DLR-IB-RM-OP-2016-358

Optical Myography Using Convolutional Neural Networks for Estimating Finger Poses

Masterarbeit

Imran Mashood Badshah

Deutsches Zentrum
DLR für Luft- und Raumfahrt

Institut für Robotik und Mechatronik	BJ.: 2016 IB.Nr.: DLR-IB-RM-OP-2016-358

MASTERARBEIT

OPTICAL MYOGRAPHY USING CONVOLUTIONAL NEURAL NETWORKS FOR ESTIMATING FINGER POSES

Freigabe:	Der Bearbeiter:	Unterschriften
	Imran Mashood Badshah	
	Betreuer:	
	Christian Nißler	
	Der Institutsdirektor	
	Dr. Alin Albu-Schäffer	

Dieser Bericht enthält 97 Seiten, 42 Abbildungen und 0 Tabellen

Ort: Oberpfaffenhofen	Datum: 22.12.2016	Bearbeiter: Imran Mashood Badshah	Zeichen:

TECHNISCHE UNIVERSITÄT MÜNCHEN

DEPARTMENT OF INFORMATICS

Master's Thesis in Informatics: Biomedical Computing

Optical Myography Using Convolutional Neural Networks For Estimating Finger Poses

Optische Myographie mittels Convolutional Neural Networks zur Fingerposenschätzung

Author:	Imran Mashood Badshah
Supervisor:	Christian Nissler (DLR), Wadim Kehl (TUM)
Advisor:	Dr. Claudio Castellini (DLR), Prof. Dr. Nassir Navab (TUM)
Submission Date:	15th December 2016

I confirm that this master's thesis in informatics: biomedical computing is my own work and I have documented all sources and material used.

Munich, 15th December 2016 Imran Mashood Badshah

Acknowledgements

When many hands work together, that is when an innovation can be propelled towards recognition. This thesis not just exemplifies the statement, but also strives towards one day bestowing those who dream of a hand to be able to lend in their own helping hand. From the family that supported me more than words can describe, to the scientists and people of other streams who inspired me to choosing the path I have chosen, and the one for supporting me in this path, I owe my gratitude. I would like to thank my past teachers and my present faculty members at the Technische Universtität München (TUM) for giving me the opportunity to learn as much as I have during my Master of Science in Biomedical Computing. I would also like to thank all my colleagues at the Deutsches Zentrum für Luft- und Raumfahrt (DLR) and especially the Bionics group within it for giving me a stage and chance to take a step forward in not just accumulating data for a Master's thesis, but for bolstering my knowledge and giving me support in every way possible to make it a true learning experience in the field of bionics. This Master's thesis could neither have a start nor an end without Dr. Claudio Castellini and Mr. Christian Nißler of DLR, and Prof. Dr. Nassir Navab and Mr. Wadim Kehl from TUM. Thank you for not just supervising me, but also for all the cooperative help in the form of knowledge and support. I am grateful to Nikoleta Mouriki for laying the foundation for this thesis. To friends and people I have come across and could not fit in this page, know that the support has always been acknowledged.

Abbreviations and Acronyms

AHP	Active Hand Prosthesis
CNN	Convolutional Neural Network
ELU	Exponential Linear Unit
FMG	Force myography
FPS	frames per second
FSR	force-sensing resistor
GB	Gigabyte
GPU	graphics processing unit
GUI	graphical user interface
I	Intensity
IP	interphalangeal
LED	Light Emitting Diode
LO	log-opponent
MCP	metacarpophalangeal
N/A	Not available
NCC	normalized cross-correlation
NIR	Near-infrared
OMG	Optical Myography
OMG_CNN	Optical Myography using Convolutional Neural Networks
RAM	Random-Access Memory
RGB	Red Green Blue
RMS	Root Mean Squared
ROI	region of interest
R_g	Red-green opponent
SMG	Sonomyography
sEMG	surface Electromyography
UDP	User Datagram Protocol
US	Ultrasound
WMA	World Medical Association

