually.

Figure 14. *Output samples of the pose estimation. It constructs heatmaps to define features of the keypoints. (Images from left to right are: the estimated final pose, a heatmap of neck, a heatmap of left elbow, a heatmap of left wrist and a heatmap of right knee)* [81].

Single pose estimation: one of the first contributions in solving the single pose estimation problem using CNNs is made by Toshev by introducing DeepPose [82]. DeepPose is simply a deep network using a cascade of CNNs of pose regressors. In other sense, this approach is based on solving a regression problem, and the learning process is done through a cascade of CNNs. Following that, Tompson attempts to solve the problem by replacing the regression problem with the prediction of heatmaps using Convolutional Networks (ConvNets) and graphical models and graphical models [83]. A heatmap is a 2D map in which each position represents a score. This score represents the likelihood of the spatial location of a certain keypoint (Figure 14). In fact, each heatmap represents a gaussian distribution of the possibility that a certain keypoint to be located in a specific location, instead of trying to estimate the exact position of a keypoint. Subsequent work done by Newell, Yang, and Deng introduced a new term called “Hourglass” network [81].

A hourglass network consists of blocks of convolutional and max pooling layers that downsample the input image to very low resolution and then the network begins up-sampling and combination of features across all different scales [81]. This bottom-up and top-down processes is repeated several times to build up a “stacked hourglass” network to generate the prediction. Wei proposes another approach by using a multi-stage architecture; first generate coarse results, and then refine the results in the following stages continuously [84]. Other approach minimizes the error to get the pose estimation and then gradually enhance the pose estimation [85].

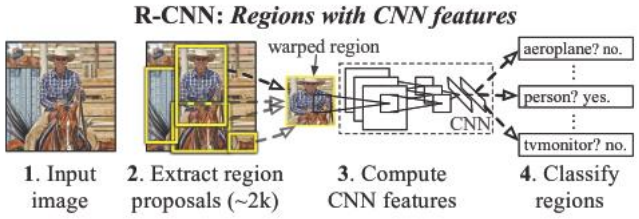

Figure 15. *The R-CNN network architecture. (1) The RCNN takes an input image, (2) then extracts 2K region proposals , (3) after that, it computes features for each of the regions by using a large (CNN), and in the final stage (4) it classifies each region using linear SVMs* [86].

Human detector: human detection algorithms sort of object detection algorithms, and they are mainly based on the R-CNN family [78], [86], [87]. Instead of trying to classify a huge number of regions in the image, Girshick proposed, in his work, a method for selecting 2000 regions in the image (called region proposals) by using the selective search algorithm, and computing features for each region proposal using a large (CNN) [86] (Figure 15).

Recent up-to-date object detectors such as [88] and [78] are composed of two stages: the first step is to generate box proposals and the second step is to crop from the feature map and use the R-CNN network to refine the box proposals to get the final boxes.

The **tool used in this thesis** belongs to the top-down approaches [15]. This means that the system uses a human detector to generate bounding-boxes around persons in the image and then estimation of location of the keypoints for each person using a single-human pose estimator. For the human detector, the tool adopts the approach of Feature Pyramid Network (FPN) for object detection [88].

Figure 16. *Estimated poses using the CPN-based pose estimation tool*

In order to train the human detector, all eighty categories from COCO dataset [89] are used for only human category. Concerning the single human pose estimation, a new approach called Cascaded Pyramid Network (CPN) is used. CPN involves two stages: GlobalNet and RefineNet [15]. GlobalNet is a feature pyramid network that can locate simple keypoints such as eyes and hands. On the other hand, RefineNet tries to locate the hard keypoints that are difficult to localize in the first stage. Figure 16 shows some results of images where CPN-based pose estimation tool is used [15].

3. MATERIALS AND METHODS

3.1 Dataset collection

This thesis relied on video recordings from a database collected by Neuroeventlabs Oy. There were some constraints that dictated the selected videos for movement analysis. First of all, the data were licensed to be used in academic research. Second, the video recordings were selected so that the events were blanket-free (i.e. the patient was not covered with blanket when he had seizures). The reason for that was to minimize external occlusion as much as possible. In addition, recordings must be stereoscopic. One single patient with epilepsy (specifically LGS) has fulfilled these conditions. He has undergone nocturnal home monitoring. 4 events were randomly selected to be processed. They are myoclonic seizures. One event representing the patient condition during REM sleep is added (i.e. non-movement event). The four seizure events are named according to their chronological appearance as: E1, E2, E3 and E4. The REM sleep event is named “REM”.

3.2 Stereo system and data preprocessing

A special camera was used to obtain videos with in-depth information. For the measurements, we have used (Intel Real Sense D435) (Figure 17).

Figure 17. *The stereoscopic camera Intel RealSense D435 [90].*

The stereoscopic camera consists of left and right camera, looking at the same object from different perspectives similar to the human eyes. Each of them produce an image with the resolution of 1280x720 pixels at frame rate of 30 FPS which is sufficient for capturing fast oscillations [90]. The stereoscopic image is stored in a top/bottom frame packing. The baseline of the camera after calibration, which describes the distance between the two imagers, is 0.0499433 cm. The focal length is 643.154 pixels.

The recorded data are rectified in the camera, in its Vision Processor D4. The rectified data are then passed through USB cable to the host for further processing. After that, videos were then pre-processed using FFMPEG framework [91] for contrast and gamma enhancement and cropping. The pre-processed videos were further processed by pose estimation to extract the skeleton keypoints. To extract the depth information, StereoSGBM algorithm was for stereo matching from OpenCV [92]. The depth dimension is then estimated for each keypoint per frame. Data cleaning procedure was applied on the output json files of the pose estimation using the library: "Pandas" in Python. The depth dimension, accompanied with the co-ordinates of each keypoints produced by the pose estimation, are then used to extract the real dimensions (X, Y, Z).

