UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Engenharia Mecânica

YU TZU WU

Optical myography system for hand posture and gesture recognition

Sistema de miografia óptica para reconhecimento de gestos e posturas da mão

CAMPINAS

2018

YU TZU WU

Optical myography system for hand posture and gesture recognition

Sistema de miografia óptica para reconhecimento de gestos e posturas da mão

Thesis presented to the School of Mechanical Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Mechanical Engineering, in the area of Mechatronics.

Dissertação apresentada à Faculdade de Engenharia Mecânica da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Mestra em Engenharia Mecânica, na Área de Mecatrônica.

Orientador: Prof. Dr. Éric Fujiwara

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELA ALUNA YU TZU WU E ORIENTADA PELO PROF. DR. ÉRIC FUJIWARA.

CAMPINAS
2018

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

W95o   Wu, Yu Tzu, 1992-
        Optical myography system for hand posture and gesture recognition / Yu Tzu Wu. – Campinas, SP : [s.n.], 2018.

        Orientador: Éric Fujiwara.
        Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica.

        1. Visão por computador. 2. Redes neurais (Computação). I. Fujiwara, Éric, 1985-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Sistema de miografia óptica para reconhecimento de gestos e posturas de mão
Palavras-chave em inglês:
Computer vision
Neural networks (Computing)
Área de concentração: Engenharia Mecânica
Titulação: Mestra em Engenharia Mecânica
Banca examinadora:
Éric Fujiwara [Orientador]
Alberto Luiz Serpa
Hélio Pedrini
Data de defesa: 24-01-2018
Programa de Pós-Graduação: Engenharia Mecânica

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
COMISSÃO DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
DEPARTAMENTO DE SISTEMAS INTEGRADOS

DISSSERTAÇÃO DE MESTRADO ACADÊMICO

Optical myography system for hand posture and gesture recognition

Sistema de miografia óptica para reconhecimento de gestos e posturas da mão

Autora: Yu Tzu Wu
Orientador: Éric Fujiwara

A Banca Examinadora composta pelos membros abaixo aprovou esta Dissertação:

Prof. Dr. Éric Fujiwara, Presidente
Faculdade de Engenharia Mecânica, Unicamp

Prof. Dr. Alberto Luiz Serpa
Faculdade de Engenharia Mecânica, Unicamp

Prof. Dr. Hélio Pedrini
Instituto da Computação, Unicamp

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno.

Campinas, 24 de janeiro de 2018.

To my parents and my brother,
the giants in my life.

To Aunt Maria and Professor Adonai,
without whom I could not have been here.

ACKNOWLEDGEMENTS

RESUMO

Nesse projeto, demonstrou-se um sistema de miografia óptica como uma alternativa promissora para monitorar as posturas da mão e os gestos do usuário. Essa técnica se fundamenta em acompanhar as atividades musculares responsáveis pelos movimentos da mão com uma câmera externa, relacionando a distorção visual verificada no antebraço com a contração e o relaxamento necessários para dada postura. Três configurações de sensores foram propostas, estudadas e avaliadas. A primeira propôs monitorar a atividade muscular analisando a variação da frequência espacial de uma textura de listras uniformes impressa sobre a pele, enquanto que a segunda se caracteriza pela contagem de pixels de pele visível dentro da região de interesse. Ambas as configurações se mostraram inviáveis pela baixa robustez e alta demanda por condições experimentais controladas. Por fim, a terceira recupera o estado da mão acompanhando o deslocamento de uma série de marcadores coloridos distribuídos ao longo do antebraço. Com um webcam de 24 fps e 640×480 pixels, essa última configuração foi validada para oito posturas distintas, explorando principalmente a flexão e extensão dos dedos e do polegar, além da adução e abdução do último. Os dados experimentais, adquiridos off-line, são submetidos a uma rotina de processamento de imagens para extrair a informação espacial e de cor dos marcadores em cada quadro, dados esses utilizados para rastrear os mesmos marcadores ao longo de todos os quadros. Para reduzir a influência das vibrações naturais e inerentes ao corpo humano, um sistema de referencial local é ainda adotado dentro da própria região de interesse. Finalmente, os dados quadro a quadro com o *ground truth* são alimentados a uma rede neural artificial sequencial, responsável pela calibração supervisionada do sensor e posterior classificação das posturas. O desempenho do sistema para a classificação das oito posturas foi avaliado com base na validação cruzada com 10-folds, com a câmera monitorando o antebraço pela superfície interna ou externa. O sensor apresentou uma precisão de $\sim$92.4% e exatidão de $\sim$97.9% para o primeiro caso, e uma precisão de $\sim$75.1% e exatidão de $\sim$92.5% para o segundo, sendo comparável a outras técnicas de miografia, demonstrando a viabilidade do projeto e abrindo perspectivas para aplicações em interfaces humano-robô.

Palavras-chave: reconhecimento de gestos, visão computacional, redes neurais, sensor a miografia óptica.

ABSTRACT

In this work, an optical myography system is demonstrated as a promising alternative to monitor hand posture and gestures of the user. This technique is based on accompanying muscular activities responsible for hand motion with an external camera, and relating the visual deformation observed on the forearm to the muscular contractions/relaxations for a given posture. Three sensor designs were proposed, studied and evaluated. The first one intended to monitor muscular activity by analyzing the spatial frequency variation of a uniformly distributed stripe pattern stamped on the skin, whereas the second one is characterized by reckoning visible skin pixels inside the region of interest. Both designs are impracticable due to their low robustness and high demand for controlled experimental conditions. At last, the third design retrieves hand configuration by tracking visually the displacements of a series of color markers distributed over the forearm. With a webcam of 24 fps and 640×480 pixels, this design was validated for eight different postures, exploring fingers and thumb flexion/extension, plus thumb adduction/abduction. The experimental data are acquired offline and, then, submitted to an image processing routine to extract color and spatial information of the markers in each frame; the extracted data is subsequently used to track the same markers along all frames. To reduce the influence of human body natural and inherent vibrations, a local reference frame is yet adopted in the region of interest. Finally, the frame by frame data, along with the ground truth posture, are fed into a sequential artificial neural network, responsible for sensor supervised calibration and subsequent posture classification. The system performance was evaluated in terms of eight postures classification via 10-fold cross-validation, with the camera monitoring either the underside or the back of the forearm. The sensor presented a $\sim$92.4% precision and $\sim$97.9% accuracy for the former, and a $\sim$75.1% precision and $\sim$92.5% accuracy for the latter, being thus comparable to other myographic techniques; it also demonstrated that the project is feasible and offers prospects for human-robot interaction applications.

Keywords: gesture recognition, computer vision, neural networks, optical myography sensor.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

3PI	Three palmar interossei
4DI	Four dorsal interossei
4L	Four lumbricales
AAM	Active appearance model
ANN	Artificial neural network
AP	Adductor pollicis
APB	Abductor pollicis brevis
APL	Abductor pollicis longus
ASM	Active shape model
CNN	Convolutional neural network
DLR	German Aerospace Center
DoF	Degree of freedom
ECRB	Extensor carpi radialis brevis
ECRL	Extensor carpi radialis longus
ECU	Extensor carpi ulnaris
ED	Extensor digitorum
EDM	Extensor digiti minimi
EI	Extensor indicis
EM	Electromagnetic
EMG	Electromyography
EPB	Extensor pollicis brevis
EPL	Extensor pollicis longus
FCR	Flexor carpi radialis
FCU	Flexor carpi ulnaris
FEM	Faculty of Mechanical Engineering
FMG	Force myography
FPaL	Flexor palmaris longus

FPB	Flexor pollicis brevis
FPD	Flexor profundus digitorum
FPL	Flexor pollicis longus
FSD	Flexor sublimis digitorum
FSR	Force sensing resistor
HCI	Human-computer interaction
HSV	Hue, saturation, value (color model)
HT	Hough transform
LIQC	Laboratory of Photonic Materials and Devices
MoCap	Motion capture
OM	Opto-mechanical
OMG	Optical myography
OP	Opponnens pollicis
OpenCV	Open Source Computer Vision Library
PDM	Pressure distributing map
RGB	Red, green, blue (color model)
RKI	Residual kinect imaging
RM	Institute of Robotics and Mechatronics
ROI	Region of interest
sEMG	Surface electromyography
UNICAMP	University of Campinas

LIST OF SYMBOLS

Latin letters

B	Convolution kernel
ch	Image channel
I_{in}	Input image
$I_{in}(x, y)$	Pixel (x, y) of the input image
$I_{in}(x, y, ch)$	Pixel (x, y, ch) of the input image
I_{marker}	Marker for morphological operation
I_{mask}	Mask for morphological operation
I_{out}	Output image
$I_{out}(x, y)$	Pixel (x, y) of the output image
M_{blue}	Binary mask of blue objects
M_{green}	Binary mask of green objects
$M_{nonskin}$	Binary mask with non-skin objects
M_{ROI}	Binary mask of region of interest
$M_{ROI}(x, y)$	Pixel (x, y) of the binary mask defining the region of interest
M_{skin}	Binary mask of skin objects
M_{tapes}	Binary mask of green tape objects
N_{mkr}	Number of valid markers per experimental set
n_c	Number of columns in the image
N_i	Number of pixels in region of interest sub region i
n_i	Relative number of pixels in region of interest sub region
n_r	Number of rows in the image
R_i	Subregion i inside region of interest
Th	Image gray threshold
$Th[ch]$	Image color threshold
x	Pixel row x
XY	Global reference frame (sensor design D)
xy_0	Local reference frame 0 (sensor design D)
xy_1	Local reference frame 1 (sensor design D)

y	Pixel column y

Logic symbols

$\wedge$	Logical AND
$\vee$	Logical OR

CONTENTS

1 INTRODUCTION

Human hand plays a fundamental role in accomplishing daily tasks and activities. For instance, one grasps and handles objects with the hand; it is also with the hand that one performs assorted chores, both simple and complex ones, such as, e.g., sewing or acrobatic manipulation of sharp objects. The hand is also an efficient and natural means, which allows one to explore and understand the surrounding environment. Through touch, it is possible to feel the texture of a surface, estimate the weight and the temperature of an object, and so forth (MELCHIORRI AND KANEKO, 2016). Moreover, gestures play an important part in the communication process, providing emphasis or extra meaning to the spoken language, or assuming itself the whole means of communication, which is the case of the sign language (WANG *et al.*, 2002).

Modern technology requires yet increasing interaction via human hand. Smartphones, tablets and even computers are now built with capacitive skin-sensitive screens, and user commands are given directly through touch. There are also intelligent systems capable of interacting with users without external mechanisms such as e.g., remote control or mouse. Instead, they adopt more intuitive means of interaction like voice or gesture control. In addition to human-computer interaction (HCI), hand pose and movement information may be used in the actual control of mechanical systems, as robotic manipulator teleoperation, assistive surgery, prostheses control, and so forth (MURTHY AND JADON, 2009). On the other hand, data recovered from actual hand pose or movements are also valuable in the entertainment field. For example, this information may be used to generate more realistic geometric models for computer graphics or animation purposes, or else as means of user interaction with virtual reality environments and its elements (WHEATLAND *et al.*, 2015).

At this point, it is worth noting that, though many of the aforementioned applications may require or not an intact hand, in some of them, a hand with a full range of movements is a must. For instance, a person with limited movement in the hands would hardly be allowed to perform a minimally invasive surgery. On the other hand, when one is talking about exoskeleton activation in rehabilitation treatments or yet, prostheses control, it is implied that the user has lost partly or fully his or her movements in the hand. In such cases, hand pose information is not very useful by itself, and the prosthetic or rehabilitation system tries to recover the motion intention instead (CASTELLINI *et al.*, 2014).

In this context, the task of hand posture or motion intention detection is a fundamental one in modern technology. A note on this topic is that, since gestures are a sequence of postures arranged along an interval of time, this task includes gesture detection already.

Glove-based systems are a common option for this task. They are characterized by an array of strategically placed sensors, an electronic setup for power supply and data acquisition/processing, and a wearable support, which is the most important trait of these systems. The sensors are usually attached over the user knuckles and play the role of transducers by converting the mechanical information such as, e.g., finger flexion, into an output quantity specified by sensor operating principle. For instance, instrumented gloves composed of resistance sensors, magnetic sensors, and fiber-optic sensors were proposed (WANG *et al.*, 2017). The glove-based systems are advantageous despite the drawbacks of each type of sensing element, being mostly easy to use, lightweight and inexpensive. However, it was noticed already that their performance can vary from user to user since there is a high diversity of hand shape. For instance, if the user fingers were shorter, the sensing elements might be moved away from the ideal sensing spot. Moreover, since the support glove is usually made of actual cloth that encloses the user hand, it might not fit the user, e.g., it might be too loose or too tight, and can be rather cumbersome in such cases (FUJIWARA *et al.*, 2014).

This problem can be avoided by computer vision-based sensors, which employ various image processing techniques to extract hand pose information from an image, and gesture information from a sequence of images. Single or multiple cameras point to the user hand and, depending on the sensing principle, active, passive or no markers at all may be attached to the skin (MOESLUND, 2000). With modern software and hardware capacity, these systems have become very flexible, being able to operate in real time, indoors or outdoors, with a high data transfer rate and inexpensively. In addition to these advantages, they provide also the comfort of use and little restriction to hand movements, thus they are an interesting alternative for hand pose or gesture estimation in modern technology (REN *et al.*, 2013). However, these sensors present low portability, since it is mandatory that the hand must be inside the camera field of view for their proper functioning (WANG *et al.*, 2017). With this shortcoming, it is implied that they are not suitable for prostheses control or any other application in which the users have restricted mobility in their hand or have lost their hand or limb, whether accidentally or by amputation.

In such cases, the adopted sensor should be able to extract the motion intent instead of the motion itself. Both sensors presented thus far are not suitable for this task, leading to a type of sensing principle based on myographic examination, which consists in measuring muscular phenomena.

Currently, a known hand monitoring technique with such operating principle is the surface electromyography (sEMG). The sensor includes an array of electrodes that measure the muscle action potential evoked by one or more muscles required to perform a pose or gesture (BRUNELLI *et al.*, 2015). Though this technique has been used for hand prostheses myoelectric control since the late 60s, its implementation in practice is still challengeable, since an accurate

acquisition of sEMG signals is required, but these signals are tempered with various kinds of background noises. For instance, both the presence of electronic equipment and the internal structure of the subject such as, e.g., the individual skin formation, blood flow velocity, skin temperature and tissue structure, all influence end sEMG signals (CHOWDHURY *et al.*, 2013). Moreover, it is reported that the placement of sEMG electrodes, as well as electrodes displacements, also influences the final signal (FANG *et al.*, 2015).

While sEMG measures the action potential of the muscles, force myography (FMG) is another type of myographic technique, which monitors pressure or radial deformation of a limb due to muscular contraction and relaxation (FERIGO *et al.*, 2017). Different setups for this type of sensor have been proposed, they are composed mainly by force or pressure sensing elements arranged around the forearm of the subject, whether radially distributed or positioned at specific locations where the radial deformation is known to be more intense due to muscular contraction. One common approach uses force sensing resistors (FSRs), but FMG setups based on optical fiber force sensors also have been proposed (CHO *et al.*, 2016; FUJIWARA *et al.*, 2017). Both setups have been demonstrated to provide reliable measurements, are inexpensive and convenient. Moreover, in comparison with sEMG, they are robust to sweating, muscle fatigue, body fat, skin temperature and external electromagnetic (EM) interference (JIANG *et al.*, 2017). Regardless of these benefits, FMG systems may cause some discomfort after a prolonged time of use, due to a potential preload, required to set the sensor to its linear behavior range. Hypothetically, if the user should keep the hand in a posture with maximum radial expansion, this deformation added to the initial preload would eventually distress the forearm of the subject. Even if he or she should not keep the hand in such posture, depending on the initial preload that is required, the sensor strap might be very tightly or more loosely fixed around the forearm, thus causing different tactile sensation to the user.

The cumbersomeness from the need of wearing a structure around a limb can be avoided by optical myography (OMG), which combines computer vision techniques with the myographic examination. Instead of pointing a camera toward the hand, it is pointed toward the forearm, and the radial deformation due to muscular activities is monitored visually. Though there is no need for the user to wear a strap around the forearm in principle, the human skin homogeneity requires additional markers as points of interest and, thus, increasing the system sensitivity (NISSLER *et al.*, 2017; WU *et al.*, Accepted for presentation). Naturally, some of the muscles responsible for moving the fingers and the thumb are found in the hand; however, these muscles are kept out of the camera scope, since OMG should be a suitable alternative also for subjects who have suffered partial or total loss of the hand or its movements (WU *et al.*, Accepted for presentation).

Though there are various approaches to recover hand posture and gesture information, this problem is not a closed case. New and better sensors are still proposed and developed, as

general technology naturally evolves. In this context, gestures are a feature increasingly used by computational systems as input command, whether as an intuitive interaction language or as a control reference of mechatronics devices in various applications. Thus, it is of utmost importance that the hand posture should be provided efficiently to computers.

In this sense, the main goal of this project comprehends the development of such a sensor, which should be able to recover not only posture information but also the user intent, expanding the sensor range of application and potential user profiles. Therefore, the sensing principle should be chosen among the myographic approaches. sEMG is prone to various types of noises and requires an expensive machinery to acquire and process the myoelectric signals. FMG, on the other hand, is more robust to external electrical noises as well as physiological issues. However, when it comes to the prolonged time of use, comfort becomes a weighty factor, therefore OMG was chosen for being the most comfortable of the three. Besides, because of its vision-based approach, its setup is usually simple, composed of single or multiple image acquisition devices and visual markers acting as points of interest. Regarding the data processing, it is performed with a regular image processing or computer vision routine, i.e., it is convenient, simple and even inexpensive for current computational technology. The simplicity, accessibility, low-cost and comfort promised by OMG all are in favor of the choice since they are contributing factors to popularize the selected technology for general users and applications.

OMG, however, is a relatively new approach, and its potential is not yet fully explored. Currently, there are only two groups researching this subject, namely one in the Institute of Robotics and Mechatronics (RM), German Aerospace Center (DLR) (NISSLER *et al.*, 2016), and another in the Laboratory of Photonic Materials and Devices (LIQC), Faculty of Mechanical Engineering (FEM), University of Campinas (UNICAMP) (WU *et al.*, Accepted for presentation). In addition to the small number of groups working with OMG, at the time when it was decided that the project would focus on developing an OMG sensor for hand posture monitoring, i.e., two years ago, there was only one published work demonstrating this technique as being promising (NISSLER *et al.*, 2015), and the LIQC, FEM/UNICAMP, was only a beginner at the subject.

Due to the lack of literature, there were very little references on the subject, so the project evolved step by step according to the hypotheses at the time, which were tested and validated, adapted or excluded.

The first consideration was regarding how the radial deformation would look like in a 2D image and whether an ordinary camera could detect the resultant visual deformation or not. Since cameras do not have infinite pixels, 3D space points that are very close to each other can be condensed into a single pixel in the photo. Thus, if the visual deformation was subtle and the camera resolution was low, the sensor would not recognize the posture change. Naturally, a high definition camera could be chosen, but then the final sensor would be less accessible

to the common user, which is not the initial goal of this project. Moreover, higher definition means also higher processing effort and required memory, which should not be a problem for current computers. In this case, however, it would be very difficult to use the sensor in real time applications. In other words, the resultant sensor would be unsuitable for a lot of current applications. Therefore, the camera specification should be defined by considering also the price-performance ratio of the sensor, and not only the image resolution.

Another issue is the selection of a suitable transducer to convert the muscular activity in the mechanical domain into optical domain for subsequent processing. Since human skin is usually homogeneous, there are few points of interest on the skin surface as a reference to monitor the radial deformation of the forearm. Therefore, additional markers are required. It was proposed by NISSLER *et al.* (2016) to use smaller visual fiducial markers (AprilTags), and by NISSLER *et al.* (2017) to use a wider plain sticker to monitor the muscular activity. In this work, other markers were used, and a total of three different markers were proposed, examined regarding the sensor price-performance ratio, and eventually replaced by the other, as the project matured.

During the development phase, both sensor design and data processing routine were fully explored, as means to achieve a simple, comfortable, reliable, accessible and easy-to-use configuration. So far, all experiments were conducted on subjects with intact hands. It should be expected, however, of the developed sensor to work for amputees and individuals with limited hand-motion, since myographic sensors are able to detect motion intent, as well as the motion itself, as demonstrated by CHO *et al.* (2016), whose work successfully tested an FMG sensor on amputees and non-amputees.

It is expected of this project a substantial contribution to OMG sensor literature. As aforementioned, there is a lack of reference works regarding the subject, so it is expected that this report grows the reference pool, and popularizing, thus, the method. Its popularization would be very beneficent to the current state of the art of hand posture and gesture recognition, mainly due to its accessibility and comfort of use, since OMG sensors require nothing more than markers and an image acquisition device to operate correctly. Moreover, as it is expected of them to recognize motion intent as well, these sensors are potential candidates for prostheses or orthoses control, improving substantially the patient quality of life. The range of application of OMG sensors goes far beyond, however, and they can provide an easy, intuitive and inexpensive experience in HCI interfaces or remote control of mechatronic systems, for instance.

To conclude this chapter, an overview of this report is presented, as follows:

In Chapter 2, the current state of the art in hand posture and gesture recognition is presented with more details, and the muscles responsible for hand and wrist movements are quickly reviewed, yielding a better comprehension of human anatomy, muscles mutual influence and, finally, the sensor response in Chapter 4. At the end of the chapter, some image processing

techniques are surveyed, especially those relevant to the development of the sensor, regarding the data preprocessing and processing procedures.

The evolution of the sensor is presented in Chapter 3, comprising all three proposed sensor designs, their characteristics, what inspired them and the main considerations regarding them, as to keep or exclude them from the project. Also, the experimental setup for each design is described, along with the tested postures, organized into distinct experimental protocols, and the guideline of the performed experiments. Finally, the data processing routine following the acquisition step is detailed, from data extraction, treatment, and processing to posture classification, implemented in the latest design of the sensor.

In Chapter 4, it is demonstrated that an ordinary camera is sensitive to visual deformation of the forearm cross-section due to muscular contraction and relaxation. Indeed, the initial experiments revolved around this matter, since one of the challenges of the project was to find an adequate camera resolution for an acceptable price-performance ratio. Once found, the sensor performance in posture prediction was evaluated, and the corresponding results are shown subsequently. All experiments with the latest sensor design were performed on both anterior and posterior sides of the forearm, for it is the most stable of all three designs. So far, the experiments on each side were performed individually due to anatomical restriction, and the sensor performance is further discussed in this chapter for each side in predicting eight different postures.

Chapter 5 concludes the report with final considerations, comparing the developed sensor with other myographic state-of-the-art techniques, as well as prospects of application and further development.

Finally, the works published during the development of this project are listed in Chapter 6.

2 LITERATURE REVIEW

In the previous chapter, the importance of hand posture and gesture recognition for numerous application fields was presented as the motivation for several types of research, which can be grouped according to their sensing principle. Since the present project had as purpose the development of such a sensor also, the state of the art on the subject must be reviewed for a better understanding of the current trends. Therefore, this chapter is opened with a more detailed description of the five techniques introduced previously, beginning with the widely used glove-based and computer vision-based sensors, and concluding the Section 2.1 with the three myographic sensors, namely sEMG, FMG, and OMG. A special care was given to review both FMG and OMG sensors, since the present project endeavors to demonstrate precisely a hand sensor based on OMG, and LIQC, FEM/UNICAMP, has prior experience with FMG sensor to monitor hand movements also, so that the related works on both subjects are especially relevant to the proposed project.

Next, Section 2.2 introduces nomenclatures common in the study of human anatomy to descript body parts and movements, as to aid further discussions in Chapter 4. Also, a survey of the muscles responsible for finger, thumb and wrist motion was conducted and summarized regarding their name, location, and main functions. The survey was particularly relevant, as the understanding of the muscles is crucial to decode the developed sensor response and, therefore, to demonstrate its consistency.

Considering the nature of OMG sensors, the knowledge of computer vision is as important as of muscles location and operation in this project. Computer vision, however, is a broad field with numerous applications that may differ drastically from one to another, thus only the principal concepts are covered in Section 2.3, more specifically, the basis of object detectors. Given a target object such as, e.g., faces, cats, or simple geometric shapes, the detectors are designed to locate them in an image, and three common approaches to performing this task, namely color-based, shape-based and texture-based object detectors, are reviewed.