Abstract

A plethora of techniques exist to rehabilitate an amputee's lost limb. Active Hand Prosthesis (AHP) is one such tool which promises aid to an amputee to gain control over daily activities. Translating the user's intent into appropriate movement of the prosthetic device, especially that of the hand is still a target to be attained by the various methods which try to acquire and interpret these biological signals. One such novel modality, known as Optical Myography (OMG) laid the proof of concept of mapping deformations on the surface of the forearm caused by muscle movement to estimate finger poses for an artificial hand. The surface movements were tracked using AprilTags stuck the surface of the forearm and by strapping the forearm to a frame in order to suppress external movement. Misdetection and missed detection of these tags can cause noise in the data acquired for the machine learning algorithm. This thesis aims to develop OMG for the estimation of finger poses by using computer vision to observe the muscle movements on the surface of the forearm thus obviating the need to rely on precisely detected tags. In order to do so, the machine learning algorithm used is a Convolutional Neural Network (CNN) trained on images of the forearm captured during the execution of the desired finger poses. Various feature extraction sources are studied before choosing the most practically applicable source to test on intact subjects.

Contents

Outline of the Thesis

Chapter 1: Introduction

Section 1.1: Motivation

This section explains the reason for the establishment of the modality and the research on its development as a topic for the thesis.

Section 1.2: State-of-the-art

This section gives a glimpse of some of the main state-of-the-art modalities related to the thesis and the reason for the need for an alternative one

Section 1.3: Overview

This section gives a brief the layout of the thesis.

Chapter 2: Problem Analysis

Section 2.1: Problem statement

This section sheds light on the requirements of the thesis.

Section 2.2: Related work

This section presents some of the literature work researched upon related to this thesis

Section 2.3: Using markers' pose as features

This section elucidates the development of Optical Myography (OMG) with the help of augmented reality markers

Chapter 3: Theoretical Background

Section 3.1: Hand and forearm anatomy

This section presents the essential theory of the anatomy pertaining to the thesis

This section focuses on the results from the finally selected feature source

Section 6.3: Other feature extraction source

This section presents the results of the remaining candidates of sources for feature extraction for OMG

Section 6.4: Comparison to previous approaches

This section compares the results of the chosen feature source to other modalities which use classification to estimate finger poses

Chapter 7: Conclusion and discussion

Section 7.1: Conclusion

This section gives a synopsis of the study conducted in the thesis along with its drawbacks

Section 7.2: Discussion

This section gives a glimpse into other experiments tested using the chosen feature source and also proposes developments to OMG

1 Introduction

Introduction

Prosthetic devices intend to rehabilitate a lost limb's functionality. The work focused on in this thesis pertains to Active Hand Prosthesis (AHP), where the subject can have more control over the rehabilitation device's finger motion. Target hands can be such as that of the DLR HIT HAND by the Harbin Institute of technology (HIT) and Deutsches Zentrum für Luft- und Raumfahrt (DLR), the i-limb series by Touch Bionics, and bebionic hands by Steeper. This thesis bolsters the work of the thesis conducted by N. Mouriki [1] by releasing the hand from the set-up frame and enabling the subject to obtain free-arm motion control over a virtual prosthetic hand. The subject should be able to replicate five finger poses based on his or her individual control. These are designed to mimic finger behaviours in cases of grasps, gestures and actions such as typing. The control would be assisted by the use of computer vision to detect the subject's intent and then use machine learning to predict the intended pose.

1.1 Motivation

Advancements in AHPs can prove to be expensive. Apart from the cost, additional costs can be incurred by the sensing device which extracts signals from the subject in order to carry out necessary movements of the robotic hand. The method chosen in this thesis to reduce the cost of the sensing device is to use computer vision to compute the movement of the forearm and map the computed movements to that of a virtual hand which would illustrate a robotic hand's response. Using an affordable camera, the thesis aims at reducing the cost of the sensing device and also attempts to obtain robust results of the prediction of the finger poses. The main goal of this thesis is to take a step forward from the work conducted by C. Nissler et al. in [1], [2] and [3] by carrying out experiments in a scenario where the subject's arm is capable of moving freely. The basis for the method used relies on the fact that residual muscle activity exists even after amputation [4], [5], [6]. The detection of the deformations of the forearm should also be reliable.

1.2 State-of-the-art

Current prosthetic devices include those developed by Touch Bionics [7], Ottobock [8], Motion Control [9] and RSL-Steeper [10]. Invasive methods generally require signals to be obtained from beneath the surface of the skin. These include intramuscular electromyography and electroneurography.