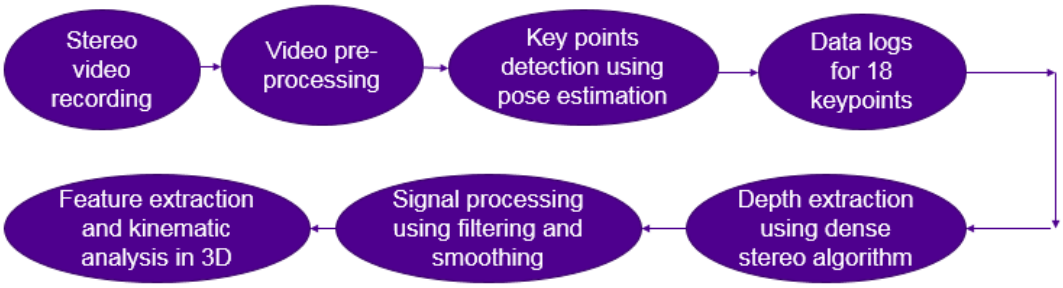

Figure 18. *Schematic drawing of the framework for the thesis*

The displacement sequences, extracted from json files, were then used for computing the first and second derivatives to calculate keypoints' velocities and accelerations. Since the data are noisy, we use an automatic algorithm for smoothing and estimation of the signals and their derivatives [93]. The signal is modelled as "a stationary double-integrated Wiener process" and derivatives are smoothed using Kalman smoother [93], [94]. The algorithm is implemented in MATLAB and freely available [93]. For further smoothing and emphases of strong and long movements, a moving average filter is used. The window size = 8 is used on the final results for further smoothing the output results (equation 6).

$$O[i] = \frac{1}{M}\sum_{j=0}^{M-1} I[i+j] \tag{7}$$

Where $I[]$, $O[]$ are the input signal from the pose estimation and the output signal respectively. M is the window size of the filter. Figure 18 summarizes the proposed system in this work.

3.3 Pose estimation

A modified version of a multi-person human pose estimation was used for conducting the patient's pose estimation during each seizure event. It is open source computer vision software that uses deep convolutional neural network (CNN) architecture to predict the location of the keypoints [15]. It uses a tensorflow implementation of human pose estimation based on Cascaded Pyramid Network (CPN).


```
{
   "version": 1.1,
   "people":[
      {
         "pose_key-
points":[760.9,22.1,0.678,698.1,47.2,0.564,679.3,22.1,0.600,688.7,12
5.7,0.594,751.5,125.7,0.394,717.0,72.3,0.528,760.9,160.3,0.472,792.3
,72.3,0.556,541.2,147.7,0.621,660.5,311.0,0.690,560.0,499.5,0.766,62
2.8,160.3,0.454,710.7,311.0,0.348,578.9,512.0,0.293,751.5,19.0,0.614
,760.9,22.1,0.624,713.8,12.7,0.602,726.4,37.8,0.51]
      }
   ]
}
```


Program 1. *A sample of data logs output from the pose estimation representing 18 key points, number of detected people=1, for each keypoint 3 values are produced. They represent, in order, x-coordinate, y-coordinate and a number indicating reliability values > 0.5 are reliably estimated values while values < 0.5 are not reliable due to occlusions and invisibility.*

The human pose estimation is done in two phases. The first phase uses Tensorflow implementation of Faster R-CNN to locate the bounding boxes of human body in a picture. The second phase, 18 key points are located from the detected human body per frame in json files (see Program 1).

Keypoints are markers of the human joints locations this system can predict 18 key-points (Figure 19).

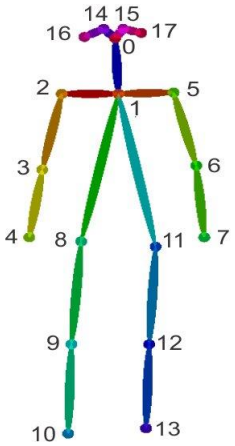

Figure 19. *Pose estimation skeletonizes the human body into 18 key* [95].

We divided those 18 keypoints into two categories: primary and secondary keypoints. Primary keypoints are directly involved in the seizure movement and they are: right wrist, left wrist, right ankle and left ankle. Secondary keypoints are the rest. We have analyzed the 4 primary keypoints, representing the upper and lower extremities, leaving the rest for future research.

3.4 Visual video assessments

As previously mentioned in 2.1, the current evaluation of epilepsies that impair the motor function is basically relying on the visual observation of the patient's motor behavior, directly or through video monitoring. This approach could enhance the disease assessment by the automatic acquisition 3D information. Additionally, it also allows to find the keypoints of the patient for further motion analysis and more features extraction based on the acquired 3D data. A comparative study has been made in this work. Four events belonging to a patient with epilepsy have been studied, processed and analyzed. There are two main streams for this study. The seizure events were re-assessed by two epileptologists (PO, SP) to provide a clinical benchmarking for this comparative work. They were asked separately to evaluate the intensity of each seizure on a rating scale from 0-9. The events were assessed based on the motor behavior in addition to disturbance in consciousness. They assessed the events individually to avoid any source

of bias). The goal is to compare between the automatic quantitative approach and the qualitative approach.

3.5 Features extraction for motion analysis

Features can generally be defined as the measurable properties of the observed phenomenon. Since the studied phenomenon here is the motor activity during epileptic seizures, the selected features for each keypoint are obtained in 3D and they are: the motion trajectory (position over time) , velocity, acceleration, displacement travelled (Euclidean distance in 3D space), spike count SC (number of spikes above certain threshold), spike intensity SI (number of spikes per seconds), weighted spike ratio WSR (proportion of number of spikes for a keypoint to the total number of spikes of the 4 keypoints). A note worth considering is that the features extracted in this thesis are basically in the time domain, other features can be extracted in further studies.

The programming languages used for this work are: Python for data wrangling and cleaning using Pandas library, C++ for computer vision algorithms using OpenCV library and MATLAB (MathWorks, Inc.) for signal processing and data visualization.

4. RESULTS AND DISCUSSION

4.1 Signal processing of seizure and REM events

The stereo matching algorithm, used for depth estimation, produces uncertainty in the depth map. Additionally, the pose estimation tool processes each frame separately, therefore the estimated keypoint locations are unstable. In addition, occlusion and invisibility of the keypoints, affects the accuracy of keypoint detection, which can be sometimes even impossible if the selected keypoint is not clear. All these factors combined introduces noise to the signals. Therefore, the applied signal processing technique discussed in section helps to a great extent minimize the noise (see Figure 20, Figure 21).