Noises, however, are inevitable in photos, including illumination, shadow, blur, and so on. Since noises are better when avoided, some image processing techniques for noise reduction are reviewed in Section 2.4. In this report, the review focused on morphological operations, including the famous dilation and erosion, opening and closing, and the morphological reconstruction. Naturally, there are several techniques for noise reduction other than morphological operation, nevertheless, they were selected for being suitable in terms of the sensor price-performance ratio and were effectively implemented in the software of the developed sensor.

2.1 Sensor for hand posture detection

Given the importance of human hand, a matter of interest is to provide its postures or, by extension, the performed gestures to computerized systems, which can use the received information for HCI, decision making, control loop, and so forth. This section is dedicated to an overview of the state of the art techniques for human hand monitoring, including the popularized glove-based and computer vision-based approaches, as well as the current myographic sensors. While some of them have already commercial versions for end-users, others are still in the development phase under controlled environments.

2.1.1 Glove-based sensor

Since their first appearance in the 1970s, the wearable glove-based sensors have been widely explored, therefore a number of different designs can be found. Despite their differences, these systems are fundamentally characterized by a group of sensing elements attached to a support glove and directly over the finger joints, a power supply circuit, and a system converting the joint angles into electrical or optical signal for posterior processing (WANG *et al.*, 2017).

The most iconic work is the Digital Entry Data Glove, designed by Gary Grimes and patented in 1983, and which commercialization boosted applied research on instrumented gloves and popularized them worldwide (DIPIETRO *et al.*, 2008). The Data Glove explored four types of sensor: two touch or proximity sensors to evaluate whether the thumb is touching fingers or another part of the hand, four "knuckle-bend" sensors to measure thumb, index and little finger flexion, two inertial sensors for measuring forearm twisting and wrist flexion, and two tilt sensors to measure hand tilt in the horizontal plane (GRIMES, 1983).

Following this general script, several instrumented gloves were designed over last few decades and a diversity of technologies were explored to measure joint angles. The Standard University CyberGlove, for instance, is equipped with 18 or 22 piezo-resistive sensors, distributed as follows: two or three flexion sensors for each finger in 18-sensors and 22-sensors model respectively, four abduction/adduction sensors, a palm arch sensor, and two sensors measuring wrist flexion and abduction/adduction (KESSLER *et al.*, 1995; NAIK *et al.*, 2014). Compared to Data Glove, CyberGlove targeted all fingers, measuring both flexion and abduction, and incorporated also the wrist abduction into its measuring range.

Other gloves, on the other hand, have been designed, and every effort was focused on developing a compact, simple, and accurate sensor to measure unconstrained finger motion. In

this sense, optical fibers, pressure sensors, flexible resistances, magnetic positioning sensors, air pressure and capacitive sensors, accelerometers, etc., have been explored to recover joint angles and each technology has its own advantages and drawbacks (DIPIETRO *et al.*, 2008; PARK *et al.*, 2015). For instance, resistance sensors are lightweight, flexible, inexpensive and suitable for wearable devices, but they are also prone to signal drift and instability. On the other hand, although magnetic sensors can provide precise hand data, they suffer from magnetic interference and require complicated models to convert the input signal into convenient angular information (WANG *et al.*, 2017). At last, fiber-optic sensors have a great advantage over the previous sensors due to their immunity to EM field, among other features such as, e.g., low cost, lightweight, and stability, these sensors have little portability due to the peripherals (FUJIWARA *et al.*, 2013).

Despite the individual drawbacks, glove-based sensors are recognized for their lightweight, easy use, flexibility, and accuracy, resulting in their widespread application (ZHOU AND HU, 2008). Wearing a glove, however, might be cumbersome or uncomfortable, and it can reduce user tactile sensitivity and even prevent interaction with capacitive touchscreens found in most smartphones or tablets nowadays (KIM *et al.*, 2012). The instrumented gloves present yet another limitation: they are user dependent. The research team in LIQC, FEM/UNICAMP, has observed during the development of an optical fiber-based glove sensor that the sensor performance varied slightly from user to user once the glove did not fit every user equally, and the transducers may be displaced from the optimum measuring point. Moreover, the glove garment should fit tightly the hand, or the sensor sensitivity would have suffered, such that it used to restrain hand motion (FUJIWARA *et al.*, 2014). One last observation about the subject is that all instrumented gloves can operate only when there is an intact hand. Thus, they are recommended for non-amputees and are suitable for rehabilitation purposes, but are not fit for prosthesis control, for example.

2.1.2 Computer vision-based sensor

In computer vision, motion capture (MoCap) refers to the process in which body movements are captured at some resolution (MOESLUND, 2000). For instance, a subject can be represented as just a moving center of mass or as a detailed entity with independent limbs. A common application for MoCap can be found in movies or animation to create natural and realistic computer graphic generated characters or movements. In particular, the human hand animation is not trivial, due to the complexity and the high number of degree of freedom (DoF) inherent to the hand. In such cases, MoCap focused on hand, whether supplied by instrumented

gloves or computer vision techniques, is handy in the sense that it provides complex movements with a natural look and less computational effort (WHEATLAND *et al.*, 2015). Since this project is about hand posture and gesture recognition, this section focuses specifically on computer vision-based hand MoCap techniques.

Visual detection of hand movements can be performed with active or passive elements. Active sensing places a set of devices, responsible for transmitting or receiving generated signals, on the targeted object and in the environment. Since the transmitter-receiver interaction is the basis of active sensing, this type of sensor is suitable for closed and controlled environments alone. On the other hand, passive sensing relies on signals found on the environment itself such as, e.g., illumination or EM waves, instead of purposely generated signals. In this case, known visual elements can be attached to the target object to facilitate the acquisition of passive signals. For example, by attaching green markers on finger joints and fingertips, it is possible to estimate the joint angle by tracking the visible markers belonging to the same finger. Depending on the surroundings, it might be easier to track the green markers rather than the fingers themselves. The use of external markers, however, restrains hand motion and is not always possible depending on the final application (MOESLUND AND GRANUM, 2001).

There are numerous alternatives to recover hand gestures from an image, and they can be grouped typically into two categories, depending on whether they have a prior knowledge of shape models or not (AGGARWAL AND CAI, 1999). Whichever technique is used, the raw image is submitted to a processing sequence composed by object detection, feature extraction, object tracking and then, action recognition (ITKARKAR AND NANDI, 2016). For instance, to recover the gesture information, the algorithm needs first to locate hand objects in the raw image and then, working on the segmented region of interest, it searches for features used to estimate posture parameters. The next stage is characterized by a tracking routine and, once the sequential posture data are gathered, the algorithm tries to recognize the performed action and concludes by assigning to it a semantic description, closing the general framework for human motion analysis (WANG *et al.*, 2003).

Usually, the hand object is extracted from the raw image based on the color information, given the characteristic color of human skin. This approach is very simple and convenient since it is processor friendly and is not affected by occlusion and orientation. But a problem arises when skin-colored objects other than the skin itself, such as sand, are in the image since the color-based detection relies on color cue alone. Another approach is based on the motion cue, which is quite effective in HCI, for instance. In this case, it can be assumed that only one subject is interacting with the system at any given time and that both the individual and the background remain relatively static; therefore, the main component of the motion in the image is usually the performed action itself. This method may fail, however, when multiple subjects interact with the system at once or the background is dynamic. Both approaches can be combined to boost

the detector performance (PAVLOVIC *et al.*, 1997).

The object detector plays an important role, since it defines the region of interest for the subsequent processing routine, and everything else will be excluded. The hand posture is described by a set of features defined accordingly to the adopted hand model. In the cases where there is no explicit hand model, 2D simple and low-level features are adopted since the algorithm does not know what is a human hand (GAVRILA, 1999). On the other hand, when an explicit hand model is available, the algorithm endeavors to match the region of interest to the predefined model, which could be 2D or 3D. The former is also known as appearance-based model, since it is defined by how the hand looks on the plane, regardless of the 3D configuration. In this case, markers can be used to recover the posture information, as well as geometric parameters such as e.g., perimeter, convexity, surface, centroid, rectangularity, etc. The 3D models, on the other hand, are further classified as 3D textured volumetric, 3D geometric or 3D skeleton models, and each of them describes the human hand at some degree of simplification. Whichever is adopted, the model parameters are redesigned whereas the spatial parameters of the performed gesture are calculated to match the hand model (ITKARKAR AND NANDI, 2016).

Nowadays, hand MoCap techniques are widely explored in several fields, such as virtual reality, advanced user interfaces, motion analysis, medicine and even to establish a geometric hand model useful in computer graphics (RAUTARAY AND AGRAWAL, 2015). Moreover, with the growing HCI field and advancing computer technologies, the tendency is that machines will have more and more natural and intelligent skills to interact with humans, and the hand plays an essential role in this interaction (YEO *et al.*, 2015).

In spite of the popularity of vision-based hand MoCap techniques among several applications, they are unsuitable for prosthesis control in cases of limb or hand loss, mainly because they rely on an intact hand to operate and are deficient in detecting motion intent. Moreover, single or multiple image acquisition devices are required, reducing thus the system portability, since the measurement is possible only inside the range of view defined by the position of the cameras. It is possible, though uncomfortable, to couple a camera on the user, but the limited field of action remains, besides the whole apparatus could not be used in public due to its configuration, and it also causes fatigue after a prolonged time of use (KIM *et al.*, 2012).

2.1.3 *Myography-based sensor*

Myography-based sensors recover hand posture by monitoring muscular activity, and they can be grouped into distinct categories depending on their sensing principle. In this report, three categories of myographic sensors are reviewed, namely sEMG, FMG, and OMG. All

three measure different variables, however, derived from muscular activity, thus the system setup of each differs from one another, as depicted in Figure 2.1. Compared to glove-based or vision-based sensors, the myography-based ones have a great advantage of being able to detect motion intent, and not the performed motion alone, since they focus on the muscles instead of the hand. Therefore, these sensors are suitable for both amputees and non-amputees, as well as rehabilitation patients (RADMAND *et al.*, 2016).

sEMG is the most popularized one in the myographic sensors family and can be found in most commercialized prostheses myoelectric control (GIJSBERTS *et al.*, 2014). The sensing system correlates a hand posture or action with an electromyography (EMG) signal pattern, which is generated by the sum of action potentials produced to contract the muscles (FANG *et al.*, 2015). These signals can be measured non-invasively at the skin surface or invasively with intramuscular recordings. The invasive measurement requires needles and wires, which can lacerate muscle tissues, causing pain during motion and requiring a medical crew to install them (HUG, 2011). On the other hand, the non-invasive sEMG acquires the EMG signals by means of an array of electrodes strategically placed over the skin surface of the forearm.

For instance, two electrodes placed over antagonistic muscle pairs of the residual forearm is a common configuration found in most commercialized prostheses on/off myoelectric control with a single DoF (DALLEY *et al.*, 2012). Whereas dexterous prosthetic hands with multiple DoFs have been developed, however, a single pair of electrodes has proven to be unable to provide an intuitive and satisfying control interface for a highly articulated hand. Other than designing robust learning algorithms to correlate multiple grasps with the corresponding sEMG signal pattern, it has also been proposed multichannel sEMG sensor configurations, with multiple pairs of electrodes placed over the forearm, to provide a full multigrasp control. Nev-
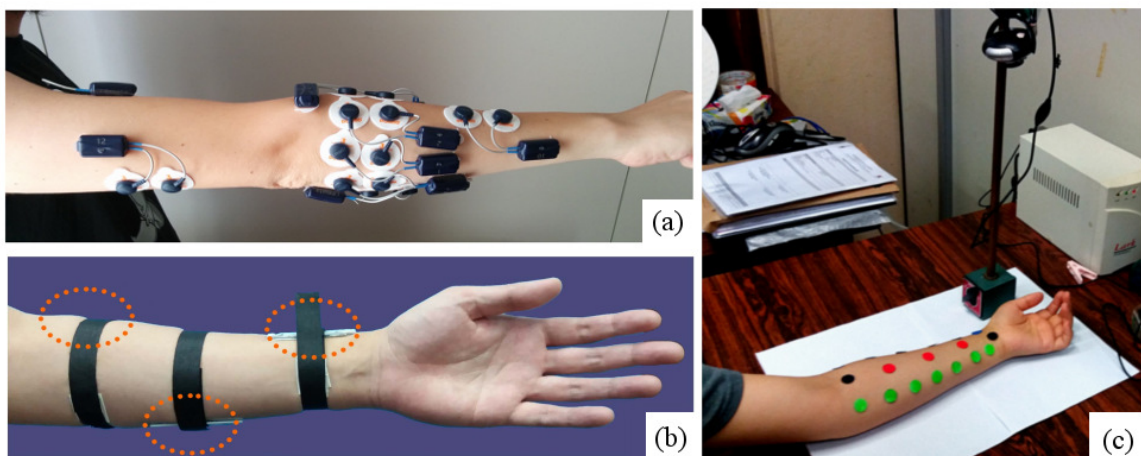

Figure 2.1: Examples of myographic sensors: (a) sEMG (PIZZOLATO *et al.*, 2017); (b) FMG (FUJIWARA *et al.*, Accepted for presentation); and (c) OMG (WU *et al.*, Accepted for presentation).

ertheless, this configuration increases system cost, size, and complexity, and is impracticable as prostheses control interface if the user residual forearm has little skin surface (FUKUDA *et al.*, 2003; TAVAKOLI *et al.*, 2017).

There are yet other factors contributing to the little acceptance of myoelectric control interfaces, clinically and among the users. For instance, sEMG signals are non-linear and quasi-random by nature, dependent on individual anatomical and physiological characteristics, susceptible to electrodes inherent and external EM noises, including the power used to run the sensor, not to mention the crosstalk among muscles. In addition to that, sEMG degrade over the time and with eventual electrodes displacements, contributing to the poor efficiency and reliability of the myoelectric control interface (CHOWDHURY *et al.*, 2013; FARINA *et al.*, 2014).

An arising member of the myographic sensors family is the FMG, also known as pressure distributing map (PDM) or residual kinect imaging (RKI). Unlike sEMG sensors, which measure myoelectric activities, FMG correlates the performed action or posture with the pressure signature, i.e., the radial deformation verified on the forearm due to the muscular contraction, the movement of the bone, and even the displacement of underlying soft tissue (WININGER *et al.*, 2008; LI *et al.*, 2012). Since FMG is based on a mechanical approach rather than an electrical one, the measurements are immune to EM signals, variation on skin impedance due to sweating, and crosstalk among muscles, thus they are a promising alternative compared to sEMG (JIANG *et al.*, 2017).

Various works on the subject have followed the pioneer, presented by PHILLIPS AND CRAELIUS (2005), in which the forearm pressure signature is measured with pneumatic pressure sensors attached on a customized silicone sleeve, and meaningful tests regarding the finger flexion/extension volition were conducted. Instead of pneumatic sensors, a common FMG sensor configuration adopts an array of force sensing resistors (FSRs), distributed around the forearm by means of an external wearable structure, which could be either a strap, a sleeve, a bracelet, or a socket (XIAO *et al.*, 2014; FERIGO *et al.*, 2017). For instance, LI *et al.* (2012) has demonstrated a feasible FMG system using an array of 32 force sensing resistors (FSRs) to classify 17 finger postures in non-amputee subjects. Meanwhile, CHO *et al.* (2016) conducted meaningful tests on impaired subjects, on both sound and residual forearms, to classify 6 to 11 combinations of grips using an array of eight FSRs, and concluded that it is possible to differentiate up to six primary grips important to daily tasks.

An alternative configuration for FMG sensors is based on optical fiber pressure sensors, which have yet the advantage of immunity to EM noises in comparison with FSRs (FUJIWARA *et al.*, Accepted for presentation). As a matter of fact, the team in LIQC, FEM/UNICAMP, has evaluated such configuration to classify 11 static postures by adopting four optical fiber pressure sensors to monitor sequentially a pressure map composed by 24 spots, including both front and back surfaces of the forearm (FUJIWARA *et al.*, 2017). The sensor was successfully

demonstrated on non-amputee subjects, with an average classification accuracy of ∼89.9%, but it should be expected of it the detection of motion volition on impaired individuals as well, as long as they have suffered neither muscle injury nor loss (Castellini *et al.*, 2014).

Besides the immunity to EM noises and physiological issues after prolonged time of use such as, e.g., muscle fatigue and sweating, FMG sensors generally provide other advantages as lightweight, low cost, easy biomimetic use and, of course, the ability to detect both the performed motion and the motion volition as well (Cho *et al.*, 2016). Despite the advantages, the user still must wear an additional structure to hold all force sensors in place, and it can be assumed from the sensing principle that the user will feel the individual sensors on the contact point with the skin. Besides, to improve the overall performance, preload is applied over the pressure sensors to operate in the linear range (Li *et al.*, 2012; Fujiwara *et al.*, 2017). Therefore, though it is very easy to put the FMG sensor on and off, it is cumbersome and uncomfortable after a prolonged time of use.

OMG, on the other hand, provides substantial comfort during the use as a combination of myographic and computer vision techniques. As such, the sensing system must have an image acquisition device, but unlike the ones described in Subsection 2.1.2, the camera here points to the forearm rather than the hand (Wu *et al.*, 2016). Naturally, there are intrinsic muscles responsible for hand motion, but they are kept out of the sensor scope since the goal is to estimate hand posture and motion intent by analyzing only the extrinsic muscles.

This technique of OMG was first proposed by Nissler *et al.* (2015), which presented a system using a simple webcam to monitor the visual deformation of the forearm anterior surface due to hand posture variation on intact individuals. Ten visual fiducial AprilTag markers were distributed along the visible forearm area in two rows of five markers each, and the test subjects were asked to perform a set of previously defined postures, with the forearm fixed on a support with Velcro straps, whereas the camera filmed the forearm with all markers. Next, markers location and orientation are extracted via offline procedure and related to the corresponding posture via a Ridge Regression. This first proposal of OMG system was tested for four postures, presenting an average classification error of ∼0.17%.

The same experiments were repeated in Nissler *et al.* (2016), and the sensor performance was compared for both linear and non-linear machine learning approaches. The non-linear model presented a slightly better performance, with an average classification error of ∼0.14%, being thus comparable to more mature methods as sEMG and FMG (Ravindra and Castellini, 2014).

Nissler *et al.* (2017) proposed yet a variation of the sensor design with the camera appended to the user limb. Instead of multiple AprilTag markers, the new setup adopted a single rectangular plain sticker as OM transducer, which covered approximately a third of the intact forearm surface. In this context, both slight position changes and bounding box deformations of

the sticker are related to the corresponding posture via a convolutional neural network (CNN). The sensor performance was then evaluated in terms of classification of five different finger configurations, namely four static postures plus the rest position with all fingers relaxed, reaching an average accuracy of $\sim$97.8%.

The results regarding the OMG sensor, as presented in all three related works from DLR, showed that the method has a comparable performance as sEMG and FMG to classify similar finger postures (RAVINDRA AND CASTELLINI, 2014). Nevertheless, further experiments with dynamic gestures are required, since practical applications hardly will have the user forearm fixed in one place during the whole time. The sensor performance in such closer-to-reality situations is still unknown, as no related work has been reported yet for OMG systems.

As for the sensor design, OMG is characterized by its low cost, accessibility and ease of use, though both setups presented by the team in DLR had Velcro straps fixing the user forearm on a rigid support. In addition to that, the OM transducers apparently do not need to be attached precisely to specific points of the skin. Despite the simplicity and convenience, OMG presents little portability due to the camera position. Even if the camera is fixed on the limb, the final setup is awkward, the device hinders hand motion, and the setup applicability depends still on the size of the residual limb.

Considering the available literature on OMG, the method is underexplored at the moment. As it has proven to be a potential alternative to other myographic sensors, it deserves to be better investigated, as to demonstrate whether it is suitable for practical applications.

2.2 Muscles controlling hand movement

The human hand is highly dexterous and has approximately 27 DoFs. Therefore, numerous postures can be performed as a combination of primary movements on each digit, thumb, and wrist. In this report, however, the wrist is kept at the neutral position in all experiments, i.e., it is aligned with both central axis and the horizontal plane of the forearm. In this sense, the proposed sensor was demonstrated to postures resulting from the combination of thumb and digits primary movements, which are depicted in Figures 2.2 and 2.3.

These movements, as well as all others, are controlled by a collection of muscles located in both forearm (extrinsic) and hand (intrinsic). Those found in the forearm are yet subdivided according to the side they are found, so they could be either in the anterior or posterior compartments (BOWEN, 1919). As shown in Figure 2.4, the anterior side corresponds to the underside, while the posterior corresponds to the back of the forearm. There are yet another two important nomenclatures for subsequent discussions: ulnar and radial side. The former corresponds to
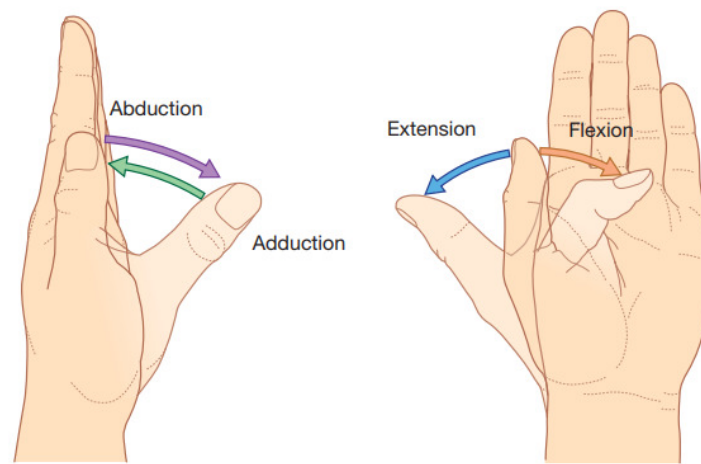
Figure 2.2: Thumb flexion/extension and adduction/abduction (DRAKE *et al.*, 2015).
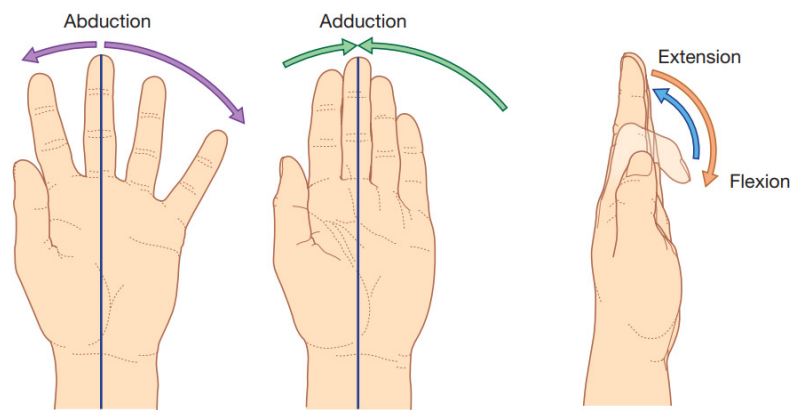
Figure 2.3: Digits flexion/extension and adduction/abduction (DRAKE *et al.*, 2015).

the region closer to the ulna (following the line of the little finger, all along the forearm), and the latter indicates the region closer to the radius (following the line of the thumb, all along the forearm).

The anterior compartment is composed of flexor muscles distributed over three layers, namely the superficial, intermediate and deep layers, whereas the posterior compartment holds extensor muscles distributed over superficial and deep layers. These muscles are responsible for wrist joint movements, flexion and extension of the digits and the thumb, and also supination and pronation movements (DRAKE *et al.*, 2015). The intrinsic muscles, on the other hand, are all found in the hand, such that there are out of OMG sensor scope, and are responsible for finer grip control of fingers and thumb movements. Meanwhile, the extrinsic muscles are mainly responsible for the power grip.

In this section, the intrinsic and extrinsic muscles are reviewed, regarding their name, localization and function, as presented in Tables 2.1–2.3. A note about digits extrinsic flexors and extensors is that they are responsible for all digits flexion and extension. One can still,

however, flex or extend fingers individually due to the four tendons going from the flexors and extensors to each finger, respectively, beginning at the lower fourth of the forearm. Each tendon is activated by separate groups of muscle fibers, allowing one to flex and extend the fingers separately as well as at once. However, since there are three fibrous bands connecting the tendons across the back of the hand, one cannot extend the fingers independently to the same degree as one could flex them (BOWEN, 1919).
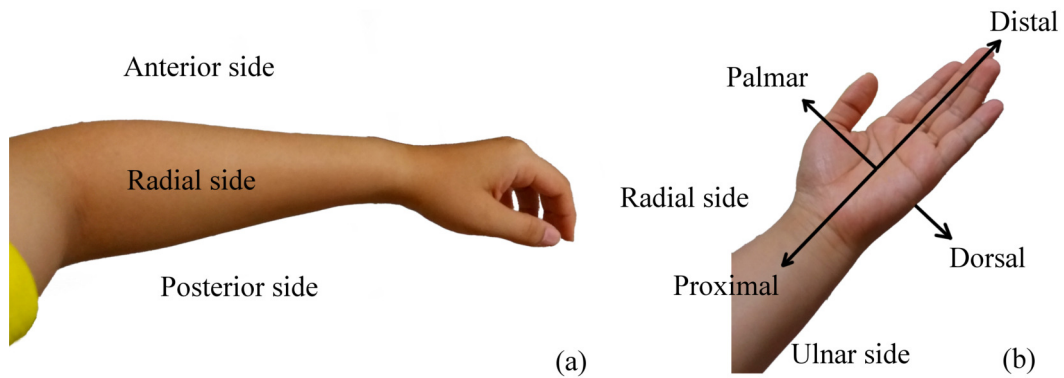
Figure 2.4: Anatomical nomenclature of location (a) in the forearm and (b) in the hand.