(a) Touch Bionics' i-limb ultra [11] (b) RSL-Steeper's bebionic hand [12]

Figure 1.1: Active Hand Prosthetic devices

Methods such as surface Electromyography (sEMG) [13], Sonomyography (SMG) [14] [15] [16] [17], and force myography (FMG) [18] [19] are under development and classified as non-invasive. Using computer vision in hand prosthesis [2] is a novel approach. The signals are collected non-invasively using a camera and markers stuck to the surface of the skin of the forearm. This is achieved by placing the arm fixed to a set-up frame using straps so that only the muscle movement can be observed. The success of the feasibility study has generated further development in the direction of testing the methodology as a more free-arm system.

Most modalities do have its advantages and drawbacks. sEMG, despite its progress in research with respect to predicting finger and hand movement, still face physical issues such as the increase in sensitivity of the sensors due to the accumulation of sweat (which alters the skin impedance and conductivity due to its salinity) between the sensor and the skin it is placed on. Although ultrasound enables one to infer changes in the deep muscles, the long term biological effects of ultrasound exposure is yet to be proved biocompatible. The force-sensing resistors (FSR) used to acquire FMG faces signal drift due to the hysteresis loss caused by the heat of the sensors over time.

1.3 Overview

The chapters constituting this thesis are outlined as follows. The main objectives of this thesis along with similar state-of-the-art modalities and a brief introduction to the development of Optical Myography (OMG) is described in Chapter 2. The necessary background knowledge pertaining to the work done in this thesis is specified in Chapter 3. The methods and materials used are elucidated in Chapter 4. Chapter 5 describes the experimental procedure used to test the work of the thesis. The results are portrayed in

Chapter 6 in the form of confusion matrices and plots depicting modality- and pose-wise accuracy, precision, specificity and sensitivity. The conclusion section in Chapter 7 gives a synopsis of the studies conducted during the thesis while the discussion section offers a glimpse of experiments tested on robustness and practicality, and potential research and development to make OMG closer towards being an alternative source of control for AHPs.

2 Problem Analysis

Problem Analysis

2.1 Problem statement

An optical modality to estimate signals to control a prosthetic hand can obviate issues related to the acquisition of physiological signals, such as those by surface Electromyography (sEMG), Force Myography (FMG) and B-mode Ultrasound (US or SMG). Using computer vision with a simple web-camera can also help in reducing the cost of the modality used, such as in Optical Myography (OMG). Placing markers on the forearm has established the proof of concept of mapping observable deformations on the surface of the forearm to the estimation of finger poses of the same. Augmented reality markers can be a good source of precise information about these deformations resulting from muscle movement. However, its small size and curvature when stuck to the forearm can lead to misdetections and imprecision in its orientation. Moreover, not all tags can be aligned to be clearly visible to the camera leading to missed detection.

This thesis aims to bolster Optical Myography by reducing the reliability of the system on acquiring precise information about the features and instead by focusing on information about the features from a holistic point of view. The thesis hence studies various sources of features (natural and artificial) and methods to extract them in order to find a suitable feature to be sent as input to the prediction model. The system must also allow for free-arm movement in order to be used in practice.

2.2 Related work

A plethora of development can be found on using surface Electromyographic (sEMG) techniques in the estimation of finger movements. The location of the signal acquisition is of key importance in sEMG. N. Celadon et al. [20] concentrated their work on finding acceptable electrode locations in the lower-arm region for finger movement. This was analysed using high-density sEMG so that individual muscle activity could be selectively captured regardless of spatial and temporal overlap of neural signals. Using Linear Discriminant Analysis as the classification tool over a non-causal Butterworth band-pass filtered signal, a good level of accuracy for selected locations of the sensors could be estimated.

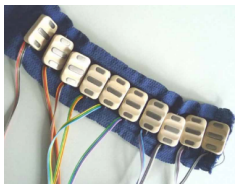
On a similar note, T. Hiyama et al. [21] used a concentric ring sEMG to distinguish between individual finger activation. In their study, they used concentric-ring sEMG as such systems have higher spatial selectivity and have less susceptibility to approximal muscle interferences as opposed to conventional EMG. The usually weak EMG signals which are susceptible to noise were high pass filtered and amplified. Root Mean

Squared (RMS) values of the EMG signals from various muscles controlling finger movement were compared in this study.