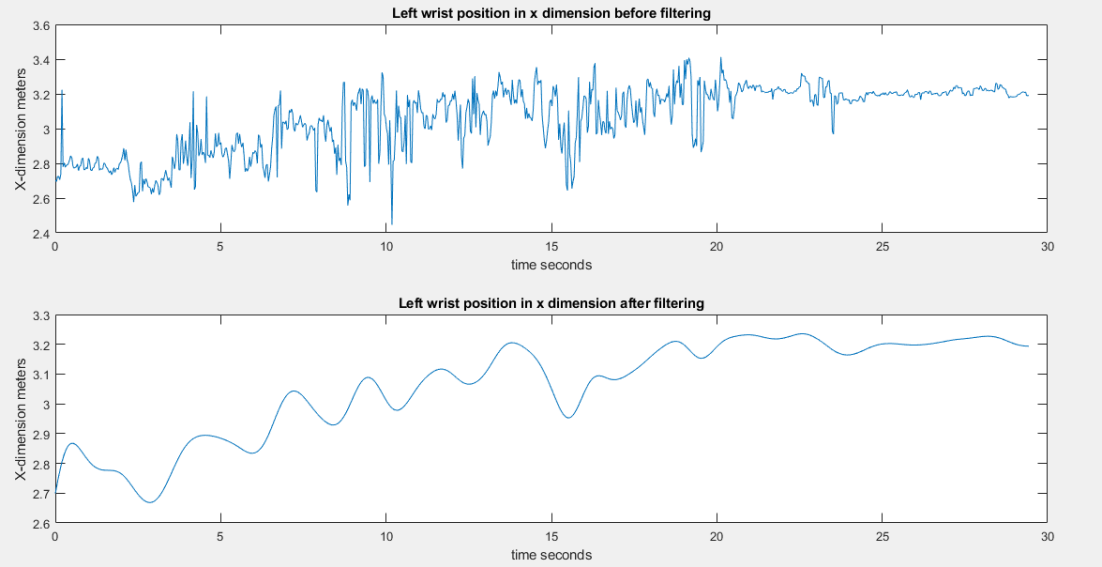

Figure 20. *Left wrist position in x-dimension in meters with time in seconds in E1 before and after filtration.*

Figure 20 shows the left wrist keypoint position over time in x-dimension before and after filtering during E1. Additionally, differentiation amplifies catastrophically the noise specially when the sampling frequency is high.

Figure 21 shows velocities and accelerations of the epileptic seizure: E1. We notice that with higher order of differentiation, the lower signal to noise ratio is.

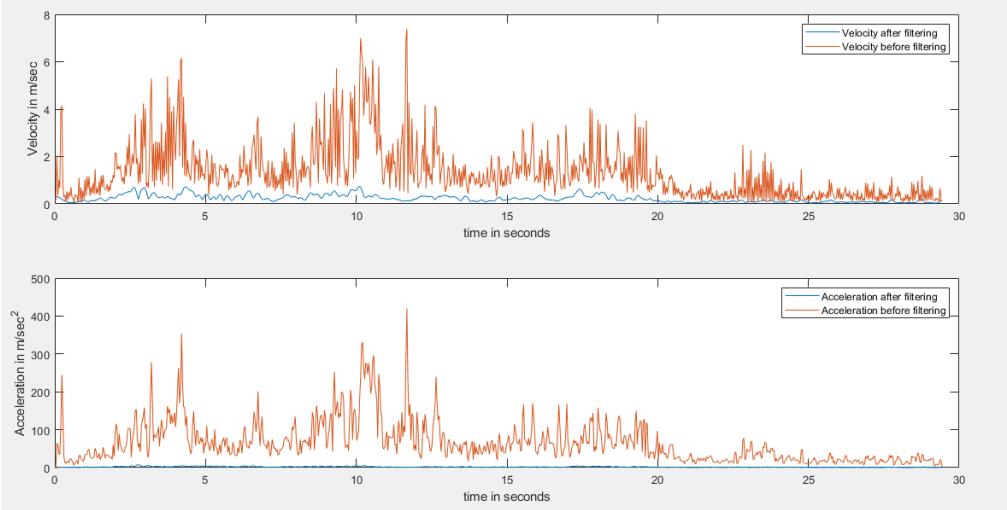

Figure 21. *Velocities and acceleration of E1 before and after filtering.*

In order to assess the inter-rater reliability in this work. The four events were visually assessed to provide a clinical benchmarking. Table 1 shows intensity ranking using a rating scale seizure intensity evaluation on a scale from 0-9.

Table 1. Visual assessments of the four seizure events with a rating scale from 0-9.

	E1	E2	E3	E4
PO	4	5	8	5
SP	4	5	7	5

As we notice in Table 1, there is no consensus on all the four seizures; E3 has two different values relative to the observer's opinion. On the other hand, they agreed that E1 has the least intensity and E3 has the highest intensity. E2 and E4 are equally rated, while other seizures took different values. There was agreement on the ranking of the four seizures in terms of intensity. The descending order is as follows: E3→ E2,E4→E1.

4.2 Motion quantification of seizure events

In this thesis, a novel 3D method of motion quantification is present to study movements during myoclonic epileptic seizures of a patient with LGS using video recordings. The videos recordings represent four seizure events and a REM sleep event, which represents a seizure-free event. In the time domain, metrics such as the motion trajectory (position over time) , velocity, acceleration, displacement travelled (Euclidean distance in 3D space) are computed.