Table 2.1: Muscles responsible for wrist movements (adapted from BOWEN (1919) and DRAKE *et al.* (2015)).

Name	Location	Function
Flexor carpi radialis (FCR)	Extrinsic, anterior, superficial.	Wrist flexion and slight abduction.
Flexor palmaris longus (FPaL)	Extrinsic, anterior, superficial.	Wrist flexion.
Flexor carpi ulnaris (FCU)	Extrinsic, ulnar, superficial; it can be seen from both anterior and posterior view.	Wrist flexion; also contracts in voluntary wrist adduction.
Extensor carpi radialis longus (ECRL)	Extrinsic, posterior, superficial.	Wrist extension and abduction.
Extensor carpi radialis brevis (ECRB)	Extrinsic, posterior, superficial.	Wrist direct extension.
Extensor carpi ulnaris (ECU)	Extrinsic, posterior, superficial.	Wrist extension and adduction.

Table 2.2: Muscles responsible for thumb movements (adapted from BOWEN (1919) and DRAKE *et al.* (2015)).

Name	Location	Function
Extensor pollicis longus (EPL)	Extrinsic, posterior, deep.	Thumb last phalanx extension; it can extend other joints as well, drawing the thumb into the plane of the rest of the hand.
Extensor pollicis brevis (EPB)	Extrinsic, posterior, deep.	Thumb first phalanx extension. When the movement is strongly made, the whole thumb is abducted and wrist ECU comes into action to prevent wrist abduction.
Abductor pollicis longus (APL)	Extrinsic, posterior, deep.	Responsible for thumb metacarpal bone extension and abduction, it acts similarly to EPB over the whole thumb, but it pulls a little more toward the back of the hand.
Flexor pollicis longus (FPL)	Extrinsic, anterior, deep.	Thumb last phalanx flexion.
Flexor pollicis brevis (FPB)	Intrinsic, thenar eminence, intermediate.	Thumb first phalanx flexion; it is activated when one moves the entire thumb toward the little finger.
Opponnens pollicis (OP)	Intrinsic, thenar eminence, intermediate.	Thumb opposition (concomitant flexion and rotation, allowing thumb to touch digits fingertip).
Abductor pollicis brevis (APB)	Intrinsic, thenar eminence, superficial.	Thumb abduction.
Adductor pollicis (AP)	Intrinsic, thenar eminence, deep.	Thumb adduction.

Table 2.3: Muscles responsible for fingers movements (adapted from BOWEN (1919) and DRAKE *et al.* (2015)).

Name	Location	Function
Flexor sublimis digitorum (FSD)	Extrinsic, anterior, intermediate.	Second phalanx flexion; it also flexes the first phalanx and the wrist.
Flexor profundus digitorum (FPD)	Extrinsic, posterior, intermediate.	First phalanx, followed by wrist, extension. It extends the other phalanges when the first one is held flexed, but has little effect on them when the first phalanx or the wrist are extended.
Extensor digitorum (ED)	Extrinsic, posterior, intermediate.	First phalanx, followed by wrist, extension.
Extensor digiti minimi (EDM)	Extrinsic, posterior, intermediate.	Little finger independent extension.
Extensor indicis (EI)	Extrinsic, posterior, deep.	Index finger independent extension.
Four lumbricales (4L)	Intrinsic, palm.	First phalanx flexion; second and third phalanges extension.
Four dorsal interossei (4DI)	Intrinsic, back of the hand.	First phalanx flexion; second and third phalanges extension; fingers abduction.
Three palmar interossei (3PI)	Intrinsic, palm.	First phalanx flexion; second and third phalanges extension; fingers adduction.

2.3 Object detection

Object detection is a recurrent procedure in numerous applications in modern technology. The main goal is to find likely regions in the image, where a particular object can be found. Therefore, it is often the foundation of more complex problems such as, e.g., face recognition, pedestrian or vehicle detection for security purposes, automatization of the harvest routine, athletes tracking during a game, classification problems, and so on. As the basis of many procedures, some of them operating in real time, the time of response became an important factor, and special purpose detectors may be preferred over general ones (SZELISKI, 2010). Whichever is the case, most object detection algorithms are based on color, shape or texture information, either only one or a combination of the three features, as described in this section.

2.3.1 Color-based detector

Color-based detectors apply a threshold over the entire image, and all pixels satisfying the rule are selected (BOVIK, 2009). Thus, given a single channel input image I_{in} of size $n_r \times n_c$ and a scalar threshold Th, the detector outputs an image I_{out} of the same size such that:

$$I_{out}(x, y) = \begin{cases} 1 & \text{if } I_{in}(x, y) \geq Th \\ 0 & \text{otherwise} \end{cases} \qquad (2.1)$$

where $1 \leq x \leq n_r$ and $1 \leq y \leq n_c$.

Expanding the problem to three channel color images, the rule is still the same; the only difference now is that each pixel is a 3×1 vector instead of a single scalar. Thus,

$$I_{out}(x, y) = \begin{cases} 1 & \text{if } I_{in}(x, y, ch) \geq Th[ch] \text{ for all channels } ch \\ 0 & \text{otherwise} \end{cases} \qquad (2.2)$$

where $1 \leq x \leq n_r$, $1 \leq y \leq n_c$ and $1 \leq ch \leq 3$.

As it can be observed, the output value of a pixel is not influenced by neighbor pixels, but rather depends solely on its initial value. Therefore, it is very common for the output images to have a granulated look, such that it is usual to implement within these detectors a small region removal routine (THAKUR *et al.*, 2011).

Moreover, the working color space is an important factor in color-based detectors. The RGB color space is the default for most image processing tools; and, though it is quite alright to set a threshold as a triple of RGB values, this color space is not suitable for object detection. For instance, if two different pixels share the same color, but varies from one unit in just one channel, they will look the same for human eyes; however, if the threshold for that one channel cuts right on that one unit variation, a pixel will be set to "1" and the other to "0". Besides, the RGB values vary drastically with environment illumination, making it very difficult to set a static threshold rule (KAKUMANU *et al.*, 2007).

The simplest way to filter out the illumination is to convert the image from RGB to perceptual HSV color space. HSV stands for hue, saturation, and value. Hue specifies the color pigment, i.e., whether it is red, green, yellow, etc.; saturation defines how intense or grayed out is the color; and value controls how bright or dark the color looks (VEZHNEVETS *et al.*, 2003). Therefore, the illumination effect is concentrated on the value channel, such that if the main goal is to find a group of colors, it is enough to create a threshold for the hue channel. However, it is not always enough; one might be looking for a shaded blue spot in the image, for example, then a threshold for both hue and value channels should be specified.

2.3.2 Shape-based detector

Shape-based detectors select likely regions that contain the desired object by analyzing the geometric features of a group of pixels in the image. Therefore, the vicinity information becomes important and the detection principle is no longer pixel-by-pixel.

A classical shape detector is the Hough transform (HT), characterized by a voting procedure to determine the regions with the highest probability for a shape to be. For a given shape of interest, its main parameters are combined within a range of values, and each combination holds an accumulating variable, which counts the number of true pixels fitting that combination of parameter values. If an accumulator is above a threshold value, the pixels belonging to those parameter values are given as an object candidate. This method is already implemented in most image processing libraries for line and circle detection, besides being robust against image imperfections of digital images as non-continuity of edge representation (TRUCCO AND VERRI, 1998).

The HT algorithm, however, is not enough to detect more complex shaped objects as human face or medical images, which also vary in shape or appearance inside the target object class. For instance, human faces can change visibly due to expression, age, gender, and so on. In such cases, an alternative is to use deformable models, such as active shape model (ASM) or active appearance model (AAM) algorithms, to find the target object (SZELISKI, 2010).

Both techniques try to match iteratively a statistical model of the target object to the new image as closely as possible. The difference between them is that, whereas the ASM matches only the shape model, the AAM learns an appearance model by combining simultaneously the shape and texture information. Whichever is the case, the model is trained using a set of images with the feature points marked by hand, as exemplified in Figure 2.5 (COOTES *et al.*, 1999).
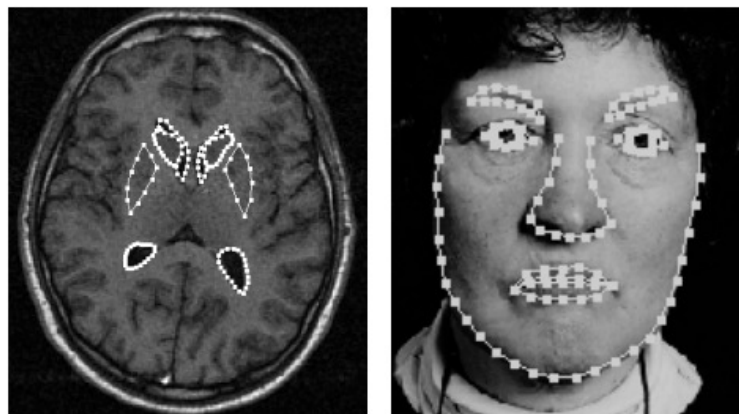

Figure 2.5: Example of the training set images with feature landmarks labeled by hand used to generate statistical shape or appearance models (COOTES *et al.*, 1999).

The model is matched to the new image with a two steps routine using the ASM. First, the shape model is matched to the boundary features in the image, followed by a texture reconstruction using an eigenface model in a shape-normalized frame. Meanwhile, the AAM proposes to fit a synthesized image as closely as possible of the novel image, considering also that the optimization problem is similar every time this task is performed, such that the similarities can be learned offline, increasing the algorithm performance and convergence speed even for problems with high dimensionality search space (COOTES *et al.*, 2001).

In comparison, ASM is faster than ASM but does not guarantee the optimal fit, since the reconstruction of the shape-normalized texture map from eigenface models depends on how well the shape model was matched to the novel image. On the other hand, AAM results in better matches with the new image the texture errors (COOTES *et al.*, 1999; COOTES *et al.*, 2001).

2.3.3 Texture-based detector

Visually, textures are related to image entropy. For example, homogeneous regions have low, whereas heterogeneous regions have high local standard variation. But this property alone is not enough to define a texture, because, for an image to be considered as a texture, it must obey some statistical properties with similar structures repeated over and over again with some degree of randomness (TUCERYAN *et al.*, 1993).

The group of pixels that are repeated in a texture image is called textel. Depending on the pattern, textels can vary in size, orientation, and randomness, as exemplified in Figure 2.6. Given an input image, it is possible to detect a particular texture by searching for the textels inside the image, a technique known as structural analysis (TOMITA *et al.*, 1982). Despite this approach resembling how human detect textures, it is not so trivial for machines to do so. Unless the pattern is very uniform, textels usually vary in size and orientation, hindering the identification routine. Moreover, depending on the type of input image, e.g. grass or hair, it is very difficult to define a universal textel that describes the pattern throughout.

While it is very difficult or almost impossible to detect textures in real images by the structural approach, it is easier to do so by statistical analysis. If the input image is colored, the first step is to flat it down to a single channel; next, an edge map is computed over the entire image; and finally, by counting the number of edge pixels inside a fixed-size region, one knows how busy or homogeneous is that region. Besides the edge density per unit area, edge direction also helps to define a texture. Notwithstanding this approach is not as intuitive as the structural analysis, it is applicable to all images and also computationally efficient. Moreover, by evaluating edges density and direction, it is possible not only to classify textures but also to
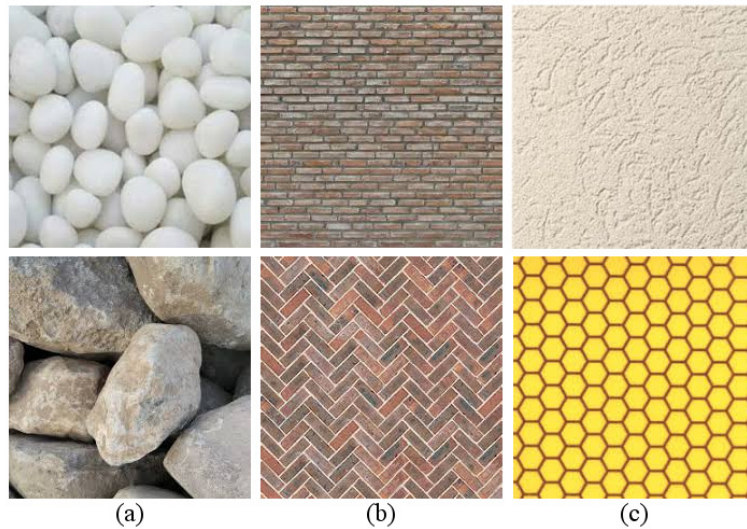
Figure 2.6: Examples of textels varying in (a) size, (b) orientation and (c) randomness.

perform texture-based image segmentation (SZELISKI, 2010).

2.4 Morphological image processing

Morphological operations refer to a number of image processing techniques used for image pre- or postprocessing, such as dilation and erosion, and region shape description, as thinning, skeletonizing and boundary extraction. Some of these operations are useful to remove noise, find image gradients, isolate individual elements and join disparate ones, find bumps or holes in an image, and so on, making them an alternative to image pre-processing (BOVIK, 2009).

The operation principle resembles that of 2-D convolution. Therefore, at least two inputs need to be informed, i.e., a structuring element B and the original image I_{in}. The latter is usually a binary image, or at most an intensity image, whereas the former is normally a rectangular, cross or ellipse-shaped matrix with the anchor set at the center. Many image processing libraries, however, allow new shape structuring elements to be defined, as well as the anchor location. Depending on the chosen operation, the output image I_{out} varies, but the principle behind is the same. As aforementioned, there is a resemblance with 2-D convolution, which lies in the fact that the anchor of B is looped over all pixels in I_{in}, and, for any original pixel $I_{in}(x, y)$, the corresponding transformed pixel $I_{out}(x, y)$ is given as the final value of the chosen operation applied over all neighbor pixels of $I_{in}(x, y)$ inside the region defined by B (GONZALEZ AND WOODS, 2007).

In this section, some of the main morphological operations are reviewed, especially those

used in the software development of the proposed sensor.

2.4.1 Morphological dilation and erosion

Dilation is a growing operation, as the name indicates. Given a binary or intensity image I_{in} and a structuring element B with a known anchor, usually its center, the dilated output I_{out} is defined as the local maxima of I_{in}. That is to say, each $I_{out}(x, y)$ is defined as the highest value of I_{in} inside the region defined by B when its anchor is located at (x, y).

On the other hand, erosion has an opposite effect, as it reduces the regions inside I_{in}. Just like how dilation can be seen as a local maxima operator, erosion can be compared to a local minima operator, for each eroded $I_{out}(x, y)$ is defined as the lowest value of I_{in} inside the region defined by B when its anchor is located at (x, y) (BRADSKI AND KAEHLER, 2008).

In both dilation and erosion operations, the structuring element B can assume any shape, however, it is common to use a rectangle, an ellipse or a cross, and by extension, a square or a circle. It is possible, though, to customize the structuring element in order to achieve the desired end effect (GONZALEZ *et al.*, 2003).

2.4.2 Morphological opening and closing

Morphological opening and closing operations resemble erosion and dilation, in a sense that they isolate individual or join disparate elements in an image. This effect can also be achieved via dilation and erosion, though they are not recommended, for the final region differs considerably from the original in terms of visible area. On the other hand, opening and closing keep the region size approximately constant before and after the transformation. The former first erodes the original image I_{in} by kernel B, and then dilates the eroded image by B, whereas the latter is the opposite: it first dilates and then erodes (SOILLE, 2013).

Figure 2.7 shows opening and closing operations graphically for better understanding. Suppose the region A with triangular peaks holds the true values inside I_{in} and the shaded ball B is the structuring element B. The complete opening region is defined as all pixels which B overlaps while translating inside A. On the other hand, the complete closing is defined as the complement of the image, formed by all pixels overlapping B while it translates outside and around the original region A (GONZALEZ *et al.*, 2003). Thus, as it can be seen from Figure 2.7, opening operation isolates and closing operation joins poorly connected elements.
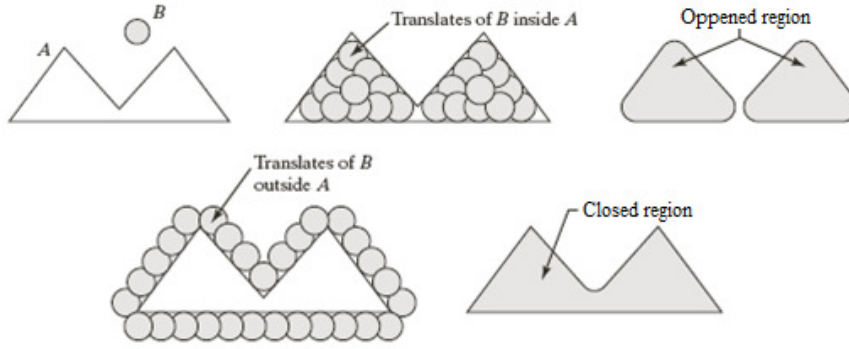
Figure 2.7: Morphological opening and closing principles (adapted from GONZALEZ *et al.* (2003)).

2.4.3 Morphological reconstruction

In morphological reconstruction, two input binary images and a single kernel are needed. One of the input images serves as the marker I_{marker}, the transformation starting point, and the other serves as the mask I_{mask}, which constrains the transformation. Meanwhile, the structuring element B defines the pixel connectivity during the operation (GONZALEZ *et al.*, 2003).

The following steps define the morphological reconstruction:

1. Initialize a temporary image I_1 to be the marker image I_{marker};
2. Define the kernel B;
3. Set another temporary image I_2 as:

$$I_2 = \text{dilation}(I_1, B) \wedge I_{mask} \qquad (2.3)$$

where $\wedge$ defines a pixel-wise AND operation;

4. Copy value of I_2 to I_1 if they were different;
5. Repeat step 3 and 4 until $I_2 = I_1$.

Because of the pixel-wise AND operation in step 3, the marker image must be a subset of the mask image, or else all connected regions will be removed from the output image. Moreover, as shown in Figure 2.8, morphological reconstruction can be used to select regions with desired features by defining a suitable marker image (SOILLE, 2013).

This property is very useful in color-based detection routines, which are based on pixel-wise thresholding, as already mentioned in Section 2.3.1. Besides the granulated appearance of the output image, regions of color transition are not chromatically "pure" due to camera resolution. Suppose the scene is composed by a red ball and a yellow background, the pixels at the ball borders may seem orangish in the image, since the camera has neither an infinite color
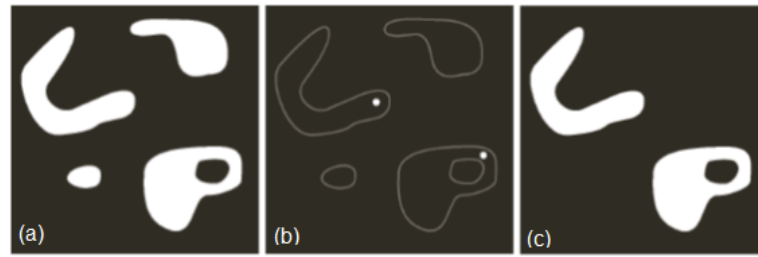
Figure 2.8: Morphological reconstruction (adapted from GONZALEZ *et al.* (2003)). (a) Original image (mask). (b) Marker image. (c) Final result.

range representation nor infinite pixels. Moreover, depending on the environment illumination and the device white balance, some background spots may become orangish in the image and be detected as a valid object according to the color detector rule. To solve this problem, it is possible to define the color threshold as a redder value, but paying the price of losing ball area, and the object may even be cleared by small region removal routine, disappearing completely from the output image. As an alternative, it is more efficient to implement two color detectors: one for orangish or even yellowish red colors and another for purer red colors. The former results in the mask image with lots of false positives, since the threshold range contains colors from pure red to yellowish oranges; while the latter results in the marker image with just the reddest regions, which can be found in the middle of the object. By reconstructing this mask with this marker, it is possible to remove most false positives from the background, keeping the region corresponding to the ball.

Morphological reconstruction is a very useful tool in image pre-processing. By setting a suitable seed marker image, it can also be used to fill holes inside connected components, clear objects touching image borders, filter elements skeleton, and so on (SOILLE, 2013).

3 MATERIALS AND METHODS

In the previous chapter, an overview of the current technology for hand monitoring was presented along with the considerations in choosing the OMG for the proposed sensor, as well as the muscles responsible for hand motion for their importance in analyzing the sensor behavior and response. Moreover, some image processing techniques were reviewed, especially the ones used to implement the sensor software to process the input data and retrieve the forearm visible deformation due to muscular activity.

Grounded in the literature review done so far, the present chapter details the proposed sensor from its design to the final posture predictor, following the steps shown in Figure 3.1.

The muscular activity is recovered by analyzing the visual deformation of the forearm and, since human skin is usually homogeneous, some external OM transducers are required to mark interesting points on the forearm. Therefore, the first step in developing the proposed sensor was to define an adequate transducer to provide enough reliable visual information about the muscular contraction and relaxation. Several setups for the OM transducer have been tested since the beginning of the project, three of which are detailed in this report for contributing significantly to the sensor evolution. As it can be further noticed, not all three designs have been tested to the same extent. For instance, the first one, which was based on the spectrum analysis of evenly spaced lines imprinted on the skin, has been discarded as soon as it was implemented, for empirical observations showed its impracticability even before posture classification experiments were conducted. This and other considerations, which led to the evolution of the sensor design, are detailed in Section 3.2 with further discussions.
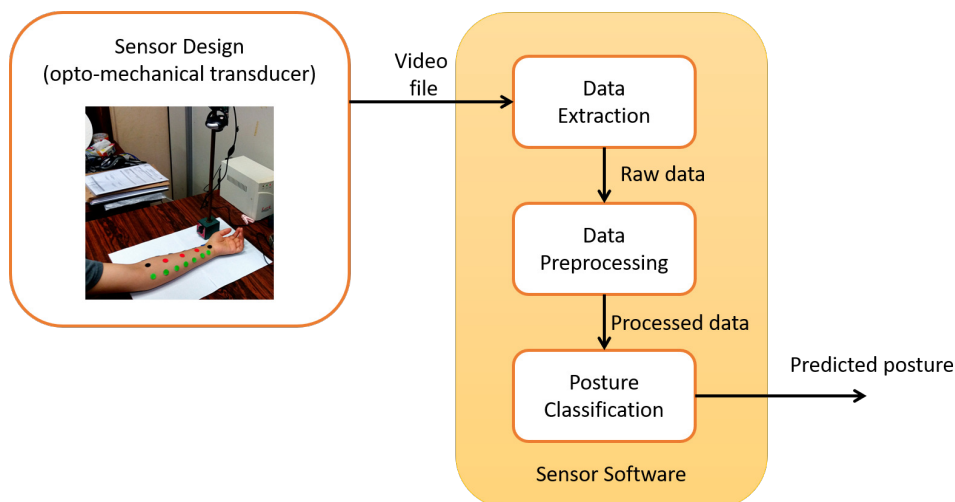

Figure 3.1: Overview of the developed sensor.

Once a suitable OM transducer has been defined, the sensor response should be evaluated in practical posture classification application, and a set of postures with a standard and reproducible protocol was therefore specified to validate the experimental results. During the experiments, the test subjects are asked to reproduce a sequence of postures in a predefined order, and the end video recording with the forearm visual deformation is then submitted to a series of processing routines, comprehending the steps two, three and four in Figure 3.1.

It is worthy to note that the experiment recordings contain all sort of information, namely the OM variable of interest and everything else, including the background and all surrounding elements. The first task of the sensor software is thus to filter out all noises and keep only the desired data to monitor muscular activity. Subsequently, the extracted raw data is submitted to a preprocessing routine to guarantee that they are suitable for the final posture classifier.

The data extraction and preprocessing routines depend on the sensor design since each of the three designs are based on a different type of OM transducer, meaning that the variable of interest varies for each design in the data extraction step. It is therefore natural to expect that the required preprocessing routine will also depend on the adopted sensor design because it would be strange to submit different variables to the same process. The particulars of the implementation are described in Section 3.3.