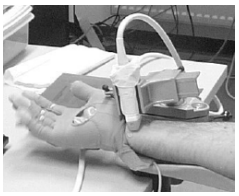
A. Gijsberts et al. [22] developed an estimation technique which could quickly adapt to natural changes in the muscle activity without much manual intervention. They did so by combining Incremental Ridge Regression and Radom Fourier Features in the prediction of hand movements. The reasons for this development were the change of myoelectric signal over time due to muscle fatigue, changing conductivity, electrode displacement and difference in the patterns produced by the user.

An alternative to EMG bases systems is force myograph (FMG). FMG measures the radially directed forces caused by the volumetric changes of musculo-tendinous complexes over which the FMG pressure sensor cuff is placed. M. Wininger et al. [23] studied grip forces of non-disabled subjects using FMG. These readings were compared to a simultaneously recorded grip force dynamometer (GFD) data in order to demonstrate that it can be used as an alternative to EMG. A study to test the feasibility of FMG signals from amputated subjects to control a robotic hand was conducted by E. Cho et al. [24]. About 6-11 grips were tested out of which 6 grips were classifiable.

Another interesting approach towards the estimation of finger motion is using ultrasound (sonomyography or SMG). S. Sikdar et al. [25] attempted to use a wearable ultrasonic system to predict finger movements of ten healthy subjects using a classification algorithm. The study prevailed that the change in ultrasound echogenicity is proportional to the flexion speed of the digits. C. Castellini et al. [15] used a B-mode ultrasound at a resolution of 1024x768 at 60 Hz. The low-pass Butterworth filtered signal was then sent through a Linear Regressor for the prediction of finger positions.



(a) surface Electromyograph [22]

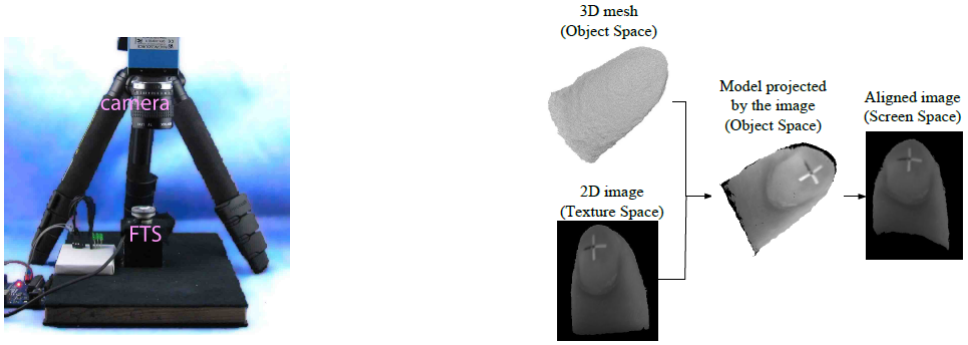


(b) B-mode ultrasound probe [15]

(c) Force myograph [24]

Figure 2.1: Other modalities researched upon for the prediction of finger poses

A work using similar modality as this thesis is that of the estimation of finger grip forces and torque by N. Chen et al. [26]. The basis for this method was to detect the colour change in the fingernail (bed) as the fingers exert forces on an object during grip. In order to do so, the hand was faced in front of a fixed camera such that the nails were in view. The images were captured during the grip at 15 frames per second (fps) with a resolution of 1024 x 768 pixels. A Convolutional Neural Network (CNN) was used here to cope with the slight movements of the finger during the various exertion of forces along time. This is done by predicting a 3-D model transformation matrix out of various 2-D images of the finger tip. The transformation matrix was formed by obtaining the orientation of a physical marker placed on each fingernail used during both training and testing. Using the 3-D model of the finger to detect and track the fingertip and further processing, the force and torques of the fingers during grip were estimated. However, the end goal for this requires fingers of the hand and can hence not be applied to estimate finger poses for amputees.