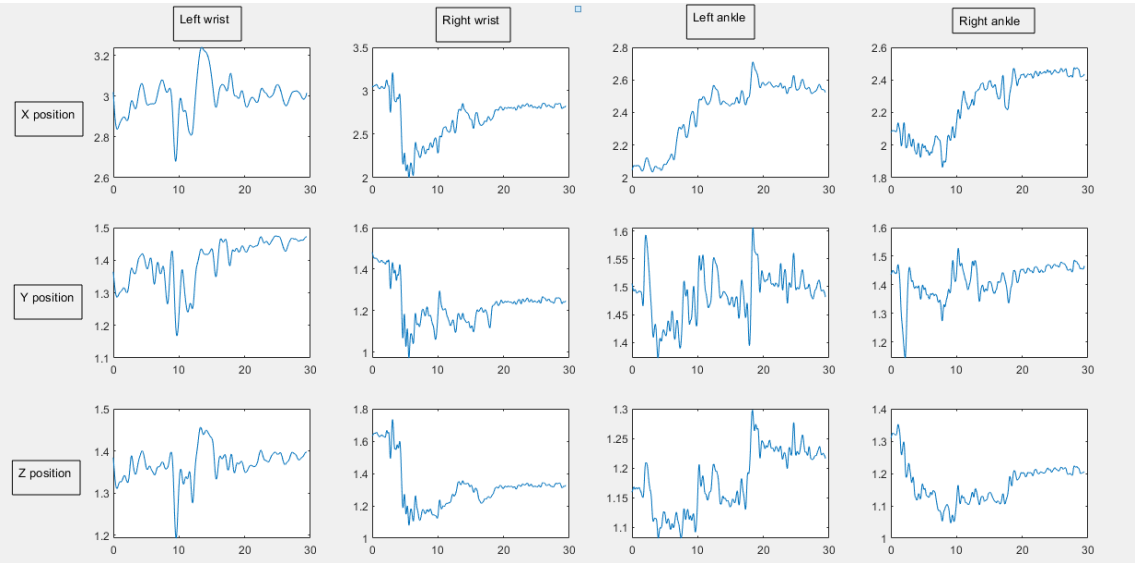

Figure 22. *Positions of the four keypoints over time in x, y, z coordinates in E1. Horizontal axes represent time and vertical axes represent the position with respect to the camera in meters.*

Moreover, metrics that can estimate the oscillatory feature of seizures are calculated such as: spike count SC (number of spikes above certain threshold), spike intensity SI (number of spikes per seconds), weighted spike ratio WSR (proportion of number of spikes for a keypoint to the total number of spikes of the 4 keypoints). Figure 22 shows keypoints' positions in 3D for E1.

Table 2. *Right and left wrists velocities in the 4 seizures and REM.*

	Velocity of Keypoints (ms^{-1})			
	Right wrist		Left wrist	
	Mean	Standard deviation	Mean	Standard deviation
E1	0.3278	0.4448	0.1686	0.1616
E2	0.2090	0.1573	0.1452	0.1728
E3	0.3356	0.3866	0.3856	0.4130
E4	0.1558	0.1443	0.0819	0.0944
REM	0.0715	0.0489	0.0460	0.0438

Table 2 and Table 3 show the magnitude of velocities (mean and standard deviation) of the keypoints in the five events. It is noticeable that the keypoints have asymmetric motion quantities, which indicates the asymmetric nature of myoclonic seizures. In other sense. body extremities do not quantitatively follow the same pattern of movement.

Table 3. *Right and left ankles velocities in the 4 events and REM.*

	Velocity of Keypoints (ms^{-1})			
	Right ankle		Left ankle	
	Mean	Standard deviation	Mean	Standard deviation
E1	0.2102	0.1857	0.1704	0.1399
E2	0.1437	0.1442	0.1576	0.1761
E3	0.1643	0.2142	0.1194	0.1532
E4	0.1342	0.1317	0.1036	0.1230
REM	0.0612	0.0439	0.0324	0.0325

Table 4. *Right and left wrists Accelerations in the 4 events and REM.*

	Acceleration of Keypoints (ms^{-2})			
	Right wrist		Left wrist	
	Mean	Standard deviation	Mean	Standard deviation
E1	3.1043	4.2360	0.7217	0.6719
E2	1.0247	0.7250	0.9444	0.9764
E3	2.5767	2.6820	3.4311	3.2910
E4	0.3982	0.3758	0.1915	0.1913
REM	0.3663	0.2705	0.1729	0.1288

Table 4 and Table 5 show the magnitude of accelerations of the keypoints in the five events. It is also noticeable that keypoints, for each event, differ quantitatively. Which also indicates asymmetric nature of myoclonic seizures.

Table 5. *Right and left ankles accelerations in the 4 events and REM.*

	Acceleration of Keypoints (ms^{-2})			
	Right ankle		Left ankle	
	Mean	Standard deviation	Mean	Standard deviation
E1	1.8355	1.5960	1.2740	0.9857
E2	0.5989	0.5254	1.0714	0.8868
E3	1.6916	1.8832	0.3856	0.4406
E4	0.4792	0.4747	0.4075	0.4075
REM	0.295	0.148	0.1545	0.1001

Table 6. *Fused velocity and acceleration of E1,E2, E3, E4 and REM.*

	Velocity (m/sec)		Acceleration (m/sec^2)	
	Mean	St	Mean	St
E1	0.1468	0.1165	0.7068	0.4340
E2	0.1189	0.0857	0.7068	0.4340
E3	0.1387	0.1636	1.2564	1.2592
E4	0.0971	0.0759	0.2270	0.2260
REM	0.0503	0.0278	0.137	0.139

A way to fuse all the parameters of the 4 keypoints in one single parameter to represent the whole event is to take the median of the keypoints per each event. In other sense, we took the median of the magnitude of velocities for all keypoints to represent the fused velocity of the event. Additionally, we took the median of the magnitude acceleration for all keypoints to represent the fused acceleration of the event.