Finally, the posture classification is performed via artificial neural network (ANN) model, with sensor calibration and posture prediction, notwithstanding the representation as a single process in Figure 3.1. The adopted ANN model is further described in Subsection 3.3.4, comprising the considerations behind the chosen model and its features, concluding thus the present chapter and the sensor software routines.

3.1 Region of interest

Intrinsic and extrinsic muscles together are responsible for hand and wrist movements, but the former is out of the sensor scope in view of potential applications on individuals with limited or no movement of the hand. Hence, it is of utmost importance the exact definition of a region of interest (ROI) of work in the early stage of the project, and then the sensor design can be proposed within the established boundaries.

Theoretically, all skin surface around the forearm is part of the ROI, though only a fraction of it is visible to the camera in practice, whether it is the anterior, posterior, radial or ulnar side, due the reduction of the representation dimension from 3-D to 2-D space. In this project, only the anterior and posterior sides of the forearm are studied, so the ROI is either the region between the wrist and the cubital fossa or the one between the wrist and the elbow, respectively,
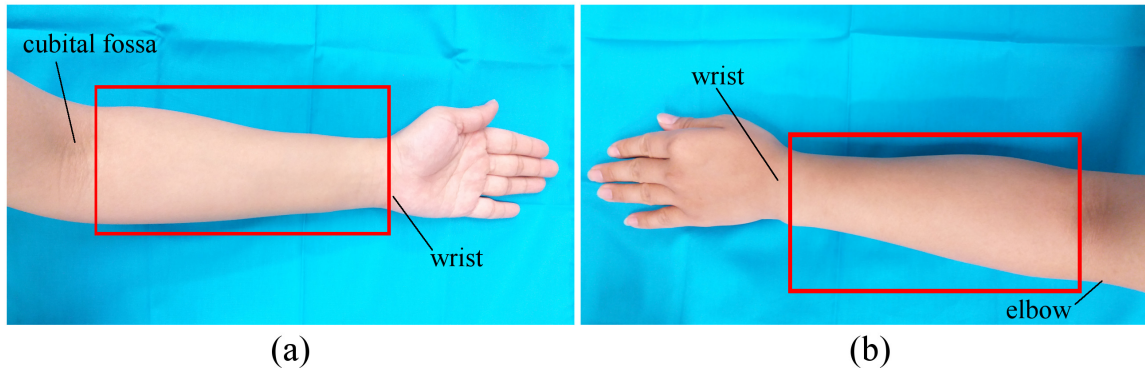
Figure 3.2: Forearm ROI in (a) the anterior and (b) the posterior side.

as shown in Figure 3.2.

3.2 Sensor design

OMG sensors are characterized by an image acquisition device pointed to the user forearm, usually with some type of OM transducers attached to the skin surface. A simple experiment consisting of nothing more than just watching the forearm while opening and closing the hand shows a perceptible variation on the borderline due to muscular activity, and the observed deformation on xy-plane, including the one in the z-axis, could be correlated to a given posture in the 3-D analysis. Nevertheless, cameras reduce 3-D real world to 2-D space and depth data is lost during the acquisition routine. If human skin had heterogeneous texture, it would hold naturally many interesting points and the depth data might even be recovered, but the truth, however, is that human skin is usually homogeneous. Therefore, any visual deformation in the image is perceptible only at forearm borders, which is why artificial OM transducers are attached on the skin, acting as interesting point landmarks to help identify deformation at the homogeneous region far from borders.

Besides the landmarks, all experiments were also conducted with a uniform background, whose color was chosen carefully to bear no resemblance with human skin, since the deformation at the forearm border will be lost if part of the background was misclassified as forearm element. The false detection of forearm surface may be particularly problematic depending on the adopted OM variable, as in the case of the sensor design B, detailed in Subsection 3.2.2.

Figure 3.3 shows the experimental setup for the last version of the proposed sensor for both anterior and posterior sides of the forearm. The experimental procedure is detailed in Section 3.4.
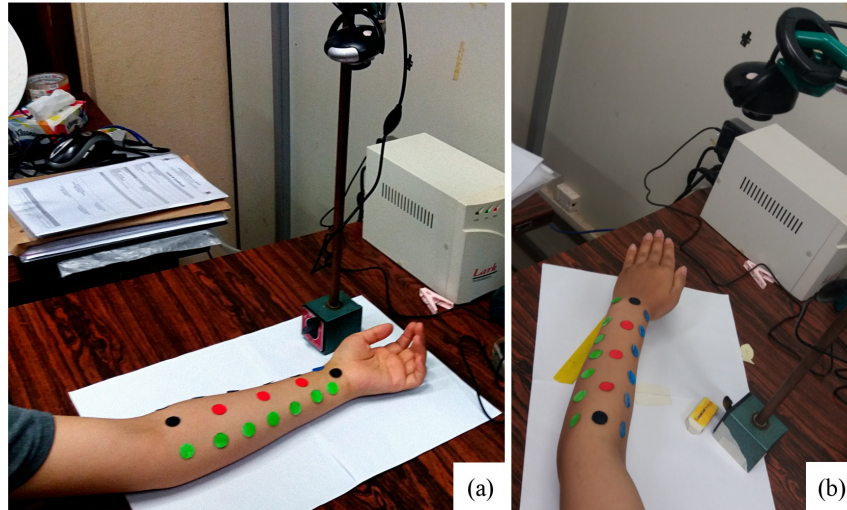
Figure 3.3: The experimental setup with the camera monitoring (a) the anterior and (b) the posterior surfaces of the forearm (WU *et al.*, Accepted for presentation).

3.2.1 Camera specification

Two different cameras were evaluated during the experiments, namely a smartphone built-in camera and an ordinary webcam, whose specification is indicated in Table 3.1.

Table 3.1: Camera specification.

Property	Smartphone camera	Webcam
Resolution	1280×720 pixels	640×480 pixels
Frame rate	30 fps	24 fps
White balance	auto/manual	auto

It is easier to set the webcam in the experimental apparatus; however, the preference was given to the smartphone built-in camera due to its better resolution and manual control of white balance. The automatic white balance used to change sometimes the appearance of the colors in the image along the experiments, leading to false detections and missing data, since the developed color-based object detector used static rules rather than dynamic thresholds. Besides, the first proposed sensor design required higher resolution for the proposed OM transducer was a set of imprinted stripes on the skin, and it was crucial that the stripes should be detected individually instead of a single entity. Even after the first proposal was proved impracticable, the second proposed sensor design was tested still with the smartphone camera at the maximum resolution. It was soon observed, however, that a resolution as high as 1280×720 pixels consumed too much from the memory and processing time. Thus, the resolution was reduced

empirically in search of a satisfactory price-performance ratio, and finally, it was found that the resolution of 640×480 pixels offered an acceptable data amount whereas being processor friendly. Grounded on this observation, the camera of the experimental apparatus was replaced by the webcam, since its inferior frame rate does not hinder the data amount and sensor performance. Regarding the automatic white balance, an additional step before each round of experiments was to certificate that the image is not tempered with a blue filter. The procedure is very simple; thus the choice of camera did not harm the sensor overall performance.

3.2.2   OM transducer designs

The proposed OM transducers are arranged chronologically and named alphabetically in this report to aid the description and further discussions.

Design A

The first endeavor to retrieve muscular contraction was by stamping a pattern of evenly spaced stripes on the forearm, inspired by spatial frequency analysis. Since the used pattern is uniform, it is expected to get specific frequencies by applying a 2-D Fourier transform over the acquired image; and they should vary with spatial distortions due to muscular activity.

To ensure sensor reproducibility, a stripe template was defined and is faithfully reproduced in Figure 3.4a (over a flat surface) and Figure 3.4b (over the forearm anterior surface). This design was submitted to initial testing with the smartphone camera set to maximum resolution, after which the following conclusions were reached:

1. Since forearm does not have a perfect rectangular shape, the stamped lines do not follow natural body curves.
2. The pattern is stamped on the skin with an easily washed out blue ink, such that the user can cleanse the forearm without stress afterward and the lines can be easily segmented from skin pixels by color information. However, once dried, the ink fades and merges with skin color, hindering the stripe pattern segmentation despite the camera high resolution.
3. To enhance the stripes pattern, an alternative is to add more ink on the stamp. Normally, human skin has fine wrinkles, and the ink spreads across these fine chaotic lines until it is relatively dry. Thus, whether one stamps the pattern with less ink, and the lines fade away, or one puts more ink, and neighboring lines will merge into a single line.

4. Depending on the amount of ink on the stamp, the lines can be bolder or thinner on the skin.

5. Over a flat surface, as shown in Figure 3.4a, the stripes seem to have enough gap in between. Nevertheless, it was very difficult for the object detection algorithm to identify the stamped lines as distinct entities in the image because the ink had the tendency to spread across the skin-colored gaps due to human skin natural properties such as, e.g., fine wrinkles and skin elasticity that can change with aging.

The observations above show the non-triviality of stripes segmentation when they are stamped on the skin, since the dried ink merged with skin color, and was interpreted as a dirty skin rather than an independent object. Moreover, the skin elasticity and brusque movements contributed to crack the dry stamped lines, such that fine blue particles could be found over the forearm; and the wet ink tends toward spreading across the fine wrinkles of the skin, reducing thus the contrast between the blue lines and the skin colored gaps.

Considering that the proposed OM transducer here is based precisely on spatial frequency analysis, it is of utmost importance the precise detection of each individual line. The initial tests, however, showed that this task was impracticable, even with a high-resolution camera, due to the low contrast between the skin and lines, which turned a region that is expected to be binary into a large blurry region. Therefore, the design A was rejected as soon as it was proposed, even before any standardized experiments with a well-defined posture protocol were conducted to evaluate the sensor performance in hand monitoring, and a new approach was proposed.
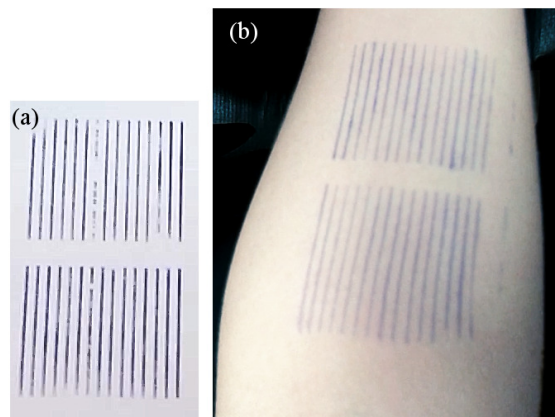

Figure 3.4: Stripe pattern for OM transducer design A on (a) paper and (b) forearm.

Design B

The design B proposes to track the muscular activity by counting all skin pixels in the image inside the ROI, as defined in Section 3.1. Hence, the ROI boundaries information becomes essential, as skin pixel beyond wrist and cubital fossa should not be counted to draw the function of skin pixels in terms of hand posture. Moreover, the forearm should be precisely segmented from background since imperfections as false skin detection and skin data loss can seriously change the results for the same posture. This level of pixel-wise precision is impossible to be achieved in digital image segmentation, so this design operates under highly controlled experimental conditions alone.

For instance, as shown in Figure 3.5a, the ROI is limited all around by distinctive blue background and green straps, both made of a minimally reflective material to avoid potential white spots due to light reflection under direct illumination. The experiments are conducted so that the camera sees only three colors, namely blue, green and nude, including any variants due to illumination condition, to facilitate the segmentation process: every blue pixel is classified as background object, while the green pixels are used to mark the head or end of a section on the forearm, and every other pixel is classified as skin pixel. This approach was possible because neither blue nor green components can be found in the human skin.

While the blue background isolates the forearm in the image, the green objects are used to divide the forearm into sections: the utmost right and left vertical straps mark the limits of the ROI, i.e., the skin pixels beyond are not counted, and the horizontal green tapes at the center of the forearm are used to divide the ROI into nine subregions, as shown in Figure 3.5b. The tapes are made of highly reflective smooth plastic, thus direct illumination must be avoided; otherwise, the skin pixel counter will have very different values for the same posture, characterizing a type of instrumental error.

Given an image with a known hand posture, the total of visible skin pixels inside the ROI

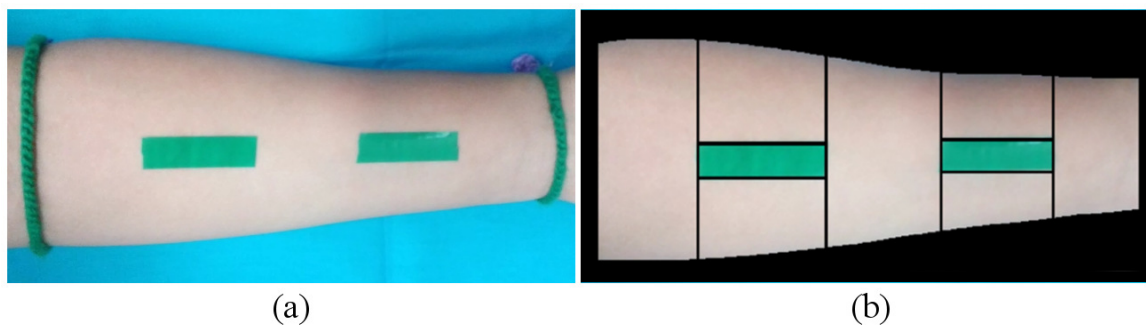(a)   (b)

Figure 3.5: OM transducer design B: (a) configuration and (b) segmentation of the ROI (WU *et al.*, 2016).

without sectioning should be enough to characterize the performed posture, as noticed at the early stage of experiments. It was necessary to divide the ROI into sections to boost the sensor performance, however, since the visual expansion near the cubital fossa could be compensated by the retraction around the wrist and vice-versa. If the variation suffered by the sum of all visible pixels were nil with the volume compensation, two distinct postures would be related to the same amount of skin pixels, reducing the number of postures the sensor can classify. By segmenting the ROI into nine sub-regions, instead of relating each posture to a single value, they are correlated to a combination of nine values, boosting the sensor performance compared to the other scenario.

Thus, special care must be given to control the experimental environment, especially to avoid direct illumination over the setup. Otherwise, white reflection spots can be observed on the surface of green tapes, reducing the area that the green color detector can identify. Depending on the size of the detected area, an entire tape could be overlooked at the worst scenario, and it would be impossible to align that frame with others, since there would be less than nine detected subregions and the frames are aligned according to the sum of visible skin pixels in each of the smaller regions. Even if the tape was not overlooked, the bad experimental condition can reduce the area corresponding to the tapes and increase the area of neighboring skin regions in consequence. Therefore, the sum of pixels inside the tape and surrounding regions will vary, not due to hand motion, but because of the bad experimental condition.

Compared to the previous design, the design B is more feasible and was submitted to some tests with five standardized static postures, as described in Section 3.4, to validate the OMG technique to monitor hand motion. Nevertheless, this design also has unstable performance that depends strongly on highly controlled experimental condition. Therefore, it can hardly be used on real applications, whose surroundings are tempered with several noise sources and, most importantly, is impossible to control the lighting condition. In consequence, a third and last OM transducer design was proposed.

Design C

The design C tracks external markers distributed over the forearm skin to monitor muscular contraction and relaxation rather than counting the number of skin pixels. Hence, its performance is more stable than the one of design B, which is susceptible to external conditions and depends on the precise segmentation of ROI and its subregions to reckon the skin pixels as clean as possible of instrumental errors.

In this case, the OM transducer is the colored markers, cut from paper and painted by hand. All markers are round with ~1.5 cm diameters, and there are four options of color,

namely blue, green, red and black. As aforementioned, human skin does not have blue nor green components, therefore blue and green markers can be easily detected. Regarding the red markers, though the human skin has a large amount of red component, hardly a person would have a pure $(R, G, B) = (255, 0, 0)$ red skin color under normal conditions, so the choice of red for visual markers does not hinder the processing routine afterward. Similarly, black color can be easily distinguished from a skin region, unless a heavy shadow is cast over ROI, a situation that is avoided by controlling environment illumination conditions. At last, the markers are covered with a layer of transparent adhesive tape, which protects them against moisture and allows them to be reused throughout experiments despite the resultant highly reflective smooth surface.

The markers are distributed over the ROI and attached to the skin with a thin layer of double-sided tape so that it is very easy to put them on the skin, as well as to take them off afterward. The markers follow forearm natural curves along three distinctive lines, one near the ulnar side, other near the radial side and a third one running through the center, as shown in Figure 3.6. Moreover, as it can be observed, the markers beside the radial and ulnar sides are attached almost right on the borderline between the background and the forearm, as to maximize the sensor sensitivity because border regions are richer of information.

In order to bestow more freedom of motion on the test subjects, the markers on the same line share a single color. This setup can be used to solve the orientation problem since a color marks the same group of muscles despite the forearm orientation in the image. Moreover, black markers are defined as local reference frame origin, which is useful to filter out forearm overall translation. For instance, each black marker in Figure 3.6 is the origin of a local reference frame, so there are three reference frames in total: the global XY (in red) and two locals xy_0 and xy_1 (in cyan).

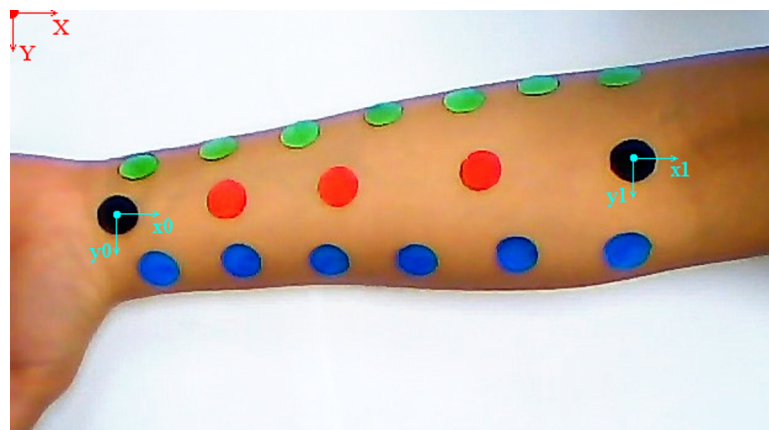Observing again the Figure 3.6, the design C dismissed the use of additional straps de-

Figure 3.6: A frame seen from the camera using the OM transducer design C with all three reference frame origins (anterior side) (WU *et al.*, Accepted for presentation).

limitating the ROI near the wrist and cubital fossa, as was the case of the design B, since it is no longer essential to recover all skin pixels. The ROI, however, still needs to be identified, though not with such a high accuracy, because the markers are by definition valid only if they are inside the ROI, explaining why the forearm is yet isolated from surrounding elements with a homogeneous background. In this case, rather than a blue one, a white background was chosen, for white has as much contrast with nude as blue, regarding both color and value data. Besides, since the white balance value of the webcam is set to automatic, the image color is less awkward with the white background and required less initial setups. A last observation about the background is that, unlike the case of design B, in which the background had to fill the whole field of view of the camera, the background here does not need to fill the video frame. It is enough to isolate the forearm from other surrounding elements such as, e.g., the table on which the test subject lies the forearm, and the sensor software can extract markers information inside the isolated ROI.

Finally, all proposed OM transducer designs are summarized in Table 3.2 comprehending the chief characteristics and drawbacks.

The main advantage of the design A is the sensing principle, based on 2-D Fourier analysis of a standardized stripes pattern. It was expected that hand motion information could be recovered via spectral analysis, but the design was proven to be impracticable due to the behavior of the ink on the skin. This elaborate design was then replaced by a very simple approach, based on skin pixel counting, characterizing the design B. Its main advantage was the simplicity and intuitive operation; however, it demanded severe restrictions on the experimental condition, and so the design B is unsuitable for real applications. Therefore, the design C was proposed as a solution to the problems found with design B such as, e.g., the high dependency of lighting condition and precise detection of skin pixels inside the ROI. In this aspect, the last design is indeed more robust to digital image processing imperfections, such as false detection or data loss due to color transition in object boundaries, illumination, moisture, and so forth. This is very important, and a significant advance compared to both previous designs since it would allow the sensor to be used in real applications. Because, if the developed sensor could not be used in real applications, it would be obsolete, and all the work would be wasted.

3.3 Sensor software

In the previous section, the sensor design was detailed in terms of the adopted camera hardware and the proposed OM transducer designs to monitor the muscular activity. In this section, on the other hand, the processing routine is presented, comprehending the block of

Table 3.2: Summary of the proposed OM transducer designs.

	Design A	**Design B**	**Design C**
OM variable	Evenly spaced stripe pattern	Sum of visible skin pixels	External colored markers
Operation principle	2-D Fourier analysis	Direct comparison	Marker position tracking
Background	Undefined	Blue monochromatic cloth	White monochromatic paper
ROI delimiter	None	Background plus green yarn straps	None
ROI divider	None	Green plastic tapes	None
Experimental condition	Undefined	Highly controlled	Controlled

sensor software in Figure 3.1, with all three stages of data extraction, data preprocessing and posture classification.

Because the proposed OM transducers have different sensing principles, the dedicated software to process the recorded data to predict the posture is different for each design. In this section, the stages will be described as individual subsections, and further subdivided according to the sensor design. It is worth noting that not all sensor designs have been developed for all stages of the sensor software branch; for instance, a specific data processing algorithm was not developed for the sensor design A, since it was rejected at the early stage of research.

3.3.1 Development tools

In total, three development environments were used to accomplish different tasks in the data processing routine, as summarized below.

MATLAB R2013b from Mathworks

The initial draft of all proposed sensor designs was tested in MATLAB to validate its practicability. The tool is dedicated to performing matrices operation efficiently and the built-in image processing library also offers several classical functions in digital image processing,

so that it is suitable for quick checks to validate whether a sensor design is practicable or not before spending more time developing a more sophisticated algorithm or improving the sensor hardware. In fact, all processing routine for the sensor design B and the data preprocessing for the sensor design C were conducted in MATLAB environment.

Microsoft Visual Studio 2017

MATLAB is efficient to perform matrices operation and can deal with common image processing techniques, but sometimes it consumed too much from the memory and processing time depending on the required functions. For instance, a video recording of less than one minute using the sensor design B consumed up to 1 GB of memory and the processing was too slow for future real-time applications.

To solve this problem, the alternative was to develop a more sophisticated image processing algorithm under Open Source Computer Vision Library (OpenCV) in Visual Studio, which was the case of the data extraction routine created for the sensor design C. In comparison, the OpenCV-based approach was much faster, demanded less from the processor and the extracted data consumed less memory.

Jupyter Notebook from Project Jupyter

The final posture classification is performed by an ANN sequential model created with Keras API with Tensorflow backend (CHOLLET *et al.*, 2015). The whole programming was done in Python in Jupyter Notebook environment, for the simplicity of data manipulation and visualization the tool provides.

3.3.2 Data extraction

At the end of each experiment, the recorded video file contains every bit of information as a sequence of frames, and not all data is required to the sensor. Thus, the recordings are first submitted to a data extraction routine that selects useful information accordingly to the type of sensor design in use.

Design B – Skin detection

The input of the design B is the number of visible skin pixels in the image. The data extraction algorithm, developed in MATLAB, is a two-step code: it first segments the ROI, then counts the number of skin pixels, as described below.

ROI segmentation

The whole background is blue by definition, such that no other object detached from the forearm region can be blue in a frame. As such, a blue color detector results in a binary mask M_{blue} with true value "1" for background objects, leaving out all foreground objects, namely the ROI, part of the hand and arm, two green yarn straps and two green tapes.

On the other hand, a green color detector is also applied over the original color image, resulting in another binary mask M_{green}. This one holds "1" value where both green yarn straps and both green tapes are located. Due to experimental setup, one can set the ROI boundaries along the x-axis by taking the rightmost and leftmost connected regions in M_{green} as delimiters. Hence, M_{green} is updated so that all columns from image left border till the rightmost vertex of the leftmost connected region are set to "1" as if they were all green, as well as all columns from the leftmost vertex of the rightmost connected region till image right border. The first modified region corresponds to all pixels beyond the cubital fossa, whereas the second one corresponds to all pixels beyond the wrist.

The resulting mask $M_{nonskin}$ with all pixels outside ROI plus the two subregions below the green tapes inside ROI is defined as follows, and can be fairly considered as the mask with all non-skin pixels:

$$M_{nonskin} \; = \; M_{blue} \; \vee \; M_{green} \tag{3.1}$$

where $\vee$ defines a pixel-wise OR operation.