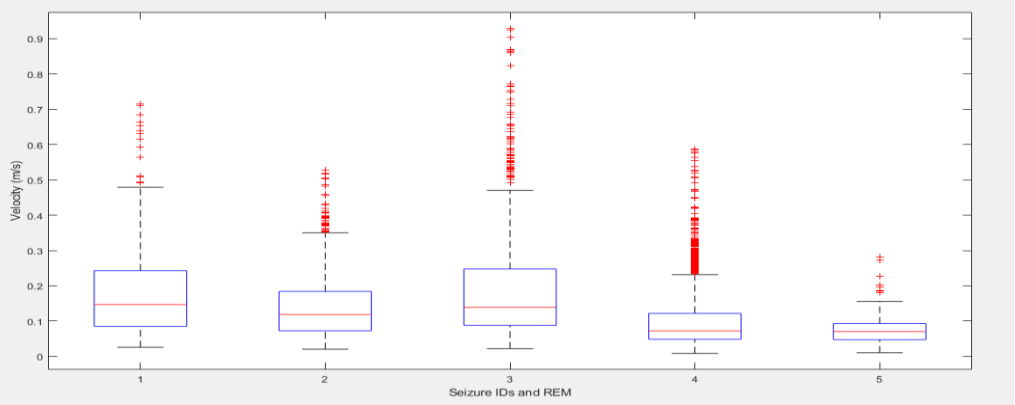

Figure 23. *Box plots of the fused velocity for the 4 events and REM: 1, 2, 3, 4, 5 represent E1, E2, E3, E4 and REM respectively.*

Table 6 shows fused velocity and accelerations of each event. It is noticeable that the descending order of seizures' fused velocities and accelerations is: E3→E1→E2→E4. This is different from the ratings of the visual assessment. Figure 23 shows box plots of the fused velocity for the 4 events and REM.

***T**able 7.* *Total spikes, duration and spike intensity of the 4 events and REM using threshold value= 0.25 m/s*

	Total number of spikes(TS)	Duration (sec)	Spike Intensity (spike/seconds)
E1	50	29.5	1.74
E2	113	49.9	2.26
E3	278	31.1	8.96
E4	242	112.8	2.16
REM	5	20	0.25

In order to estimate the oscillatory feature of a seizure, is to count spikes during seizures with respect to threshold. The threshold in our measurements is arbitrarily defined as the value that is higher than 98% of the fused REM velocity values. Table 7 shows total spikes for each event. It also shows the spike intensity of the five events. The descending order of spike intensity for the five events according to Table 7 is: E3→E2→E4→E1→REM.

Table 8. *Number of spikes, weighted spike ratio in the four seizures and REM.*

	E1		E2		E3		E4		REM	
	SC	WSR (%)	SC	WSR (%)	SC	WSR(%)	SC	WSR (%)	SC	WSR (%)
Right wrist	33	66	12	10.6	104	37.41	124	51.23	4	80
Left wrist	10	20	38	33.6	138	49.64	94	38.8	1	20
Right ankle	3	6	24	21.39	20	7.2	14	5.78	0	0
Left ankle	4	8	26	23	16	5.755	10	4.132	0	0
Total	50	100	113	100	278	100	242	100	5	100

Table 8 shows spike counts for the 4 keypoints per each event. It also shows that each keypoint has different motion intensity. For instance, in E3, left wrist has the highest intensity while left ankle has the lowest value.

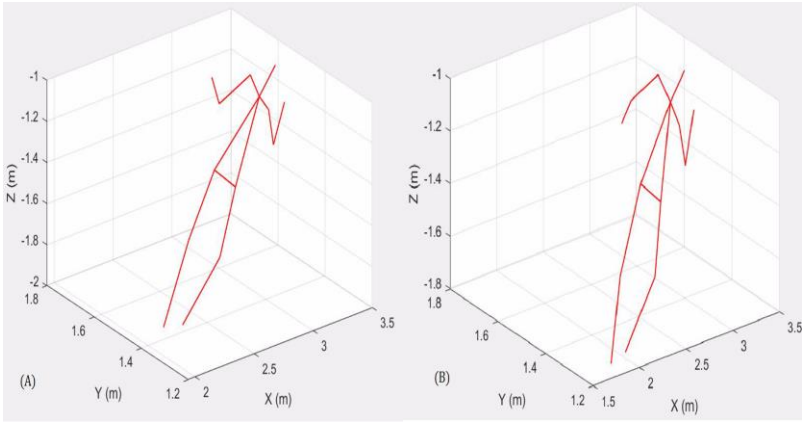

Figure 24. *3D model of E4 at 2 instances: (A) at time = 1 second, (B) at time=3 seconds*

A 3D model (the real dimensions) of the skeletonized body can be fully extracted for each of the four seizures. Figure 24 shows the skeleton of the test subject before the seizure at (t=1 Second) and during the seizure at (t=3 seconds).

5. CONCLUSION

In this thesis, a state-of-art automatic video-based method was used to conduct 3D quantification for the motoric aspects of myoclonic epileptic seizures. Based on the conducted literature review, this is the first time to use a deep learning-based human pose estimation tool to estimate seizure intensity in 3D space from home monitoring video recordings. The data used in the study are 4 myoclonic seizure events recorded during sleep. They belong to a patient with LGS. In addition, a non-movement event recorded during REM sleep was used as the ground truth of the experiment.

The seizure events were visually evaluated by two epileptologists. A comparative study was conducted between the two approaches. The comparative results show low level of inter-rater agreement with the clinical evaluation concerning the features: velocity and acceleration. On the other hand, features that represent the oscillatory characteristics of the seizures, such as: spike intensity, have better inter-rater agreement. which makes automatic detection a potential tool for motion quantification of seizures using video-based methods.

This method can be used for seizure detection and seizure quantification. It does not require markers nor sensors to be attached to the patient, which makes it comfortable and portable, not only for seizure recording, but also for other motor impairment diseases and sport medicine. Moreover, it provides an automatic way for motion capture in 3 dimensions, which gives more information about the motor behavior than the other traditional 2D methods.

On the other hand, challenging cases to pose estimation still exist such as keypoints occlusion (i.e. clothing and blanket, etc.), complex erratic motions, which cannot be accurately tracked and can dramatically affect the results. Additionally, the pose estimation tool, localizes the keypoints frame by frame which introduces effective noise to the system. Other factors such as: the videos quality and the other used estimation techniques in this thesis introduced additional noise. Furthermore, the pose estimation tool, the computer vision algorithms and the signal processing tools require high computational resources.