Thus, the ROI with the skin pixels alone is defined by the complement of $M_{nonskin}$:

$$M_{skin} = \text{not}(M_{nonskin}) \tag{3.2}$$

A morphological hole filling operation over M_{skin} results in M_{ROI}, a mask with all elements inside ROI, including both skin and green tape pixels. A mask with the tapes can be found by:

$$M_{tapes} = M_{ROI} \wedge M_{green} \tag{3.3}$$

Then, M_{ROI} is segmented into nine subregions by setting to zero the columns corresponding to the rightmost and leftmost vertices and the rows corresponding to the topmost and bottommost vertices of both connected regions inside the updated M_{tapes}. Before segmentation, however, perimeter subtraction is applied over the mask M_{ROI} for a few times to remove occasional shadow at the connected region borders.

Figure 3.5b shows an example of ROI with all nine subregions already segmented and with their original color. During the processing routine, however, all images following the blue and green detectors are binary images, for it is cheaper and faster to work with logical operations.

Skin counter

The pixels inside each of the nine subregions represented in M_{ROI} are counted as

$$N_i = \sum M_{ROI}(x, y) \, if \, M_{ROI}(x, y) = 1 \tag{3.4}$$

where $i = 1, \ldots, 9$ and $(x, y) \in R_i$.

For subsequent processing routine, the absolute values N_i are yet divided by total visible skin pixels inside ROI, yielding some value between 0 and 1 (Equation 3.5). This number can be interpreted as how much of the whole ROI is taken by each subregion in a given frame. It is expected that, depending on the input posture, n_i will vary according to how skin region is visually arranged due to muscular activity. This variation, however, is not enough to compare different postures, since n_i is defined on common ground for subregions in the same image and not across frames. For this kind of comparison, in order to define a subregion growth and reduction behavior across different postures, an additional processing step is required, namely the data normalization. In this case, one is interested in how should n_i vary from posture to posture given a fixed subregion. So, the average n_i from all evaluated postures is normalized in between, setting them on equal ground and allowing the comparison of different postures.
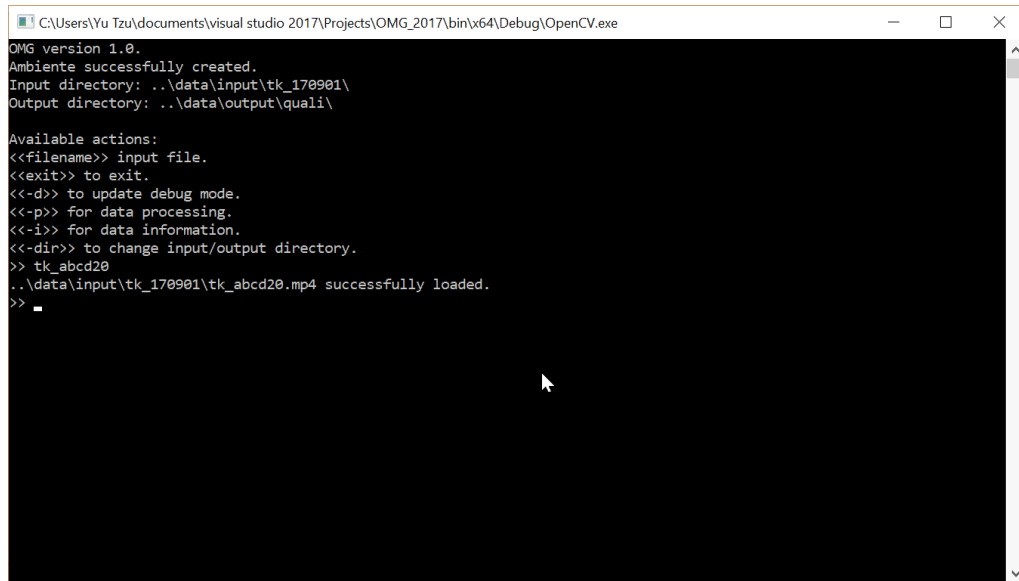
$$n_i = \frac{N_i}{\sum_{i=1}^{9} N_i} \tag{3.5}$$

Design C – Marker detection

The position of the markers is extracted offline with an image processing algorithm developed under OpenCV in Visual Studio, which is much faster, processor and memory-friendly when compared with the aforementioned algorithm, developed in MATLAB.

The processing routine is called with a simple and intuitive user interface (Figure 3.7), and its algorithm is essentially that of an object detector. The targets are the four color markers and the ROI skin pixels, which are differentiable by color, such that five color detectors with small region removal algorithm were defined.

False positives and noise, e.g. lighting condition, reflection and shadow, are considerably reduced with an image preprocessing routine consisting of morphological operations applied over the entire image. For example, morphological closing can remove white points due to

Figure 3.7: Screenshot of the user interface developed under OpenCV for marker extraction.

light reflection and morphological reconstruction can avoid false positives. However, though efficient, they are still not enough, so two more conditions were defined:

1. Markers candidates are valid only if they have neighboring skin pixels;
2. Border objects cannot be markers.

These two conditions result from the experimental setup, as explained in Subsection 3.2.2, since the whole arm should fit entirely inside the camera field of view, and markers are distributed over the skin area to track any visual displacement due to muscular activity.

In every frame, visible markers are located (Figure 3.8) and the following information is saved for each one of them: frame ID, centroid (X, Y) position and color information. These raw data are then submitted to a preprocessing routine as described in Subsection 3.3.3. The algorithm here described is summarized in the block diagram shown in Figure 3.9.

Regarding the extracted centroid position (X, Y), one might notice that each marker is reduced to one single pixel. Though such reduction simplifies subsequent processing routine, the markers are actually much bolder than a single pixel. Moreover, since the forearm is not a flat surface, markers normal vector is not always aligned to camera depth axis, so markers visible area is not always the same. In fact, the green and blue markers closer to forearm borders seem much more like ellipses than actual circles most of the times. Another contributing factor for differences in the detected area is the finite resolution of the camera, which gives transition regions a mixed color, combining both outer and inner colors, that can be out of the color detection range. On top of it, reflection white spot also hinders the marker detection, for white color is not included in any implemented detection rule. Though this last problem is relatively well solved by morphological operations, if the reflection spot should be wider, then it shall
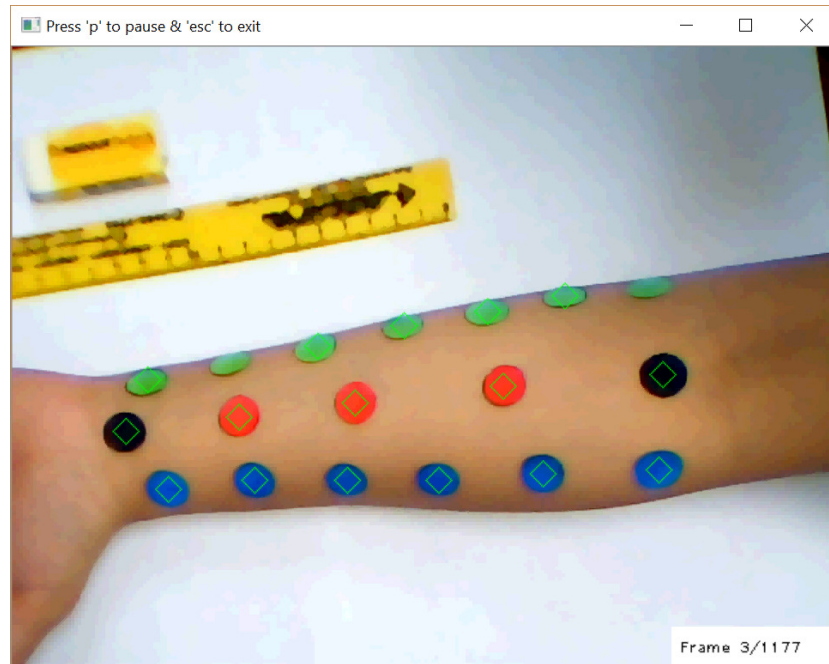
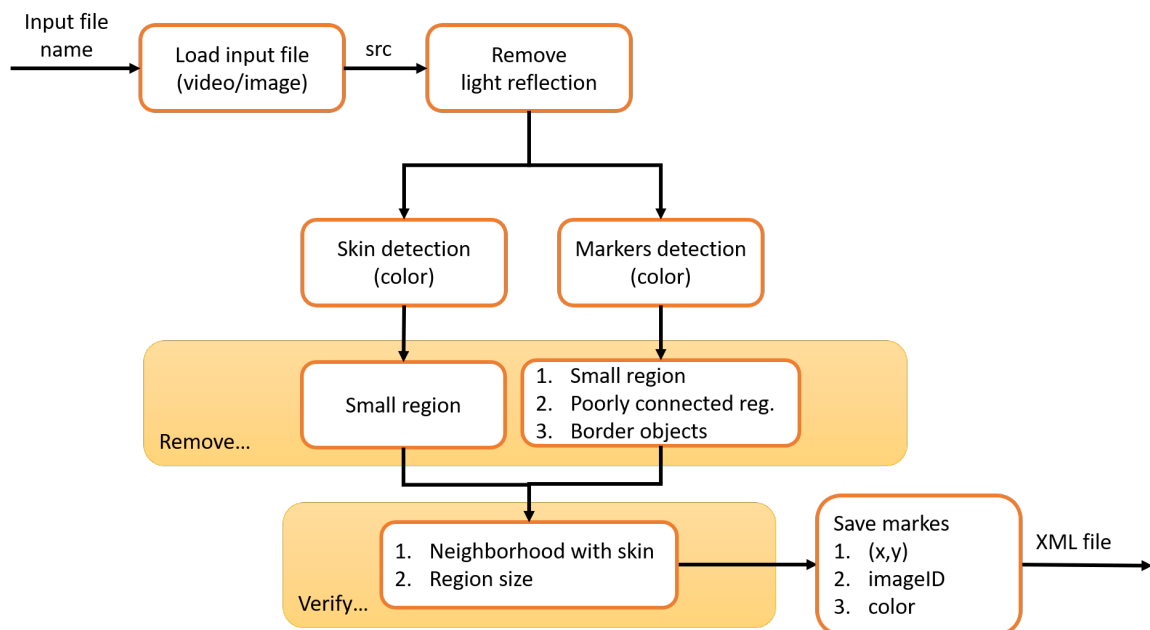Figure 3.8: Screenshot of a frame sample with the detected markers.

Figure 3.9: Procedures performed by marker detector algorithm developed under OpenCV.

remain in the image till the very end, and even markers on the center line, which mostly seem as round as possible, can lose their apparent roundness. In view of these points, (X, Y) does not match markers true centroid, calculated over their actual surface. Instead, (X, Y) is evaluated as the visible area centroid, i.e., given a connected region inside the image, that marker centroid is calculated over the visible connected component, nonetheless, it might not match the actual centroid.

3.3.3 Data preprocessing

Sensor design B requires no extra preprocessing routine. Given the simplicity of the input data, they can be compared directly as they are.

However, the same does not apply to the inputs from design C. Since it depends on marker tracking to monitor muscular activity, it is essential to attribute an ID to each marker on the forearm and update their position along the frames. Thus, a program implemented in MATLAB preprocesses raw marker data at the end of the image processing step described in Figure 3.9 with an object tracking algorithm, but before, position data are converted from Cartesian to the more straightforward polar domain, so markers can be given simply as a module and direction pair.

In the case of object tracking, some rules must be followed for a marker to be considered the same in consecutive frames. First, its definite color cannot change; but the code will allow a combination of an unknown with a known color. Second, the new position of the marker centroid must lie within a maximum range of $\rho \leq 10$ and $\theta \leq 10°$. If more than one candidate fit the above rules, then the closest one to the old position would be considered as the same marker. A note about these rules, however, is that the maximum allowed variation of 10 units for both ρ and θ variables is scale dependent. In other words, for most of the conducted experiments, this value was observed to be satisfactory to track the markers correctly, but since it is static, the tracking algorithm would fail if there was a significant scale variation, as zooming in or out.

In the end, a matrix with *valid frames* rows $\times$ *markers features* columns is generated. Moreover, the posture class of each frame is added manually to the final matrix and is subsequently used as ground truth for the supervised training of the ANN classifier.

Second and third matrices are generated from the first one by converting all samples to local reference frames with the origin set on each black marker, as described previously and shown in Figure 3.6. This approach removes global perturbation or subtle inherent vibrations or spasms of the forearm by compensating the uniform translation over all markers.

The Figure 3.10 summarizes how raw data from former marker extraction step are preprocessed till they are suitable for final classification routine.

3.3.4 Classification Routine

As already mentioned, sensor design B has undesirable drawbacks as dependency on how well the ROI was segmented. For example, if a part of ROI was cut out of a given frame, the
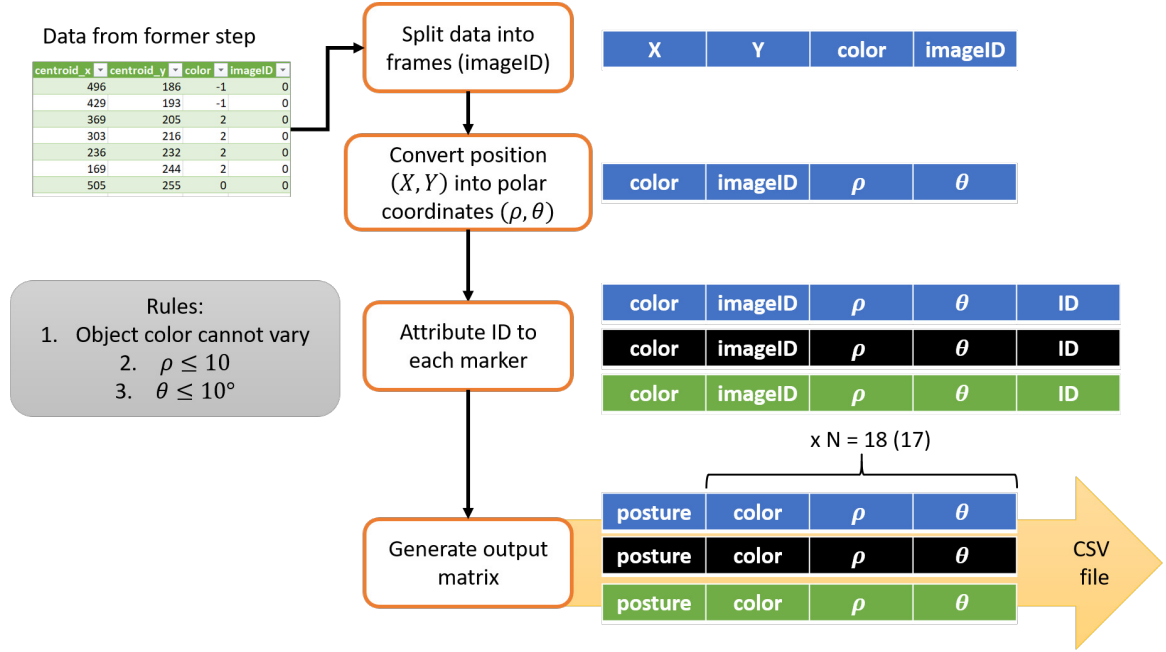
Figure 3.10: Procedures performed by marker tracking algorithm developed under MATLAB.

number of visible skin pixels will vary considerably in comparison with the previous or next frame, but the posture may be the same all along. This performance instability is not desirable so that further tests as posture classification were not conducted with the design B. Instead, they were conducted for the design C alone, using a sequential ANN model created with Keras API with Tensorflow backend (CHOLLET *et al.*, 2015).

The Figure 3.11 shows the chief properties of the implemented model. It has 54 or 51 input features, two hidden layers with 64 and 32 output nodes and dropout of 0.5 each, and an output layer with eight categorical outputs, corresponding to the total of eight monitored postures from protocol II, defined in Section 3.4.

The sequential model was adopted as an extension of previous works related to FMG sensors, in which sequential ANN models have shown high precision and performance to classify several hand postures with supervised training.

Regarding the input layer, the number of 54 or 51 features were calculated as $3N_{mkr}$, where N_{mkr} corresponds to the number of valid markers in the set and can amount to either 18 (global referenced set) or 17 (local referenced set; the local origin is not counted). The factor of 3 is also defined by the experimental definition, for the classifier is fed with three features per marker, namely one color and two position information, as described in Subsection 3.3.3.

The number of implemented hidden layers was defined empirically, as well as the number of nodes in each of them and the dropout value. The latter defines the fraction of neurons that will be turned off at random during the training step to improve the classifier robustness to eventual noises. In this case, for example, the dropout is set to 0.5, meaning that for each training round, only 32 neurons in the first hidden layer and 16 neurons in the second one are
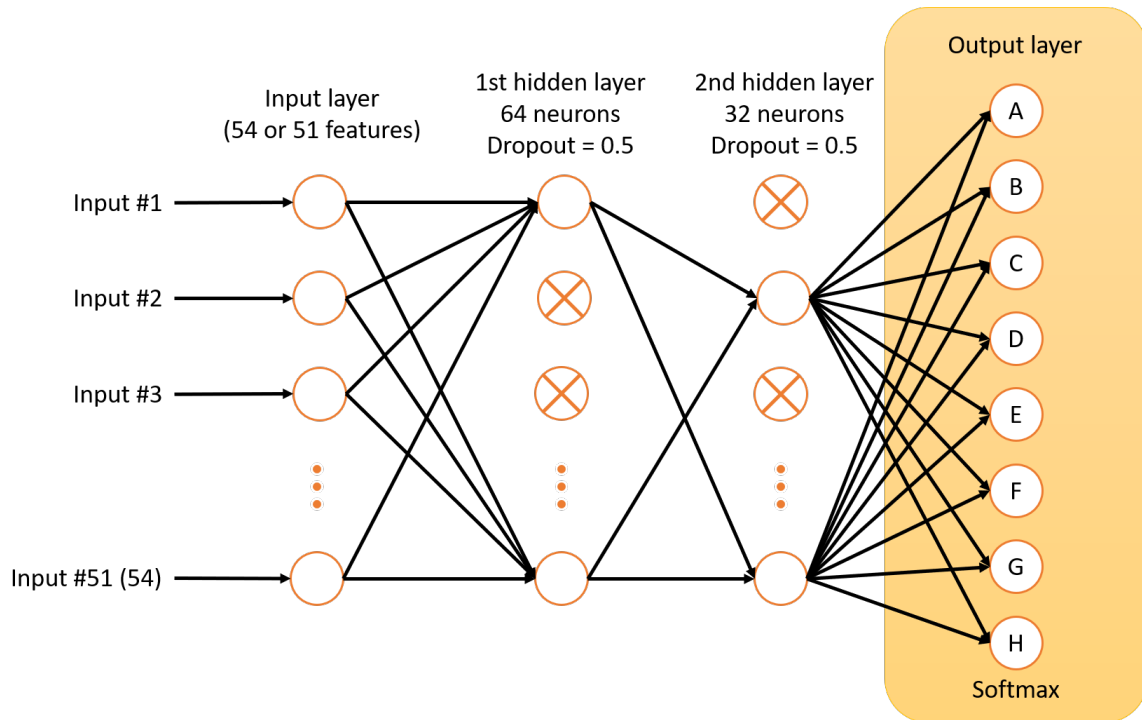
Figure 3.11: ANN classifier model.

active, while the other half is off. Because the choice of active neurons is random for each round, this approach serves to improve classifier robustness when all neurons were activated at the final calibration (KRIZHEVSKY *et al.*, 2012).

Besides the dropout, an early stopping was implemented to control model complexity and prevent it from overfitting (BISHOP, 2007). For instance, all experimental data are split evenly into calibration and test data sets, and the former is yet divided into training and validation sets. Whereas the ANN is trained on the training set, its performance is also evaluated on the validation data. Early stopping means that the model training will be halted if the model performance for validation points becomes worse than the one for training points, a situation known as overfitting. In this scenario, the model fits better to the training set but will behave horribly to unknown data. Hence, the calibration data here was yet randomly split into training and validation sets with a ratio of $0.9/0.1$ due to early stopping implementation.

The model performance, however, can vary according to samples inside each set, thus a 10-fold cross validation routine was also implemented, so all samples ($\sim$9000 points per subject and experiment) can be validation data at least once (RODRIGUEZ *et al.*, 2010). The ANN model final performance is given by averaging individual performances of all ten folds, per posture and also for all postures together, as well as per subject and for all subjects at once.

3.4 Experiment description

In this section, the experimental protocol and the set of tested postures used to evaluate the sensor performance, with the hardware and software described above, are presented.

During the experiments, the camera was set above and pointing perpendicularly to the forearm with the OM transducers attached to the skin. The test subject is asked to sit comfortably at the desk and to rest the forearm over the homogeneous background. When the recording begins, the user is asked to perform a set of predefined postures in rounds, holding each posture for a standardized time. All particulars regarding the conducted experiments for each sensor design are summarized in Table 3.3, with a note that the design A was not submitted to any elaborated experiment.

As described in Table 3.3, the sensor design B was tested on five postures defined with the protocol I (Figure 3.12). The postures included digits full extension /adduction and thumb adduction – open hand (P1), fingers and thumb full flexion – closed fist (P2), individual extension of index finger (P3) and both index and middle fingers (P4) with full flexed thumb, plus digits full flexion and thumb extension – thumbs-up (P5). The experiments were conducted with all five postures, shuffled in random and predefined order, and repeated twice. Each posture was held for $\sim$5 s, totalizing a $\sim$50 s recording per subject per experiment. Moreover, all experiments per subject were carried out in the same sequence to remove external variables, such as illumination, camera positioning, ROI dividers position, arm orientation, etc. The results of these experiments are presented in Section 4.1.1.

The sensor design C, on the other hand, was evaluated for eight postures, defined in the protocol II (Figure 3.13). This time, the postures were selected so that they complement each

Table 3.3: Summary of the conducted experiments.

Property	Design B	Design C
Camera	Smartphone camera	Webcam
Resolution	640 × 480 pixels	640 × 480 pixels
Frame rate	30 fps	24 fps
Total of tested postures	5 (Protocol I)	8 (Protocol II)
Tested postures per set	5	4
Repetitions per set	2	3
Duration of each posture	$\sim$5 s	$\sim$4 s
Video duration per experiment	$\sim$50 s	$\sim$48 s

Protocol I Postures
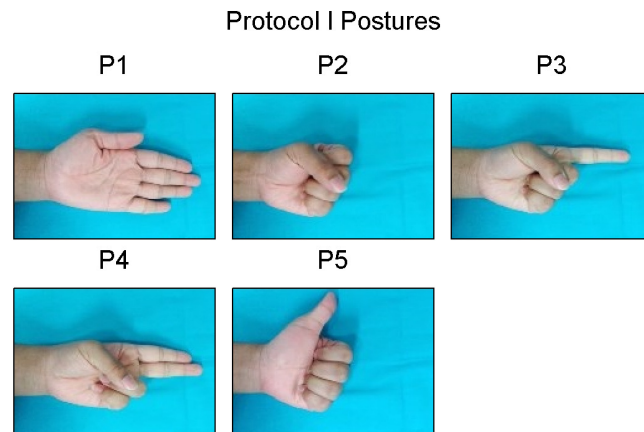
P1   P2   P3

P4   P5

Figure 3.12: Protocol I postures (Wu *et al.*, 2016).

other, i.e., they must be analyzed in pairs AB, CD, EF, and GH. Each pair activates or deactivates any number of muscles while locking all the rest, hence allowing the independent analysis of muscles responsible for a specific movement. For example, by comparing postures A and B, one analyzes the influence of muscles in thumb full flexion and extension when the fingers are kept extended and adducted. The postures C and D, on the other hand, investigate whether there is any difference in muscles contraction when the thumb is over or under fingers in closed fist configuration. The postures E and F, on the other hand, watch the influence of individual fingers flexion/extension, so that the thumb is kept fully flexed for both postures and only the middle finger varies from one to another posture. Lastly, the pair GH analyzes the thumb abduction/adduction, keeping the fingers fully extended and adducted in both postures, varying only the thumb.

The experiments for the sensor design C were conducted by selecting four of eight postures in protocol II, i.e., two complementary pairs, which are repeated thrice in a predefined order. Taking, for example, a set composed of postures CDGH, the four postures are repeated thrice in this order. Each posture is held for $\sim$4 s, totalizing thus a $\sim$48 s recording per subject and per experiment.

As it can be seen from Table 3.3, though the design B was tested with the smartphone camera, its resolution was downgraded to 640×480 pixels to provide a better processing time and memory consumption. Considering that the webcam is easier to fix on the experimental apparatus, the optimum resolution with the smartphone can be found on the webcam as well, and 6 fps is not a significant difference, the smartphone camera was replaced by the webcam in the experiments with the sensor design C, as shown in Table 3.3.

The duo protocol II and sensor design C was validated for both anterior and posterior sides of the forearm, gathering data from six and four different subjects, respectively, comprising a population with age between 23–35 years old, male and female. Due to markers size and
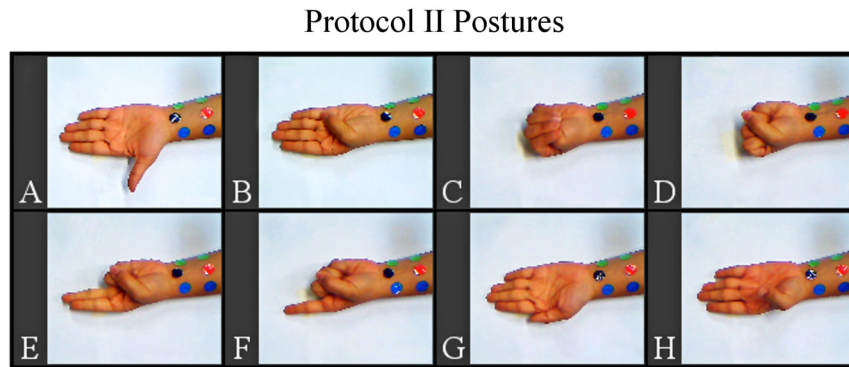
Protocol II Postures

Figure 3.13: Protocol II postures (WU *et al.*, Accepted for presentation).

difficulties in attaching three lines on each side of the forearm, neither ulnar nor radial sides were directly evaluated.

Both protocols I and II investigate fingers and thumb movements only, thus neither stated wrist movement from a posture to another. All thirteen postures kept the wrist in neutral position, i.e., neither flexed/extended nor abducted/adducted.

4 RESULTS AND DISCUSSION

In this chapter, the proposed OMG sensor is validated in terms of practicability, answering the initial question of whether it is possible to see and correlate the visual deformation on the forearm due to muscular activity with the corresponding posture. The experimental results related to the sensor validation were examined in the light of muscle anatomy, as reviewed in Chapter 2, and presented in Section 4.1.

Of the three proposed sensor designs, only the designs B and C were submitted to validation test, for the design A has shown to be impracticable in terms of subsequent data processing routine. The sensor B was successfully validated despite its high vulnerability to environmental conditions and dependence on precise image segmentation, two reasons why the design was after all disregarded.

Therefore, the results regarding the classification performance in Section 4.2 refer to the design C alone. In this case, the sensor was evaluated in terms of variability, reproducibility, and broadness, by setting the number of postures to eight and test subjects to ten. Moreover, the sensor was also tested for both front and back of the forearm, as to verify how the sensor would respond to the change of target muscles.

All experiments, for sensor calibration and performance examination, were conducted with respect to the experimental protocol presented in Chapter 3, guaranteeing the sensor reproducibility within the established experimental conditions.

4.1 Sensor validation

In this section, the experimental results were investigated in view of sensor validation, so the main analysis is to compare the OM transducer behavior for each tested posture. For instance, whereas the design B can be validated by a distribution map with the number of visible skin pixels inside each subregion of ROI, the design C compares the position of the markers and its variation.

4.1.1 Sensor design B

The results bellow correspond to the average values out of 46 experiments, following the protocol I described in Section 3.4. The subregions R_i are ordered as depicted at the bottom of Figure 4.1, which also presents the distribution map of the normalized n_i across all five postures.

The distribution map is analyzed in terms of the muscles involved in performing each posture. Then, it is important to identify which muscles are activated.

As shown in Figure 3.12, the posture P1 is characterized by full extension and adduction of the fingers plus the adduction of the thumb, as in an open hand configuration with all fingers and thumb pointing to the same direction. According to Table 2.2 and Table 2.3, a total of six muscles are activated: thumb extensors EPB and EPL, thumb adductor AP and fingers extensors ED, EI and EDM, all of them shown in Figure 4.2. As it can be observed, all activated extrinsic muscles are located at the back of the forearm, so none is directly viewed by the camera. As for the intrinsic muscle AP, it is also out of camera scope. By checking the n_i distribution over all subregions, there is a higher concentration of visible skin pixels around the center, corresponding to the bolder part of finger extensors and termination of thumb extensors. Nonetheless, the intrinsic AP muscle is out of sensor scope, its contraction must still cause any degree of perturbation on the wrist area, explaining the high value observed in R9.

The posture P2, on the other hand, is characterized by full flexion of all fingers and the thumb in a closed fist configuration with the thumb on top of the flexed digits. Thus, it contracts thumb flexors FPL and FPB, fingers flexors FSD and FPD, and also the intrinsic muscles 4L
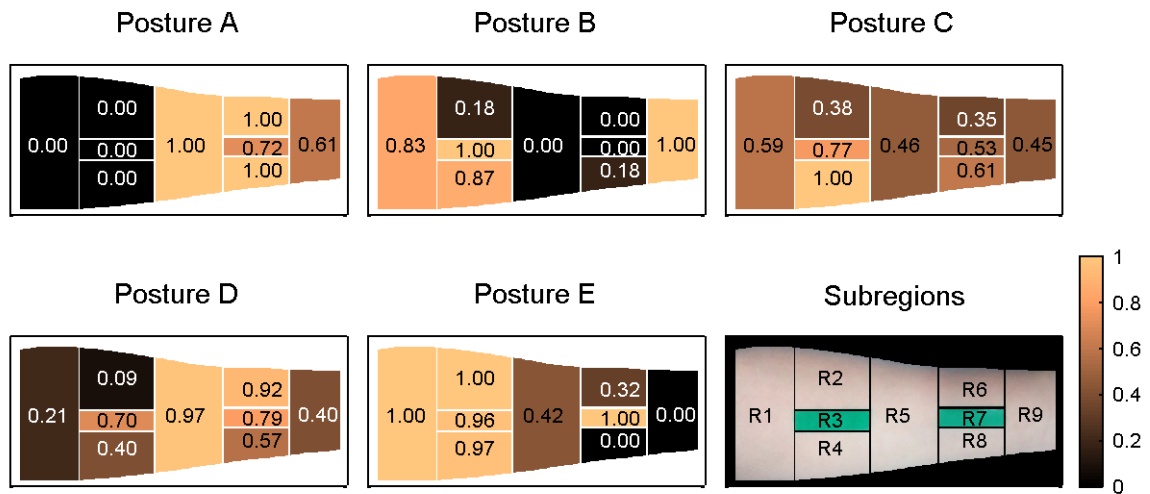

Figure 4.1: Normalized n_i of each subregion R_i for all postures from protocol I (WU *et al.*, 2016).

and 3PI, all of them are shown in Figure 4.3. All flexors are found on the anterior side of the forearm, and the natural tendency is to increase the forearm volume near the elbow when activated. This is the reason why there are more visible skin pixels in subregions R1, R3, and R4. However, the same is observed in subregion R9, which can be explained by the action of the three intrinsic muscles (FPB, 4L, and 3PI).

The postures P3 and P4 are intermediate states between postures P1 and P2 and activate the same group of muscles (Figure 4.4) with individual flexion of the index finger and the pair index-middle fingers, respectively. Since the fingers are individually flexed or extended, both
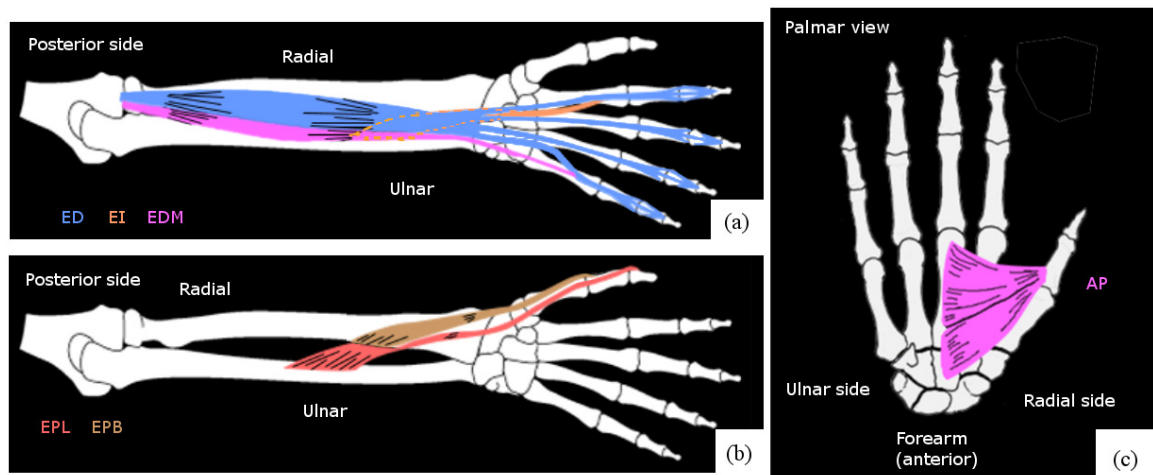
Figure 4.2: Muscles activated in posture P1 by (a) fingers extension, (b) thumb extension and (c) thumb adduction (adapted from LOUDON (1997)).
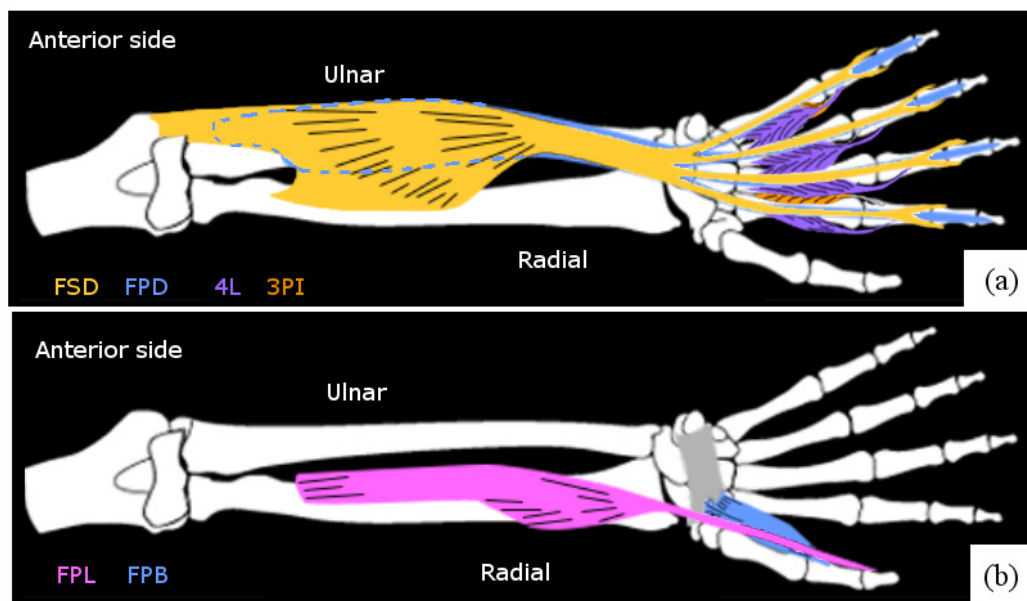

Figure 4.3: Muscles activated in posture P2 by (a) fingers flexion and (b) thumb flexion (adapted from LOUDON (1997)).
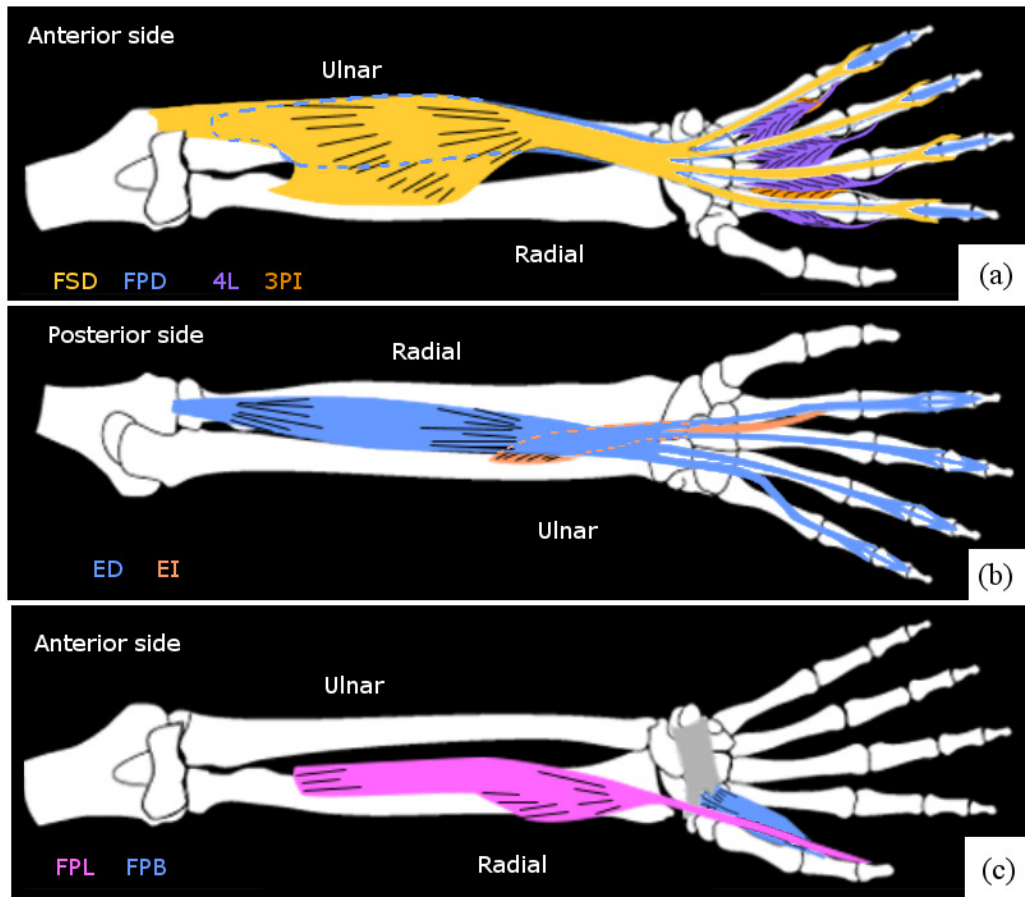
Figure 4.4: Muscles activated in postures P3 and P4 by (a) fingers flexion, (b) fingers extension and (c) thumb flexion (adapted from LOUDON (1997)).

postures activate all flexors, plus two extensors of the fingers: ED for middle finger extension and EI for index finger extension. As for the thumb, its full flexion activates both flexors: the extrinsic FPL and the intrinsic FPB. Without analyzing the actual effects of these muscles on the forearm cross-section, one can compare the results in Figure 4.1 for postures P3 and P4 as a transition from P1 to P2. Since P4 is closer to P1 than P2, the distribution map should be more similar to the one observed for the former, and indeed was the case. Similarly, the visible skin pixels in P3 are distributed as a halfway from posture P4 and P2, coherent to the expectations.

Regarding the four postures P1–P4, one notes that the resultant distribution map behaviors as a gradient, since the postures activate the same group of muscles, only in different degrees. The results in Figure 4.1 proved the OMG sensor ability to distinguish the different degrees of contraction.

As for the posture P5, it held all fingers fully flexed and the thumb extended, in a thumbs-up configuration. Hence, all flexors are activated for the fingers, as well as the thumb extensors (Figure 4.5). Since the only difference between the postures P2 and P5 is the thumb flexion/extension, the result for them should be similar. Indeed, by comparing them in Figure 4.1, one notes that both have fewer visible skin pixels around the middle, due to the contraction of
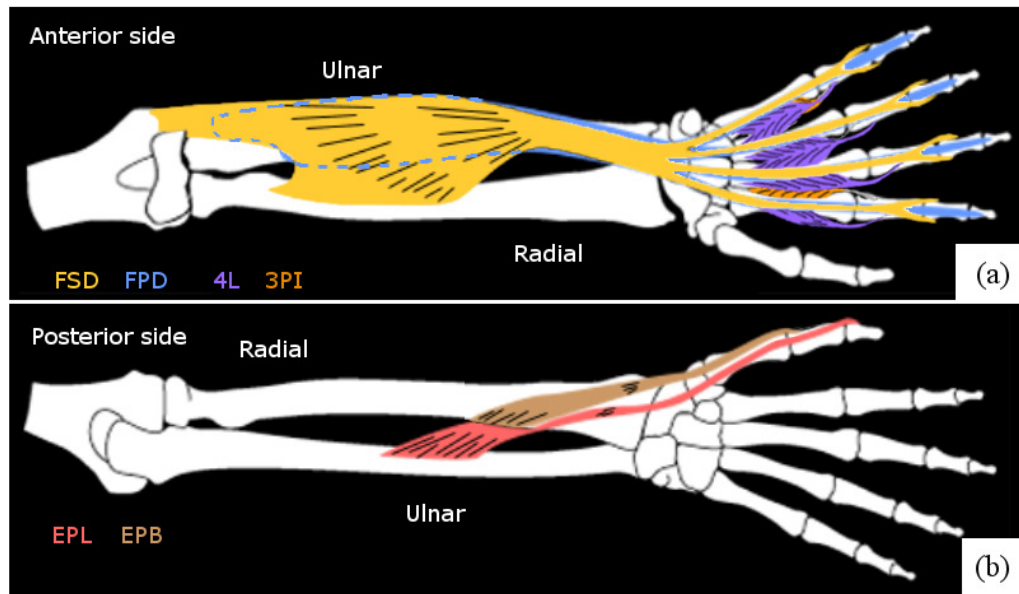
Figure 4.5: Muscles activated in posture P5 by (a) fingers flexion and (b) thumb extension (adapted from LOUDON (1997)).

flexor and extensor muscles, which increases the cross-section area near the elbow. However, the distribution maps for both postures differs significantly in subregion R9, which could be explained by the thumb configuration. Since one of the thumb flexors (FPB) is intrinsic, its contraction caused P2 to have more visible skin pixels in R9. On the other hand, since FPB is not contracted in P5 and no other intrinsic muscle is activated in this posture, R9 had fewer visible skin pixels, as expected.

Again, this OMG approach showed itself to be responsive to variation as subtle as the degree of contraction of overlaying muscles. The sensor, however, was too vulnerable to external conditions, so it was dismissed for the design C.

4.1.2 Sensor design C

The sensor design C is based on tracking the external markers to follow the muscular activity, so it can be validated by comparing the position of the markers for each of the eight tested postures in protocol II, as defined in Section 3.4.

The same tests were performed on the front and back of the forearm, so different groups of muscles can be directly monitored for all postures. All other muscles, which are hidden from the field of view of the camera, can be examined indirectly by observing the markers attached to the ulnar or the radial sides, near the forearm borderline in the image.

In this case, the average position of the centroid of all markers, seen from the local reference frame xy_0, are plotted together in complementary pairs AB, CD, EF, and GH. The choice to visualize the data in pairs instead of plotting all eight postures in a single plot is explained by the fact that, doing so, it is possible to evaluate the influence of selected muscles, as described in Section 3.4. The distribution map with the centroid of all markers are shown in Figure 4.6 for the anterior side, and in Figure 4.7 for the posterior side of the forearm. It can be observed at once that the markers moved from a posture to another, i.e. the points have both blue and orange representation nearby. By extension, the points without blue representation did not move. Through individual analysis of each pair of postures, it is possible to confirm the overall forearm deformation caused by the muscular activity the pair observes.

Postures A and B compare only the influence of thumb maximum flexion/extension so that the only difference between the muscles organization should be found on thumb flexors and extensors, shown in Figure 4.8. There are two flexors, of which only one is extrinsic; whereas there are three extrinsic extensor muscles on the back of the forearm. Thus, it is natural that the markers should present stronger displacement if attached to the posterior surface. However, if one should monitor only the anterior side, the fact that most of the activated muscles are hidden from the camera is not a severe problem. By observing the Figure 4.6, one notes that there is a subtle variation on markers position near the ulnar lateral. This is due to the fact that, when
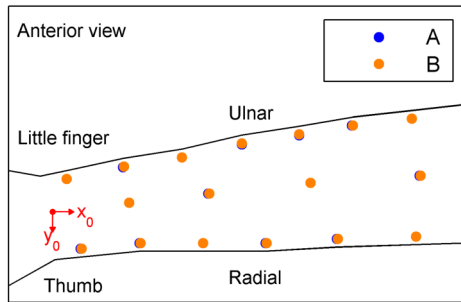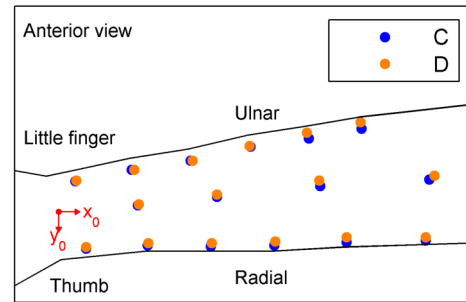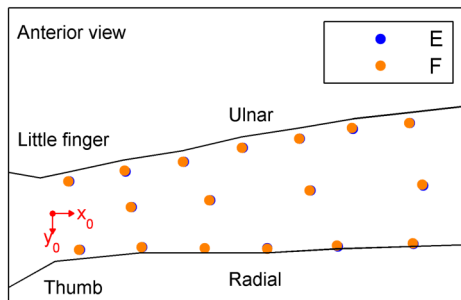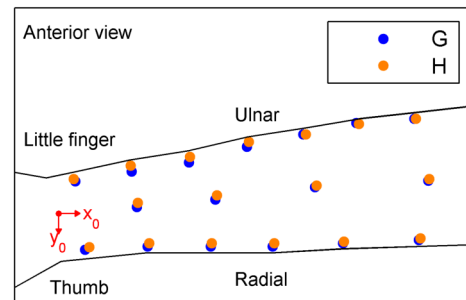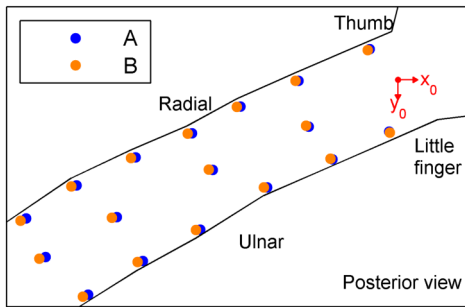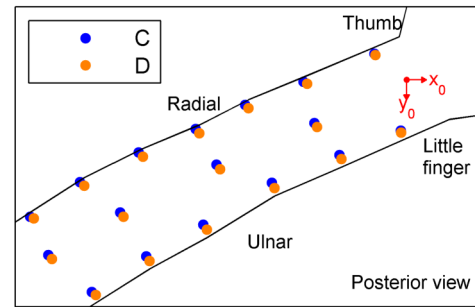

Figure 4.6: Markers average position for complementary postures from protocol II, anterior view (Wu *et al.*, Accepted for presentation).
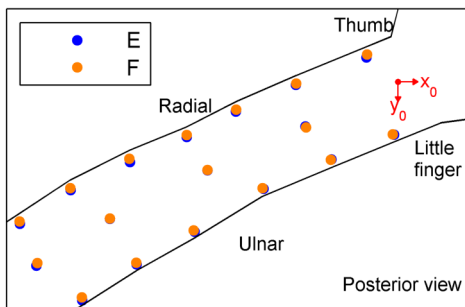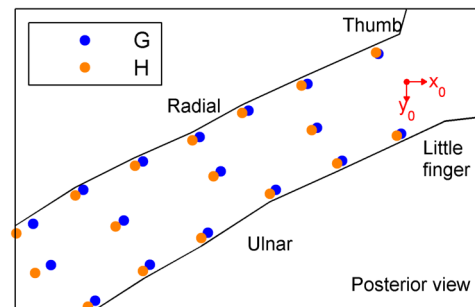
Figure 4.7: Markers average position for complementary postures from protocol II, posterior view.
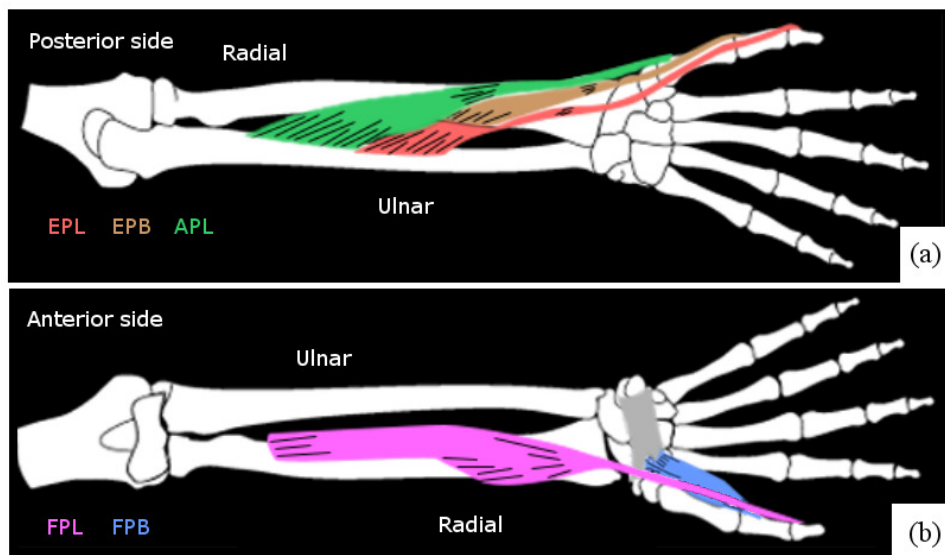
Figure 4.8: Muscles activated by postures (a) A and (b) B (adapted from LOUDON (1997)).

the thumb extensors contract, they expand the region near the ulnar lateral and closer to cubital fossa. This behavior can be observed in Figure 4.7 as well, in which the markers near elbow present higher offset.