Future work should be done to enhance the detection accuracy of the pose estimation. Further studies need to be conducted to further validate this approach. Larger datasets representing different seizure events and different patients should be used to further extract better defining features of seizure intensity. Other techniques can be combined

for better development of the overall quality of movement assessment of epileptic seizures, in specific, and other applications that involve motor assessment in general.

REFERENCES

[1] 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.

[2] P.A. Dekker, *EPILEPSY : A manual for Medical and Clinical Officers in Africa and Clinical Officers*. Geneva: World Health Organization, 2002.

[3] A. D. & W. H. O. Murray, Christopher J. L, Lopez, *Global comparative assessments in the health sector : disease burden, expenditures and intervention packages*. 1994.

[4] H. E. Scharfman, “The neurobiology of epilepsy,” *Curr. Neurol. Neurosci. Rep.*, vol. 7, no. 4, pp. 348–354, 2007.

[5] S. Noachtar and A. S. Peters, “Semiology of epileptic seizures: A critical review,” *Epilepsy Behav.*, vol. 15, no. 1, pp. 2–9, 2009.

[6] H. Lu, Y. Pan, B. Mandal, H. L. Eng, C. Guan, and D. W. S. Chan, “Quantifying limb movements in epileptic seizures through color-based video analysis,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 461–469, 2013.

[7] G. A. Baker, D. F. Smith, M. Dewey, J. Morrow, P. M. Crawford, and D. W. Chadwick, “The development of a seizure severity scale as an outcome measure in epilepsy.,” *Epilepsy Res.*, vol. 8, no. 3, pp. 245–51, Apr. 1991.

[8] K. Tufenkjian and H. O. Lüders, “Seizure semiology: Its value and limitations in localizing the epileptogenic zone,” *J. Clin. Neurol.*, vol. 8, no. 4, pp. 243–250, 2012.

[9] P. Kotagal *et al.*, “Dystonic posturing in complex partial seizures of temporal lobe onset: A new lateralizing sign,” *Neurology*, vol. 39, no. 2, pp. 196–201, Feb. 1989.

[10] J. P. Silva Cunha, C. Vollmar, Z. Li, J. Fernandes, B. Feddersen, and S. Noachtar, “Movement quantification during epileptic seizures: A new technical contribution to the evaluation of seizure semiology,” *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 1, pp. 671–673, 2003.

[11] A. Aghaei-Lasboo and R. S. Fisher, “Methods for Measuring Seizure Frequency

and Severity," *Neurol. Clin.*, vol. 34, no. 2, pp. 383–394, 2016.

[12] A. Bleasel, P. Kotagal, P. Kankirawatana, and L. Rybicki, "Lateralizing value and semiology of ictal limb posturing and version in temporal lobe and extratemporal epilepsy," *Epilepsia*, vol. 38, no. 2, pp. 168–174, 1997.

[13] Z. Ma *et al.*, "Recent Advantages of Computer Vision," *IEEE Access*, vol. 6. Institute of Electrical and Electronics Engineers Inc., pp. 31481–31485, 2018.

[14] Ian Goodfellow, Yoshua Bengio, and Aaron Courvill, *Deep Learning*. MIT Press, 2016.

[15] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded Pyramid Network for Multi-person Pose Estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7103–7112, 2018.

[16] J. S. Duncan and J. W. A. S. Sander, "The chalfont seizure severity scale," *J. Neurol. Neurosurg. Psychiatry*, vol. 54, no. 10, pp. 873–876, Oct. 1991.

[17] S. J. Frucht, S. E. Leurgans, M. Hallett, and S. Fahn, "The Unified Myoclonus Rating Scale.," *Adv. Neurol.*, vol. 89, pp. 361–376, 2002.

[18] S. Pietracupa, E. Bruno, A. E. Cavanna, M. Falla, M. Zappia, and C. Colosimo, "Scales for hyperkinetic disorders: A systematic review," *J. Neurol. Sci.*, vol. 358, no. 1–2, pp. 9–21, Jun. 2015.

[19] D. F. Smith, G. A. Baker, M. Dewey, A. Jacoby, and D. W. Chadwick, "Seizure frequency, patient-perceived seizure severity and the psychosocial consequences of intractable epilepsy," *Epilepsy Res.*, vol. 9, no. 3, pp. 231–241, Sep. 1991.

[20] M. F. O'Donoghue, J. S. Duncan, and J. W. A. S. Sander, "The National Hospital Seizure Severity Scale: A Further Development of the Chalfont Seizure Severity Scale," *Epilepsia*, vol. 37, no. 6, pp. 563–571, Jun. 1996.

[21] J. A. Cramer *et al.*, "A method of quantification for the evaluation of antiepileptic drug therapy," *Neurology*, vol. 33, no. Issue 3, Supplement 1, pp. 26–37, Mar. 1983.

[22] R. S. Fisher, G. Nune, S. E. Roberts, and J. A. Cramer, "The Personal Impact of Epilepsy Scale (PIES)," *Epilepsy Behav.*, vol. 42, pp. 140–146, Jan. 2015.

[23] H. A. Carpay and W. F. M. Arts, "Outcome assessment in epilepsy: Available rating scales for adults and methodological issues pertaining to the development of scales for childhood epilepsy," *Epilepsy Res.*, vol. 24, no. 3, pp. 127–136, Jul.

1996.

[24] H. A. Carpay *et al.*, “Parent-completed scales for measuring seizure severity and severity of side-effects of antiepileptic drugs in childhood epilepsy: Development and psychometric analysis,” *Epilepsy Res.*, vol. 24, no. 3, pp. 173–181, Jul. 1996.

[25] G. A. Baker, D. F. Smith, A. Jacoby, J. A. Hayes, and D. W. Chadwick, “Liverpool Seizure Severity Scale revisited,” *Seizure*, vol. 7, no. 3, pp. 201–205, Jun. 1998.

[26] A. Ulate-Campos, F. Coughlin, M. Gaínza-Lein, I. S. Fernández, P. L. Pearl, and T. Loddenkemper, “Automated seizure detection systems and their effectiveness for each type of seizure,” *Seizure*, vol. 40. W.B. Saunders Ltd, pp. 88–101, 01-Aug-2016.

[27] J. P. S. Cunha *et al.*, “NeuroKinect: A novel low-cost 3Dvideo-EEG ystem for epileptic seizure motion quantification,” *PLoS One*, vol. 11, no. 1, pp. 1–17, 2016.

[28] R. K. Maganti and P. Rutecki, “EEG and epilepsy monitoring,” *Contin. Lifelong Learn. Neurol.*, vol. 19, no. 3, pp. 598–622, Jun. 2013.

[29] P. Chauvel and A. McGonigal, “Emergence of semiology in epileptic seizures,” *Epilepsy Behav.*, vol. 38, pp. 94–103, Sep. 2014.

[30] A. Olivares, G. Olivares, F. Mula, J. M. Górriz, and J. Ramírez, “Wagyromag: Wireless sensor network for monitoring and processing human body movement in healthcare applications,” *J. Syst. Archit.*, vol. 57, no. 10, pp. 905–915, Nov. 2011.

[31] S. Beniczky, T. Polster, T. W. Kjaer, and H. Hjalgrim, “Detection of generalized tonic-clonic seizures by a wireless wrist accelerometer: A prospective, multicenter study,” *Epilepsia*, vol. 54, no. 4, pp. e58–e61, Apr. 2013.

[32] U. Kramer, S. Kipervasser, A. Shlitner, and R. Kuzniecky, “A novel portable seizure detection alarm system: Preliminary results,” *J. Clin. Neurophysiol.*, vol. 28, no. 1, pp. 36–38, Feb. 2011.

[33] J. Lockman, R. S. Fisher, and D. M. Olson, “Detection of seizure-like movements using a wrist accelerometer,” *Epilepsy Behav.*, vol. 20, no. 4, pp. 638–641, Apr. 2011.

[34] T. M. E. Nijsen, J. B. A. M. Arends, P. A. M. Griep, and P. J. M. Cluitmans, “The

potential value of three-dimensional accelerometry for detection of motor seizures in severe epilepsy," *Epilepsy Behav.*, vol. 7, no. 1, pp. 74–84, 2005.

[35] S. Bonnet and R. Heliot, "A magnetometer-based approach for studying human movements," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 7, pp. 1353–1355, Jul. 2007.

[36] B. Kemp, A. J. Janssen, and B. van der Kamp, "Body position can be monitored in 3D using miniature accelerometers and earth-magnetic field sensors.," *Electroencephalogr. Clin. Neurophysiol.*, vol. 109, no. 6, pp. 484–8, Dec. 1998.

[37] G. Becq, S. Bonnet, L. Minotti, M. Antonakios, R. Guillemaud, and P. Kahane, "Collection and exploratory analysis of attitude sensor data in an epilepsy monitoring unit," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 2007, vol. 2007, pp. 2775–2778.

[38] M. Pediaditis, M. Tsiknakis, and N. Leitgeb, "Vision-based motion detection, analysis and recognition of epileptic seizures-A systematic review," *Comput. Methods Programs Biomed.*, vol. 108, no. 3, pp. 1133–1148, Dec. 2012.

[39] Z. Li, A. Martins Da Silva, and J. P. Silva Cunha, "Movement quantification in epileptic seizures: A new approach to video-EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 6, pp. 565–573, Jun. 2002.

[40] Z. Li, A. Martins Da Silva, and J. P. Silva Cunha, "Movement quantification in epileptic seizures: A new approach to video-EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 6, pp. 565–573, 2002.

[41] J. P. S. Cunha, L. M. Paula, V. F. Bento, C. Bilgin, E. Dias, and S. Noachtar, "Movement quantification in epileptic seizures: A feasibility study for a new 3D approach," *Med. Eng. Phys.*, vol. 34, no. 7, pp. 938–945, 2012.

[42] E. N. Marieb and S. M. Keller, *Essentials of human anatomy & physiology*, 12th ed. 2017.

[43] P. D. J. Gordon Betts, aKelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, *Anatomy & Physiology*. 2013.

[44] Editor-in-Chief and L. R. Squire, "Encyclopedia of Neuroscience | ScienceDirect," 2009. [Online]. Available: https://www.sciencedirect.com/referencework/9780080450469/encyclopedia-of-neuroscience. [Accessed: 12-Apr-2020].

[45] B. S. Meldrum, “The role of glutamate in epilepsy and other CNS disorders,” *Neurology*, vol. 44, no. 11, pp. S14–S23, 1994.

[46] J. J. Engel and T. A. Pedley, *Epilepsy: A Comprehensive Textbook*, 2nd ed. 1998.

[47] R. S. Fisher *et al.*, “Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE),” *Epilepsia*, vol. 46, no. 10, pp. 1701–1702, 2005.

[48] M. A. Samuels and S. K. Feske, *Office Practice of Neurology: Second Edition*. Elsevier Inc., 2003.

[49] P. Bonanni, L. Parmeggiani, and R. Guerrini, “Different neurophysiologic patterns of myoclonus characterize Lennox-Gastaut syndrome and myoclonic astatic epilepsy,” *Epilepsia*, vol. 43, no. 6, pp. 609–615, 2002.

[50] G. G. Somjen, “Ion regulation in the brain: Implications for pathophysiology,” *Neuroscientist*, vol. 8, no. 3. SAGE Publications Inc., pp. 254–267, 2002.

[51] C. Vaillend, S. E. Mason, M. F. Cuttle, and B. E. Alger, “Mechanisms of neuronal hyperexcitability caused by partial inhibition of Na+-K+-ATPases in the rat CA1 hippocampal region,” *J. Neurophysiol.*, vol. 88, no. 6, pp. 2963–2978, 2002.

[52] T. Fellin and P. G. Haydon, “Do astrocytes contribute to excitation underlying seizures?,” *Trends in Molecular Medicine*, vol. 11, no. 12. Elsevier Ltd, pp. 530–533, Dec-2005.

[53] M. H. Meisler, J. Kearney, R. Ottman, and A. Escayg, “Identification of Epilepsy Genes in Human and Mouse,” *Annu. Rev. Genet.*, vol. 35, no. 1, pp. 567–588, Dec. 2001.