Postures C and D investigate the muscle arrangement in two closed fist configurations:

Figure 4.9: Muscles activated by postures C and D for (a) fingers flexion and (b) thumb flexion (adapted from LOUDON (1997)).

one for the thumb over the fingers and another for the thumb under the fingers. In the beginning, it was expected of them to show no differences, since both bear all fingers and thumb in maximum flexion and should activate the very same group of muscles (Figur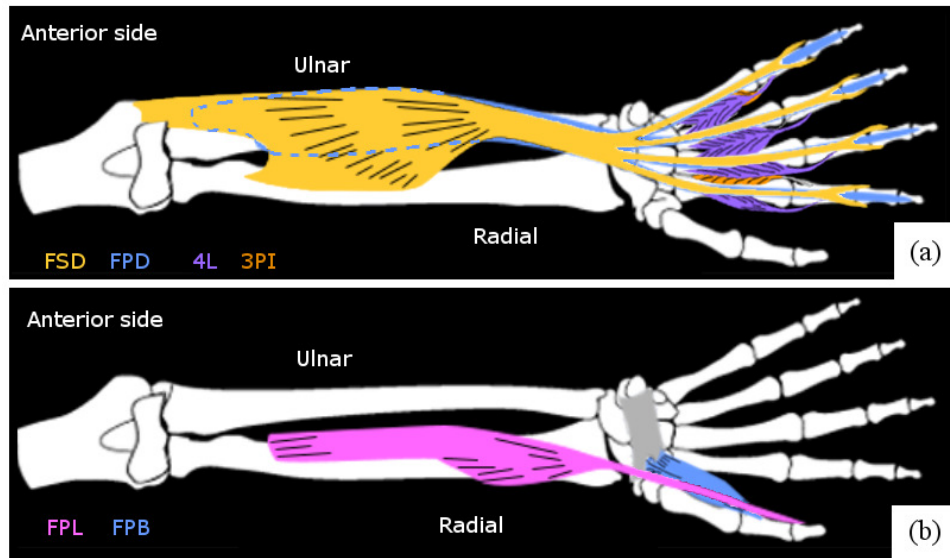e 4.9). However, as it can be seen from Figure 4.6 and Figure 4.7, the markers presented indeed offset from a posture to another, probably because one of the flexions is not the real maximum flexion due to the presence of the thumb (posture C) or the fingers (posture D) beneath. In comparison with postures AB, the duo CD presented higher offset since most of the activated muscles are extrinsic and can be found on both anterior and posterior sides. When attached to the posterior side, all markers had an offset due to extensor muscles at the back. On the other hand, when they are attached to the other side, the offset was slighter and concentrated mostly near cubital fossa, which is where the flexors expanded during contraction.

The postures E and F also activate muscles from both sides of the forearm, namely the flexor and extensor muscles of the digits (Figure 4.10). However, since this duo investigates the influence of individual extension of the index finger and middle fingers, it is expected that the flexor and extensor muscles should not be contracted at the same degree as in the previous case, with all fingers flexed at once. So, though some offset was observed for the markers attached either to the anterior and posterior surface, this offset was quite subtle, coherent to the prior expectation.

Lastly, the postures G and H investigate thumb adduction and abduction, which activate the muscles shown in Figure 4.11. Since posture G held the thumb adducted, it contracted only muscles found in the thenar eminence; whereas the posture H activated both intrinsic and extrinsic muscles. Therefore, it is not surprising that the markers moved most when attached

to the back of the forearm, where APL is found. Nonetheless, the markers on the anterior side also presented offset, especially those closer to the hand. This can be explained by the fact that the tested subject had an intact hand, and, since the human skin has natural elasticity, the contraction of so many muscles found in the thenar eminence pulled the skin towards the hand, moving thus the markers near the wrist.
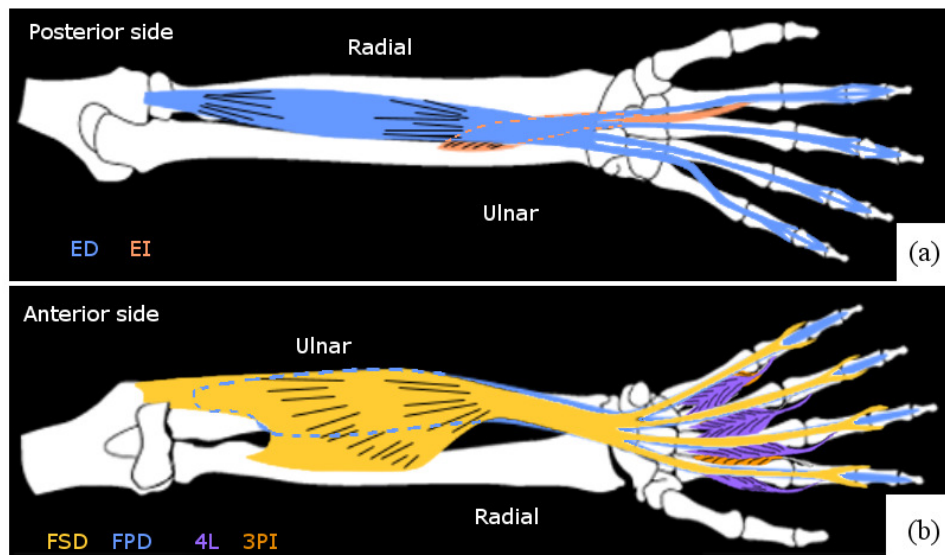

Figure 4.10: Muscles activated by postures E and F for fingers (a) extension and (b) flexion (adapted from LOUDON (1997)).
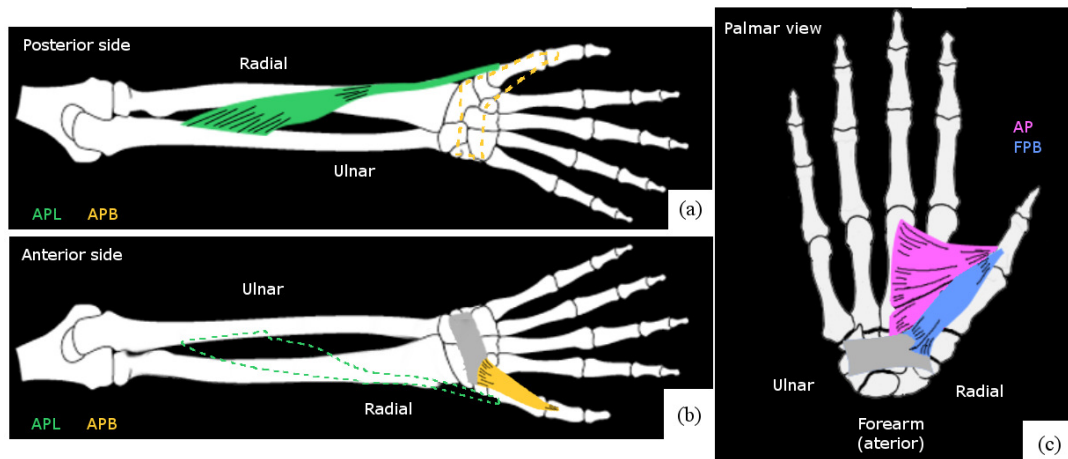

Figure 4.11: Muscles activated by postures (a,b) H and (c) G (adapted from LOUDON (1997)).

4.2 Artificial neural network classifier performance

The created ANN model, as described in Subsection 3.3.4, was calibrated with half of the total of experimental samples, whereas the other half was used to test the trained model. The classifier performance regarding the test set is discussed and evaluated in this section. The same analysis is done twice with the same data set seem either from the global XY or the local xy_0 reference frame. The other local reference frame xy_1 was not used because its origin marker was removed from too many frames during the processing routine due to light reflection in some experiments. The missing reference frame did not, however, compromise the overall performance of the classifier since the other one was preserved along the frames.

Here, the concept of accuracy and precision must be introduced, for they evaluate the performance of a predictive model. Accuracy relates to how correct the predictions were and is given by the ratio between all correct predictions (both true positives and true negatives) and the total of samples. Precision, on the other hand, relates to how relevant the predictions were and is given by the ratio between all true predictions (including both true and false positives) and the total of samples which really held positive value.

For example, Figure 4.12 shows the prediction results for a given set. They are plotted alongside the ground truth, i.e. the true class so that every point where orange and blue dots overlap is a correct prediction, also called a true positive. On the other hand, if they do not overlap, then there are two situations: from the orange dot perspective, the prediction was a false negative; meanwhile, from the blue dot perspective, it was a false positive. All those points with neither orange nor blue dots are true negatives. A more intuitive way of visualizing these elements is through the confusion matrix, in which every column corresponds to the true and every row to the predicted class. The intersection between a row and a column provides the total of samples predicted as belonging to that *row* class but are actually from that *column* class. For each one of the ten subjects, a confusion matrix was calculated as an average of 10-folds cross-validation; Figures 4.13 to 4.18 correspond to experiments with the markers attached to the forearm anterior side, whereas Figures 4.19 to 4.22 have markers on the posterior side.
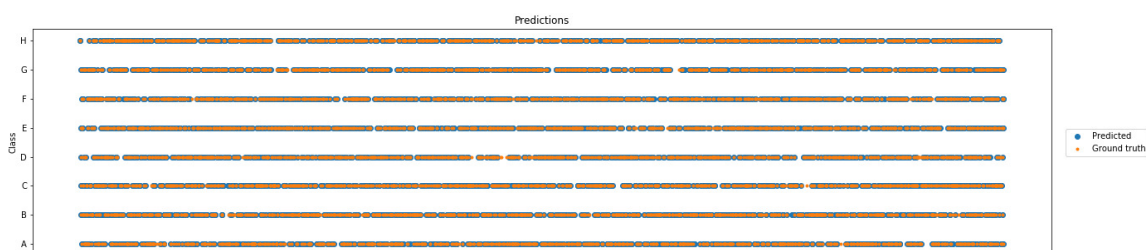
Figure 4.12: Predicted and true class for one subject.
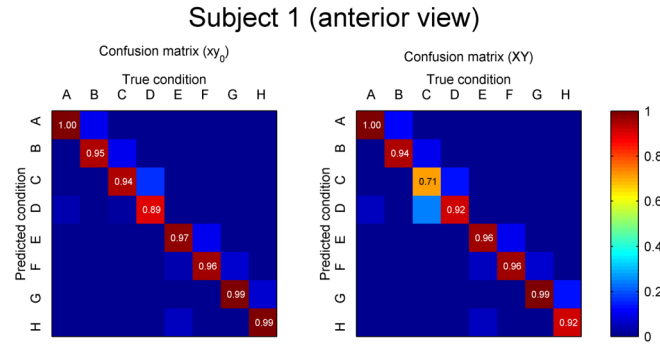
Figure 4.13: Subject 1. Confusion matrix (anterior side, seem from xy_0 and XY).
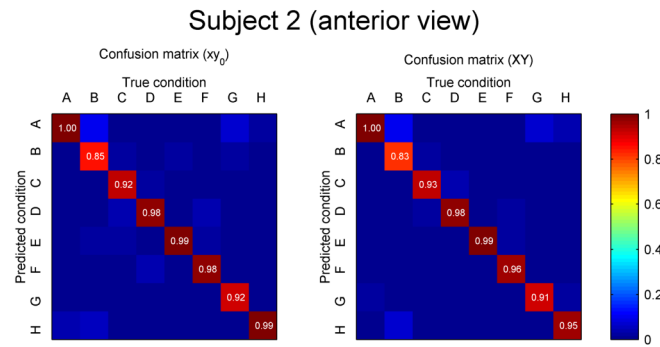
Figure 4.14: Subject 2. Confusion matrix (anterior side, seem from xy_0 and XY).
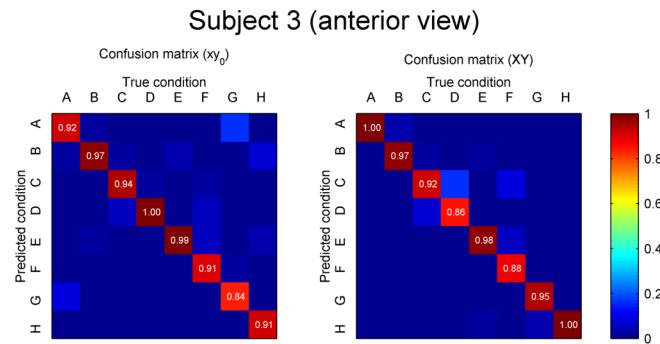
Figure 4.15: Subject 3. Confusion matrix (anterior side, seem from xy_0 and XY).
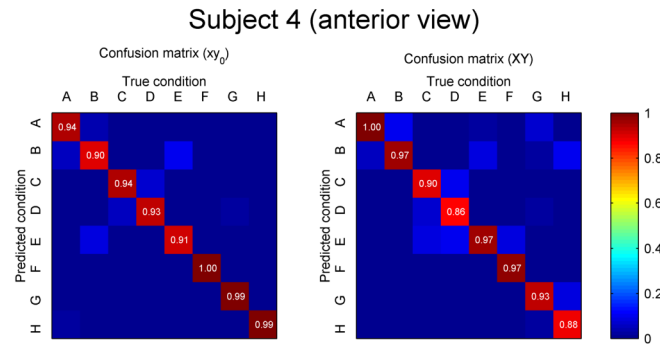
Figure 4.16: Subject 4. Confusion matrix (anterior side, seem from xy_0 and XY).
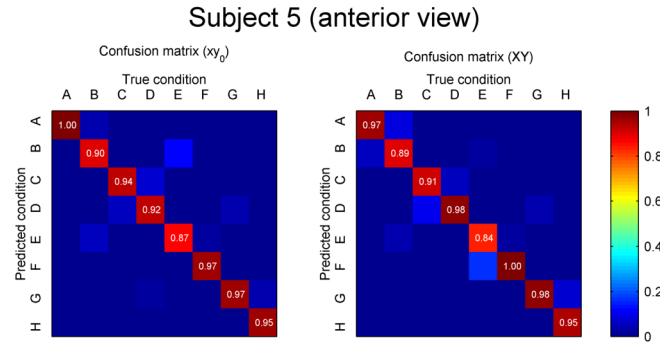
Figure 4.17: Subject 5. Confusion matrix (anterior side, seem from xy_0 and XY).
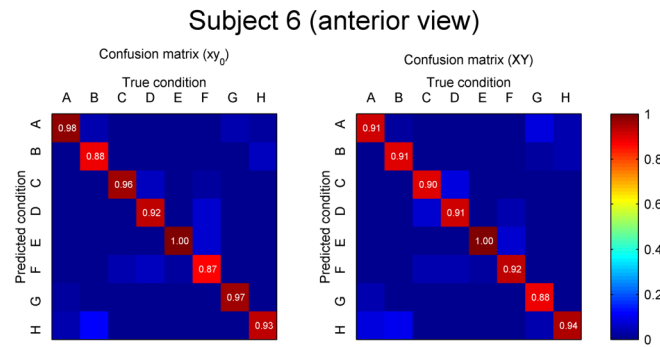
Figure 4.18: Subject 6. Confusion matrix (anterior side, seem from xy_0 and XY).
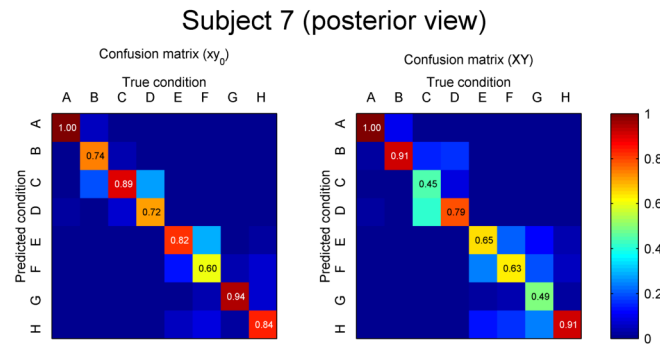
Figure 4.19: Subject 7. Confusion matrix (posterior side, seem from xy_0 and XY).
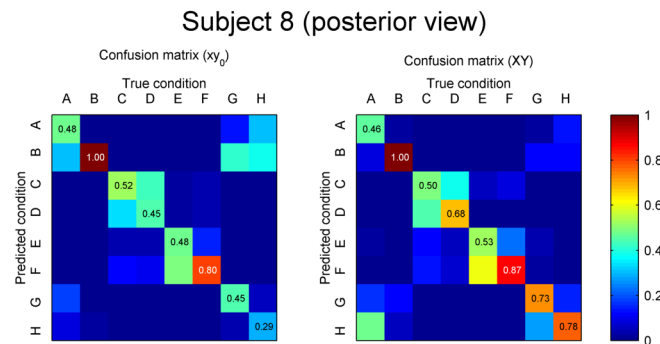
Figure 4.20: Subject 8. Confusion matrix (posterior side, seem from xy_0 and XY).
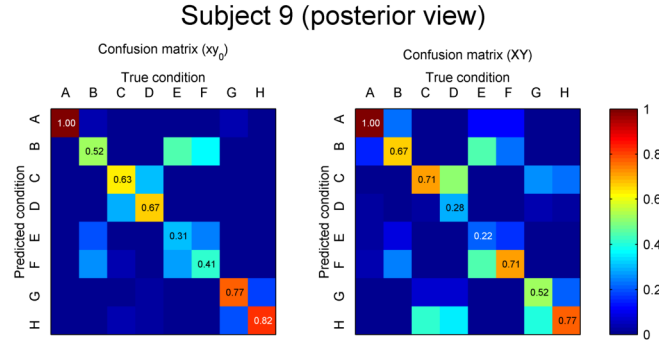
Figure 4.21: Subject 9. Confusion matrix (posterior side, seem from xy_0 and XY).
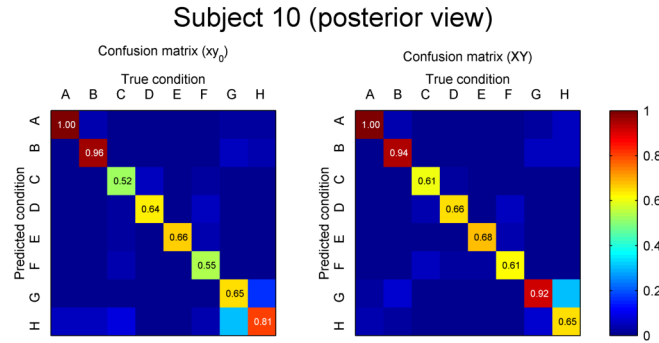
Figure 4.22: Subject 10. Confusion matrix (posterior side, seem from xy_0 and XY).

A quick analysis of the confusion matrices showed that the sensor presented a better hit ratio when the markers were attached to the anterior rather than the posterior side of the forearm. This can be confirmed by the high percentage of correct predictions found in the main diagonal of the confusion matrices in Figures 4.13 to 4.18, which corresponds to the experiments performed on the anterior side of the forearm. On the other hand, the hit ratio was lower when the markers were attached to the back of the forearm, as it can be observed from Figures 4.19 to 4.22. For instance, though the test subjects represented in Figures 4.19 and 4.22 had a uniform looking confusion matrix, with high values in the main diagonal and low values outside it, the percentage of the correct predictions is still inadequate for real applications. Regarding the test subjects represented in Figures 4.20 and 4.21, the percentage of correct predictions were low but there was also an increase of false predictions, as pointed by the presence of elements away from the main diagonal of the confusion matrix with any color other than dark blue. For instance, a significant number of samples belonging to class C was misclassified as belonging to class D and vice-versa in both test subjects.

From the confusion matrices, another important information can be extracted, that is the sensor had a better performance when the test samples were viewed from the local reference frame xy_0. As it can be verified, all confusion matrices on the left (viewed from xy_0) in the Figures 4.13 to 4.22 presented higher hit rate than the ones on the right (viewed from XY). This observation is coherent with the initial hypothesis when the sensor design C was first

proposed. The chief goal of transferring the position information of the samples from the global XY reference frame to the local one, xy_0, whose origin is set on a marker attached to the forearm itself, is to limit the variation of the centroid position within the anatomical boundaries of the forearm. For instance, if only the global-referenced position was considered, the detected variation would not derive purely from the muscular activity but also from the motion, natural spasms or slight vibrations of the forearm. On the other hand, if the position of the markers were compensated by a reference frame with origin set on the forearm, all those noises such as the motion, spasms and slight vibrations of the forearm would be filtered out, guaranteeing that the observed variation resulted from the muscular activity alone and increasing thus the classifier performance.

The sum of true and false predictions is provided by the confusion matrices, and the values of accuracy and precision of the classifier were calculated as an average of six and four subjects, corresponding respectively to the cases when the sensor monitored the front and the back of the forearm. The numerical performance was calculated for both reference frames as well, and the results are shown in Tables 4.1 to 4.4.

From the Table 4.1, one can see that the sensor was able to predict satisfactory eight postures on the anterior face of the forearm with an average precision of ∼92.4% and an average accuracy of ∼97.9% for the samples viewed from xy_0. In other words, less than 8% of the test samples were misclassified as belonging to a given class but belonged indeed to another when compared to all correct predictions. Also, the high accuracy indicates that the predictor misclassified in total less than 3% of the test samples, considering both false positives and false negatives in the whole data set.

The same experiments were evaluated for the samples viewed from XY and the resulting

Table 4.1: Classifier performance (anterior side, viewed from xy_0).

Posture	Accuracy(%)	Precision(%)
A	97.8	90.1
B	97.1	89.0
C	98.0	92.6
D	97.9	92.2
E	98.0	91.9
F	98.3	94.5
G	98.4	95.3
H	98.3	93.7
Average	97.9	92.4

Table 4.2: Classifier performance (anterior side, viewed from XY).

Posture	Accuracy(%)	Precision(%)
A	97.5	88.4
B	97.3	91.9
C	96.8	90.5
D	96.9	88.2
E	97.7	92.5
F	97.9	93.5
G	98.0	93.8
H	97.7	92.8
Average	97.5	91.4

accuracy and precision are shown in Table 4.2. It can be observed a slight decrease in precision and accuracy, reaching average values of $\sim$91.4% and $\sim$97.5%. The reduction was less than 1% and can almost be regarded as a statistical variation, but one must remember that the experimental protocol did not allow forearm translation so that, besides the muscular activity itself, only the micro spasms and vibrations of the forearm can influence the observed position of the markers. Since this variation is very subtle, the analysis from xy_0 did not outperform much the one from XY. The difference in the performance should, however, be expected to be more accentuated if the free motion of the forearm was allowed.

The sensor performance for the back of the forearm is evaluated in Tables 4.3 and 4.4 for samples viewed from xy_0 and XY, respectively. As expected from the confusion matrices analysis, there was a significant drop of precision, with an average value of $\sim$75.1% against the $\sim$92.4% obtained for the anterior face with local-referenced samples, and of $\sim$73.4% against $\sim$91.4% with global-referenced samples.

This drop can be explained by the experimental protocol and apparatus. The protocol II defined that the wrist must be kept in the neutral position, i.e. aligned to the forearm plane and axis, for all postures. It is very difficult, however, to reproduce some postures and to keep at the same time the wrist in the neutral position while resting the forearm on a flat surface with the back facing up. For instance, the wrist is slightly extended by postured C and D, and it is impossible to reproduce the posture H (thumb abduction) resting the forearm on the desk with the back facing up since the thumb cannot go through the desk. The alternative, then, is to reproduce the postures with the forearm suspended in the air but the user still has to keep the forearm parallel to the desk meanwhile. This position is tiresome, so, besides the natural spasms and vibrations of the forearm, the sensor was yet influenced by zoom. Though the zoom of the camera was kept constant, the forearm tended to move up and down when the test subject tried

Table 4.3: Classifier performance (posterior side, viewed from xy_0).

Posture	Accuracy(%)	Precision(%)
A	95.8	83.0
B	90.8	70.2
C	92.0	72.7
D	93.3	80.6
E	92.1	73.5
F	90.7	65.9
G	93.5	82.4
H	92.0	72.2
Average	92.5	75.1

Table 4.4: Classifier performance (posterior side, viewed from XY).

Posture	Accuracy(%)	Precision(%)
A	94.7	85.0
B	92.9	75.0
C	90.4	69.2
D	92.2	78.8
E	91.3	70.9
F	90.5	66.3
G	91.2	77.6
H	88.8	64.6
Average	91.5	73.4

to keep it still, but the task becomes too tiresome, simulating a zooming effect. As described in Subsection 3.3.3, the tracking algorithm is scale dependent, and the zooming effect can also modify the scale of the distribution map of the marker position, all that can compromise the sensor performance.

A decrease was also observed in the classifier accuracy for the back of the forearm, though it was not so accentuated as the one verified for the precision. This is probably because the accuracy considers all correct detections in the whole dataset (both true positives, and true negatives), whereas the precision considers only the relevant predictions.

The classifier precision per posture presented little variation when the sensor monitored the anterior surface of the forearm, as it can be noted from Tables 4.1 and 4.2. The small

standard variation can be simply the effect of statistical variation among the samples.

On the other hand, the precision per posture presented a high standard variation when the sensor evaluated the back of the forearm. As it can be observed from the Table 4.3, the precision per posture varied from 65.9% to 83.0%, with the best and worst performance for postures A and F, respectively. The same behavior can be seen from Table 4.4, with the highest precision of 85% for the posture A and the lowest of 64.6% for the posture H, followed by 66.3% for the posture F. As aforementioned, the defined postures were difficult to be reproduced with the camera pointing to the back of the forearm due to experimental protocol definitions. The fatigue, the need to keep the forearm still in the air for almost a whole minute, the forearm zooming in and out of the view, and the limited space available to reproduce the postures, all these factors disturbed the sample data. Thus, it was observed by investigating individual runs of ANN classifier that sometimes a class is missed for the entire dataset, i.e., the model fits all test samples in less than eight classes but there are samples for all eight postures. Also, this behavior was recurrent with posture F, explaining the low performance of the sensor to predict these postures as an average among all four test subjects. On the other hand, the better performance for posture A indicates that the sensor was able to classify most of the samples as one, and more than 80% of the times, the prediction was relevant. The same explanation can be applied to other postures with intermediate precisions.

The vulnerability of the sensor for the posterior side against its reliability for the anterior side of the forearm showed that the group of monitored muscles must be chosen according to the intended postures, as to provide comfort and space to reproduce them. For instance, it the thumb abduction is evaluated, then the better approach is to examine the muscles on the front of the forearm, allowing the user to rest the forearm on a flat surface. On the contrary, if one desires to monitor the overextension of the fingers or even the extension of the wrist, it is advisable to investigate the muscular activity from the back of the forearm, thus providing more space for hand motion.

Finally, the proposed sensor was compared to other myographic techniques in terms of the number of postures it supports and the classification performance.

The sensor is first compared with the currently published works on OMG, specifically the NISSLER *et al.* (2016) and NISSLER *et al.* (2017), which included recent advances on the topic.

In NISSLER *et al.* (2016), visual fiducial markers were employed as OM transducers, the forearm of test subjects was fixed on a rigid support with Velcro straps, and the sensor was demonstrated for four postures classification with a non-linear learning model and an average classification error of $\sim$0.14%. In terms of the OM transducers, the evaluation of the position and orientation of the fiducial markers was claimed to be complicated due to their small size. Though the sensor here demonstrated employs also small sized markers, they are orientation free, and only the color information is required so that they do not need to be accurately

segmented from the rest of the image.

A different approach was presented in Nissler *et al.* (2017), which replaced the fiducial markers with a wide rectangular plain sticker as OM transducer, and correlated the position changes and bounding box deformation observed on the rectangle with five postures using a CNN model, reaching an average accuracy of $\sim$97.8%.

Regarding the experimental apparatus, both Nissler *et al.* (2016) and Nissler *et al.* (2017) confined the forearm of test subjects on a rigid support with Velcro straps. In comparison, the proposed sensor tries to provide its users most of the motion freedom, so that the experience can be comfortable after a prolonged time of use and feasible for real applications.

The classification performance is difficult to be compared among all three works since they considered different experimental protocol, posture set, and measures to evaluate the classifier model performance. However, the proposed sensor supports up to eight different static postures with $\sim$92.4% precision and $\sim$97.9% accuracy, being comparable to the accuracy shown in Nissler *et al.* (2017), which supports only five postures.

Comparing yet with the FMG sensor presented in Fujiwara *et al.* (2017), it is unknown if presently the proposed sensor can detect 11 static postures correctly. Therefore, the numerical comparison of the classification performance is also not reliable, since both sensors have different bases, postures, and experimental conditions. However, the OMG sensor here reported showed a performance comparable to the $\sim$89.9% accuracy of the other sensor.

5 CONCLUSION

In this project, an OMG sensor for hand posture and gesture recognition is proposed. The developed system is here described in detail, from the early sketches till the final testing and evaluation of the to classify eight different postures. Though not implemented as part of the algorithm, the proposed sensor can recognize gestures via temporal analysis of a sequence of postures by extension.

Since the beginning of the project, the challenge has been to detect visually the forearm cross-section contraction/expansion due to muscular activity. Therefore, several OM transducer candidates are proposed and tested, from which three most relevant ones are resumed in this report. From the first prototype, based on spectral analysis, until the final one, based on object detection, there has been an evolution, considering the problems observed during the experiments, the sensor performance, and, above all, the comfort, and the ability to provide a motion as free as possible. Indeed, in comparison with the stripe analysis or the skin pixel reckoning, the final version founded on marker detection and tracking is much more robust to instrumental noises and environment conditions without loss of comfort. Moreover, since the hand is not directly investigated in this configuration, the sensor should also be responsive to impaired users, provided that they suffered neither muscle injury or loss.

Currently, the sensor can recognize correctly up to eight different postures and is responsive to either situation when the monitored surface is the underside or the back of the forearm. In the first case, the classifier presented a precision of 92.4% and an accuracy of 97.9%, whereas they dropped in the second case, with 75.1% precision and 92.5% accuracy. Nonetheless, the sensor showed a satisfactory performance, since this drop was caused not by deficiencies of the sensor, but because it was easier and more comfortable to reproduce the selected postures with the camera pointed to the anterior rather to the posterior side of the forearm.

Considering the sensor performance, its simplicity, and comfort of use, it is expected to contribute to a wide range of applications, such as HCI, healthcare, rehabilitation, entertainment and so forth.

As for future work perspective, the following can be listed:

- Real-time functions: to achieve real-time performance for real applications, the sensor can be improved still. For example, parallel programming and algorithm refinement, especially on the stage of image preprocessing, can improve the processing time. Besides, the sensor software presently is sequential and programmed to operate offline, so some adjustments are required to provide real-time functions. Moreover, the algorithm is based

on several morphological operations, which are very expensive for requiring a spatial sampling. Another alternative to improve the processing time would be to downsample the input frames, as for example, instead of acquiring image data at 24 fps, a lower acquisition rate could be adopted since there is a physical limitation regarding how much a person can move the forearm in consecutive frames. Therefore, a study about what sample rate provides a suitable price-performance ratio can be conducted.

- Addition of more detectable postures: depending on the final application in view, eight postures might be enough or not. For instance, two postures are enough for an on/off control while eight are too few for sign language recognition.

- Study of dynamic gestures: the protocols defined in this work disregarded the wrist movement, but it is tiresome to keep the wrist in the same position for a prolonged time. Besides, since any other movements, such as raising the whole arm or just the forearm, will change the muscle arrangement, it should be investigated whether the sensor can distinguish the same posture for different arm configurations.

- Effectively test the system on users with the limited use of hand or a missing hand.

6 PUBLICATIONS

WU, Y.T.; FUJIWARA, E. and SUZUKI, C.K. Optical myography sensor for gesture recognition. In **Advanced motion control (AMC), 2018 IEEE International Workshop on**. IEEE, Accepted for presentation.

WU, Y.T.; FUJIWARA, E. and SUZUKI, C.K. Optical myography system for posture monitoring. In **Consumer Electronics (ISCE), 2016 IEEE International Symposium on**, pp. 37–38. IEEE, 2016.

FUJIWARA, E.; WU, Y.T.; SUZUKI, C.K.; DE ANDRADE, D.T.G.; RIBAS NETO, A. and ROHMER, E. Optical fiber force myography sensor for applications in prosthetic hand control. In **Advanced motion control (AMC), 2018 IEEE International Workshop on**. IEEE, Accepted for presentation.

FUJIWARA, E.; WU, Y.T.; DOS SANTOS, M.F.M.; SCHENKEL, E.A. and SUZUKI, C.K. Development of a tactile sensor based on optical fiber specklegram analysis and sensor data fusion technique. **Sensors and Actuators A: Physical**, v. 263, 677–686, 2017.

FUJIWARA, E.; PAULA, F.D.; WU, Y.T.; SANTOS, M.F.; SCHENKEL, E.A. and SUZUKI, C.K. Optical fiber tactile sensor based on fiber specklegram analysis. In **Optical Fiber Sensors Conference (OFS), 2017 25th**, pp. 1–4. IEEE, 2017.

FUJIWARA, E.; WU, Y.T.; SANTOS, M.F.M.; SCHENKEL, E.A. and SUZUKI, C.K. Optical fiber specklegram sensor for measurement of force myography signals. **IEEE Sensors Journal**, v. 17, n. 4, 951–958, Feb 2017.

FUJIWARA, E.; PAULA, F.D.; WU, Y.T.; SANTOS, M.F. and SUZUKI, C.K. Optical fiber tactile sensor for user interfaces. In **Consumer Electronics (ISCE), 2016 IEEE International Symposium on**, pp. 11–12. IEEE, 2016.

REFERENCES

AGGARWAL, J.K. and CAI, Q. Human motion analysis: A review. **Computer vision and image understanding**, v. 73, n. 3, 428–440, 1999.

BISHOP, C. Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn. **Springer, New York**, 2007.

BOVIK, A.C. **The essential guide to image processing**. Academic Press, 2009.

BOWEN, W.P. **Applied anatomy and kinesiology: the mechanism of muscular movement**. Lea & Febiger, 1919.

BRADSKI, G. and KAEHLER, A. **Learning OpenCV: Computer vision with the OpenCV library**. " O'Reilly Media, Inc.", 2008.

BRUNELLI, D.; TADESSE, A.M.; VODERMAYER, B.; NOWAK, M. and CASTELLINI, C. Low-cost wearable multichannel surface emg acquisition for prosthetic hand control. In **Advances in Sensors and Interfaces (IWASI), 2015 6th IEEE International Workshop on**, pp. 94–99. IEEE, 2015.

CASTELLINI, C.; ARTEMIADIS, P.; WININGER, M.; AJOUDANI, A.; ALIMUSAJ, M.; BICCHI, A.; CAPUTO, B.; CRAELIUS, W.; DOSEN, S.; ENGLEHART, K. *et al.* Proceedings of the first workshop on peripheral machine interfaces: Going beyond traditional surface electromyography. **Frontiers in neurorobotics**, v. 8, 2014.

CHO, E.; CHEN, R.; MERHI, L.K.; XIAO, Z.; POUSETT, B. and MENON, C. Force myography to control robotic upper extremity prostheses: a feasibility study. **Frontiers in bioengineering and biotechnology**, v. 4, 2016.

CHOLLET, F. *et al.* Keras. https://github.com/fchollet/keras, 2015.

CHOWDHURY, R.H.; REAZ, M.B.; ALI, M.A.B.M.; BAKAR, A.A.; CHELLAPPAN, K. and CHANG, T.G. Surface electromyography signal processing and classification techniques. **Sensors**, v. 13, n. 9, 12431–12466, 2013.

COOTES, T.F.; EDWARDS, G.J. and TAYLOR, C.J. Active appearance models. **IEEE Transactions on pattern analysis and machine intelligence**, v. 23, n. 6, 681–685, 2001.

COOTES, T.F.; EDWARDS, G.J.; TAYLOR, C.J. *et al.* Comparing active shape models with active appearance models. In **Bmvc**, v. 99, pp. 173–182. Citeseer, 1999.

DALLEY, S.A.; VAROL, H.A. and GOLDFARB, M. A method for the control of multigrasp myoelectric prosthetic hands. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 20, n. 1, 58–67, 2012.

DIPIETRO, L.; SABATINI, A.M. and DARIO, P. A survey of glove-based systems and their applications. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 38, n. 4, 461–482, 2008.

DRAKE, R.; VOGL, W. and MITCHELL, A. Gray's anatomy for students. 3rd. **Edinburgh, Scotland: Churchill Livingstone**, 2015.

FANG, Y.; LIU, H.; LI, G. and ZHU, X. A multichannel surface emg system for hand motion recognition. **International Journal of Humanoid Robotics**, v. 12, n. 02, 1550011, 2015.

FARINA, D.; JIANG, N.; REHBAUM, H.; HOLOBAR, A.; GRAIMANN, B.; DIETL, H. and ASZMANN, O.C. The extraction of neural information from the surface emg for the control of upper-limb prostheses: emerging avenues and challenges. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 22, n. 4, 797–809, 2014.

FERIGO, D.; MERHI, L.K.; POUSETT, B.; XIAO, Z.G. and MENON, C. A case study of a force-myography controlled bionic hand mitigating limb position effect. **Journal of Bionic Engineering**, v. 14, n. 4, 692–705, 2017.

FUJIWARA, E.; D. SANTOS, M.F.M. and SUZUKI, C.K. Flexible optical fiber bending transducer for application in glove-based sensors. **IEEE Sensors Journal**, v. 14, n. 10, 3631–3636, Oct 2014.

FUJIWARA, E.; WU, Y.T.; MIYATAKE, D.Y.; SANTOS, M.F. and SUZUKI, C.K. Evaluation of thumb-operated directional pad functionalities on a glove-based optical fiber sensor. **IEEE Transactions on Instrumentation and Measurement**, v. 62, n. 8, 2330–2337, 2013.

FUJIWARA, E.; WU, Y.T.; SANTOS, M.F.M.; SCHENKEL, E.A. and SUZUKI, C.K. Optical fiber specklegram sensor for measurement of force myography signals. **IEEE Sensors Journal**, v. 17, n. 4, 951–958, Feb 2017.

FUJIWARA, E.; WU, Y.T.; SUZUKI, C.K.; DE ANDRADE, D.T.G.; RIBAS NETO, A. and ROHMER, E. Optical fiber force myography sensor for applications in prosthetic hand control. In **Advanced motion control (AMC), 2018 IEEE International Workshop on**. IEEE, Accepted for presentation.

FUKUDA, O.; TSUJI, T.; KANEKO, M. and OTSUKA, A. A human-assisting manipulator teleoperated by emg signals and arm motions. **IEEE Transactions on Robotics and Automation**, v. 19, n. 2, 210–222, 2003.

GAVRILA, D.M. The visual analysis of human movement: A survey. **Computer vision and image understanding**, v. 73, n. 1, 82–98, 1999.

GIJSBERTS, A.; BOHRA, R.; SIERRA GONZÁLEZ, D.; WERNER, A.; NOWAK, M.; CAPUTO, B.; ROA, M.A. and CASTELLINI, C. Stable myoelectric control of a hand prosthesis using non-linear incremental learning. **Frontiers in neurorobotics**, v. 8, 8, 2014.

GONZALEZ, R. and WOODS, R. **Digital image processing**. Pearson Prentice Hall, 2007.

GONZALEZ, R.; WOODS, R. and EDDINS, S. **Digital image processing using MATLAB**. Pearson Prentice Hall, 2003.

GRIMES, G.J. Digital data entry glove interface device. November 8 1983. US Patent 4,414,537.

HUG, F. Can muscle coordination be precisely studied by surface electromyography? **Journal of electromyography and kinesiology**, v. 21, n. 1, 1–12, 2011.

ITKARKAR, R.R. and NANDI, A.V. A survey of 2d and 3d imaging used in hand gesture recognition for human-computer interaction (hci). In **Electrical and Computer Engineering (WIECON-ECE), 2016 IEEE International WIE Conference on**, pp. 188–193. IEEE, 2016.

JIANG, X.; MERHI, L.K.; XIAO, Z.G. and MENON, C. Exploration of force myography and surface electromyography in hand gesture classification. **Medical Engineering and Physics**, v. 41, 63–73, 2017.

KAKUMANU, P.; MAKROGIANNIS, S. and BOURBAKIS, N. A survey of skin-color modeling and detection methods. **Pattern recognition**, v. 40, n. 3, 1106–1122, 2007.

KESSLER, G.D.; HODGES, L.F. and WALKER, N. Evaluation of the cyberglove as a whole-hand input device. **ACM Transactions on Computer-Human Interaction (TOCHI)**, v. 2, n. 4, 263–283, 1995.

KIM, D.; HILLIGES, O.; IZADI, S.; BUTLER, A.D.; CHEN, J.; OIKONOMIDIS, I. and OLIVIER, P. Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In **Proceedings of the 25th annual ACM symposium on User interface software and technology**, pp. 167–176. ACM, 2012.

KRIZHEVSKY, A.; SUTSKEVER, I. and HINTON, G.E. Imagenet classification with deep convolutional neural networks. In **Advances in neural information processing systems**, pp. 1097–1105. 2012.

LI, N.; YANG, D.; JIANG, L.; LIU, H. and CAI, H. Combined use of fsr sensor array and svm classifier for finger motion recognition based on pressure distribution map. **Journal of Bionic Engineering**, v. 9, n. 1, 39–47, 2012.

LOUDON, J.K. Hand kinesiology. 1997.
URL: *http://classes.kumc.edu/sah/resources/handkines/kines2.html*

MELCHIORRI, C. and KANEKO, M. **Robot Hands**, pp. 463–480. Springer International Publishing, Cham, 2016.

MOESLUND, T.B. **Interacting with a virtual world through motion capture**. Springer, 2000.

MOESLUND, T.B. and GRANUM, E. A survey of computer vision-based human motion capture. **Computer vision and image understanding**, v. 81, n. 3, 231–268, 2001.

MURTHY, G. and JADON, R. A review of vision based hand gestures recognition. **International Journal of Information Technology and Knowledge Management**, v. 2, n. 2, 405–410, 2009.

NAIK, G.R.; ACHARYYA, A. and NGUYEN, H.T. Classification of finger extension and flexion of emg and cyberglove data with modified ica weight matrix. In **Engineering in Medicine and Biology Society (EMBC), 2014 36th annual international conference of the IEEE**, pp. 3829–3832. IEEE, 2014.

NISSLER, C.; BADSHAH, I.; CASTELLINI, C.; KEHL, W. and NAVAB, N. Improving optical myography via convolutional neural networks. In **Proceedings on 2017 Myoelectric Controls and Upper Limb Prosthetics Symposium, MEC'17**, p. 84. 2017.

NISSLER, C.; MOURIKI, N. and CASTELLINI, C. Optical myography: detecting finger movements by looking at the forearm. **Frontier Neurobobotics**, v. 10, 2016.

NISSLER, C.; MOURIKI, N.; CASTELLINI, C.; BELAGIANNIS, V. and NAVAB, N. Omg: introducing optical myography as a new human machine interface for hand amputees. In **Rehabilitation Robotics (ICORR), 2015 IEEE International Conference on**, pp. 937–942. IEEE, 2015.

PARK, Y.; LEE, J. and BAE, J. Development of a wearable sensing glove for measuring the motion of fingers using linear potentiometers and flexible wires. **IEEE Transactions on Industrial Informatics**, v. 11, n. 1, 198–206, 2015.

PAVLOVIC, V.I.; SHARMA, R. and HUANG, T.S. Visual interpretation of hand gestures for human-computer interaction: A review. **IEEE Transactions on pattern analysis and machine intelligence**, v. 19, n. 7, 677–695, 1997.

PHILLIPS, S.L. and CRAELIUS, W. Residual kinetic imaging: a versatile interface for prosthetic control. **Robotica**, v. 23, n. 3, 277–282, 2005.

PIZZOLATO, S.; TAGLIAPIETRA, L.; COGNOLATO, M.; REGGIANI, M.; MÜLLER, H. and ATZORI, M. Comparison of six electromyography acquisition setups on hand movement classification tasks. **PloS one**, v. 12, n. 10, e0186132, 2017.

RADMAND, A.; SCHEME, E. and ENGLEHART, K. High-density force myography: A possible alternative for upper-limb prosthetic control. **Journal of Rehabilitation Research & Development**, v. 53, n. 4, 2016.

RAUTARAY, S.S. and AGRAWAL, A. Vision based hand gesture recognition for human computer interaction: a survey. **Artif. Intell. Rev.**, v. 43, n. 1, 1–54, 2015.

RAVINDRA, V. and CASTELLINI, C. A comparative analysis of three non-invasive human-machine interfaces for the disabled. **Front. Neurorobotics**, v. 8, 2014.

REN, Z.; YUAN, J.; MENG, J. and ZHANG, Z. Robust part-based hand gesture recognition using kinect sensor. **IEEE transactions on multimedia**, v. 15, n. 5, 1110–1120, 2013.

RODRIGUEZ, J.D.; PEREZ, A. and LOZANO, J.A. Sensitivity analysis of k-fold cross validation in prediction error estimation. **IEEE transactions on pattern analysis and machine intelligence**, v. 32, n. 3, 569–575, 2010.

SOILLE, P. **Morphological image analysis: principles and applications**. Springer Science & Business Media, 2013.

SZELISKI, R. **Computer vision: algorithms and applications**. Springer Science & Business

Media, 2010.

TAVAKOLI, M.; BENUSSI, C. and LOURENCO, J.L. Single channel surface emg control of advanced prosthetic hands: A simple, low cost and efficient approach. **Expert Systems with Applications**, v. 79, 322–332, 2017.

THAKUR, S.; PAUL, S.; MONDAL, A.; DAS, S. and ABRAHAM, A. Face detection using skin tone segmentation. In **Information and Communication Technologies (WICT), 2011 World Congress on**, pp. 53–60. IEEE, 2011.

TOMITA, F.; SHIRAI, Y. and TSUJI, S. Description of textures by a structural analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, , n. 2, 183–191, 1982.

TRUCCO, E. and VERRI, A. **Introductory techniques for 3-D computer vision**, v. 201. Prentice Hall Englewood Cliffs, 1998.

TUCERYAN, M.; JAIN, A.K. *et al.* Texture analysis. **Handbook of pattern recognition and computer vision**, v. 2, 235–276, 1993.

VEZHNEVETS, V.; SAZONOV, V. and ANDREEVA, A. A survey on pixel-based skin color detection techniques. In **Proc. Graphicon**, v. 3, pp. 85–92. Moscow, Russia, 2003.

WANG, C.; GAO, W. and SHAN, S. An approach based on phonemes to large vocabulary chinese sign language recognition. In **Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on**, pp. 411–416. IEEE, 2002.

WANG, L.; HU, W. and TAN, T. Recent developments in human motion analysis. **Pattern recognition**, v. 36, n. 3, 585–601, 2003.

WANG, L.; MEYDAN, T. and WILLIAMS, P. Design and evaluation of a 3-d printed optical sensor for monitoring finger flexion. **IEEE Sensors Journal**, v. 17, n. 6, 1937–1944, 2017.

WHEATLAND, N.; WANG, Y.; SONG, H.; NEFF, M.; ZORDAN, V. and JÖRG, S. State of the art in hand and finger modeling and animation. In **Computer Graphics Forum**, v. 34, pp. 735–760. Wiley Online Library, 2015.

WININGER, M.; KIM, N.H. and CRAELIUS, W. Pressure signature of forearm as predictor of grip force. **Journal of rehabilitation research and development**, v. 45, n. 6, 883, 2008.

WU, Y.T.; FUJIWARA, E. and SUZUKI, C.K. Optical myography system for posture monitor-

ing. In **Consumer Electronics (ISCE), 2016 IEEE International Symposium on**, pp. 37–38. IEEE, 2016.

WU, Y.T.; FUJIWARA, E. and SUZUKI, C.K. Optical myography sensor for gesture recognition. In **Advanced motion control (AMC), 2018 IEEE International Workshop on**. IEEE, Accepted for presentation.

XIAO, Z.G.; ELNADY, A.M. and MENON, C. Control an exoskeleton for forearm rotation using fmg. In **Biomedical Robotics and Biomechatronics (2014 5th IEEE RAS & EMBS International Conference on**, pp. 591–596. IEEE, 2014.

YEO, H.S.; LEE, B.G. and LIM, H. Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. **Multimedia Tools and Applications**, v. 74, n. 8, 2687–2715, 2015.

ZHOU, H. and HU, H. Human motion tracking for rehabilitation—a survey. **Biomedical Signal Processing and Control**, v. 3, n. 1, 1–18, 2008.