[54] R. D. Traub, H. Michelson-Law, A. E. J. Bibbig, E. H. Buhl, and M. A. Whittington, “Gap junctions, fast oscillations and the initiation of seizures,” *Advances in Experimental Medicine and Biology*, vol. 548. pp. 110–122, 2004.

[55] J. J. Riviello, “Classification of seizures and epilepsy,” *Curr. Neurol. Neurosci. Rep.*, vol. 3, no. 4, pp. 325–331, 2003.

[56] “epileptogenic lesion – APA Dictionary of Psychology.” [Online]. Available: https://dictionary.apa.org/epileptogenic-lesion. [Accessed: 04-Apr-2020].

[57] H. O. Lüders, I. Najm, D. Nair, P. Widdess-Walsh, and W. Bingman, “The epileptogenic zone: General principles,” *Epileptic Disord.*, vol. 8, no. SUPPL. 2, pp. 1–9, 2006.

[58] Thomas R. Browne and Gregory L. Holmes, *Handbook of Epilepsy*, 4th ed. 2008.

[59] H. GASTAUT *et al.*, “Childhood Epileptic Encephalopathy with Diffuse Slow Spike-Waves (otherwise known as ‘Petit Mal Variant’) or Lennox Syndrome,” *Epilepsia*, vol. 7, no. 2, pp. 139–179, Jun. 1966.

[60] “EPILEPTIC SPASMS.” [Online]. Available: https://www.epilepsydiagnosis.org/seizure/epileptic-spasms-overview.html. [Accessed: 06-Apr-2020].

[61] T. Yu *et al.*, “Surgical treatment of hypermotor seizures originating from the temporal lobe,” *Seizure*, vol. 22, no. 10, pp. 862–866, Dec. 2013.

[62] H. Lüders *et al.*, “Semiological seizure classification,” *Epilepsia*, vol. 39, no. 9, pp. 1006–1013, Sep. 1998.

[63] Peter Corke, *Robotics, Vision and Control: Fundamental Algorithms in MATLAB (Springer Tracts in Advanced Robotics Book 73), Corke, Peter, eBook - Amazon.com*, 1st ed. Springer, 2011.

[64] S. Richard, *Computer Vision - Algorithms and Applications*, 1st ed. Springer, 2010.

[65] C. Loop and Z. Zhang, “Computing rectifying homographies for stereo vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 125–131, 1999.

[66] H. Hirschmüller and D. Scharstein, “Evaluation of Cost Functions for Stereo Matching.”

[67] W. Wu, H. Zhu, and Q. Zhang, “Oriented-linear-tree based cost aggregation for stereo matching,” *Multimed. Tools Appl.*, vol. 78, no. 12, pp. 15779–15800, Jun. 2019.

[68] P. F. Felzenszwalb and D. P. Huttenlocher, “Pictorial structures for object recognition,” *Int. J. Comput. Vis.*, vol. 61, no. 1, pp. 55–79, Jan. 2005.

[69] M. A. Fischler and R. A. Elschlager, “The Representation and Matching of Pictorial Structures Representation,” *IEEE Trans. Comput.*, vol. C–22, no. 1, pp. 67–92, 1973.

[70] L. Bo, C. Sminchisescu, A. Kanaujia, and D. Metaxas, “Fast algorithms for large scale conditional 3D prediction,” in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[71] K. Grauman, G. Shakhnarovich, and T. Darrell, “Inferring 3D structure with a statistical image-based shape model,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2003, vol. 1, pp. 641–648.

[72] M. Salzmann and R. Urtasun, “Combining discriminative and generative methods for 3D deformable surface and articulated pose reconstruction,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 647–654.

[73] L. Pishchulin *et al.*, “DeepCut: Joint subset partition and labeling for multi person pose estimation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December, pp. 4929–4937.

[74] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, “Deepercut: A deeper, stronger, and faster multi-person pose estimation model,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9910 LNCS, pp. 34–50.

[75] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December, pp. 770–778.

[76] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, pp. 1302–1310.

[77] A. Newell, Z. Huang, and J. Deng, “Associative embedding: End-to-end learning for joint detection and grouping,” in *Advances in Neural Information Processing Systems*, 2017, vol. 2017-December, pp. 2278–2288.

[78] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, pp. 2980–2988.

[79] G. Papandreou *et al.*, “Towards Accurate Multi-person Pose Estimation in the Wild,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 3711–3719, Jan. 2017.

[80] S. Huang, M. Gong, and D. Tao, “A Coarse-Fine Network for Keypoint

Localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, pp. 3047–3056.

[81] A. Newell, K. Yang, and J. Deng, "Stacked Hourglass Networks for Human Pose Estimation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9912 LNCS, pp. 483–499, Mar. 2016.

[82] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1653–1660.

[83] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation," *Adv. Neural Inf. Process. Syst.*, vol. 2, no. January, pp. 1799–1807, Jun. 2014.

[84] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional Pose Machines," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 4724–4732, Jan. 2016.

[85] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human Pose Estimation with Iterative Error Feedback," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 4733–4742, Jul. 2015.

[86] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[87] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[88] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Dec. 2016.

[89] "COCO - Common Objects in Context." [Online]. Available: http://cocodataset.org/#home. [Accessed: 04-May-2020].

[90] Intel Corportation, "Intel RealSense Depth Camera D400-Series," no. September, 2017.

[91] "FFmpeg." [Online]. Available: https://www.ffmpeg.org/. [Accessed: 24-Apr-2020].

[92] Itseez, "Open Source Computer Vision Library," 2019. [Online]. Available: https://github.com/opencv/opencv. [Accessed: 24-Apr-2020].

[93] Robert Piche (2020), “derest.” [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/63925-derest. [Accessed: 20-Apr-2020].

[94] R. Piché, “Automatic numerical differentiation by maximum likelihood estimation of state-space model ⋆,” 2016.

[95] B. Yang, X. Cheng, D. Dai, T. Olofsson, H. Li, and A. Meier, “Macro pose based non-invasive thermal comfort perception for energy efficiency,” Nov. 2018.