(a) Setup used to caliberate finger grip forces and torques using a force-torque sensor (FTS) and a camera

(b) Images aligned using texture mapping based on the estimated transformation matrix from the CNN with the help of markers on the fingernail

Figure 2.2: Finger grip estimation [26]

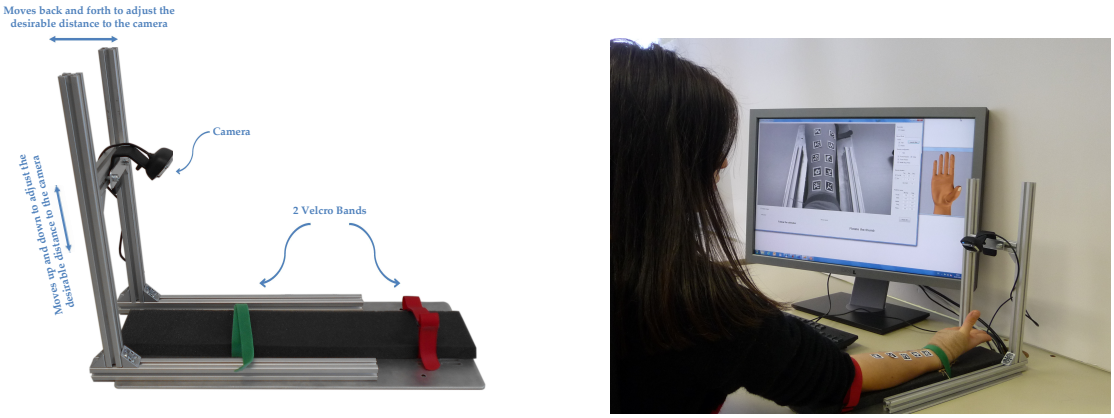
The novel approach presented by N. Nissler et al. [3] to estimate finger movements using computer vision was termed Optical Myography (OMG). By using artificial fiducials to track the deformations of the forearm during finger movement, the cost of the acquisition system could be reduced compared to previously mentioned methods. The study was focused on analysing solely the muscle movements of the surface of the forearm optically and to test its relation to that of the intended finger movements. A linear relation between the two was observed since the band-pass Butterworth filtered signals produced predictions almost as good as a non-linear when using a

linear regressor. The use of a more complex (non-linear) machine learner can hence be obviated. In order to isolate the surface muscle deformations from the gross arm movements, the subject's hand was strapped to a frame with Velcro. The fiducials (AprilTags [27]) were then tracked and used as features for the regression model. The results of the study were also tested against various optical conditions such as blur, contrast and brightness [2]. However, fixing the forearm to a set-up was just a proof of concept to test the feasibility of the novel method. The aim of this thesis is to take this work to the next stage, in letting the forearm move so that the amputee can gain wider reach with the robotic hand attached to the stump.

2.3 Using markers' pose as features

2.3.1 The existing set-up

In the work carried out in [2], [3] and in [1], AprilTags tags were stuck onto a forearm so that a mapping from the deformation of the underlying muscles to the four different finger poses (Thumb Flexion, Thumb Rotation, Index Flexion and Combo Flexion) could be used to predict the same finger poses. In order to obtain purely the muscle deformation of the forearm, gross movements of the arm were suppressed by strapping the subject's forearm to a set-up frame using VELCRO® bands such that the anterior side of the forearm, where the tags were stuck, was visible to a web-camera mounted to the frame.



(a) Set-up frame used to suppress gross arm movements using VELCRO® bands

(b) Subject participating in an experiment with aprilTags stuck on the forearm

Figure 2.3: Set-up [2] and [3] during OMG using aprilTags with the subject's forearm strapped to the frame to suppress gross arm movements

The tags were then detected off-line and their orientations (translation and Euler rotation along x-, y-, and z-axes) were computed. These signals were filtered using a second order Butterworth bandpass filter and then sent through a ridge regressor one pose at a time, i.e. the regressor is exposed to just one of the finger poses during a training session.

2.3.2 Adaptations to the existing set-up

Since one of the goals of this thesis is to develop OMG into a free-arm system, the arm is released from the set-up. In order to do this, absolute (camera to marker transformation) poses of the AprilTags cannot be used any more since the values fed to the regressor would be not just that of the muscle deformation, but also that of the gross movement of the forearm about its environment. This can introduce new values of orientation to the regressor every time the arm enters a new position during the gross movement. Having the camera attached to the forearm could remove the gross movements of the forearm, however, the tags become too affine for it to be detected by the lexicode reader. The only option left in this scenario is to have the camera fixed away from the forearm and restrict the forearm within the view of the camera.

With this set-up, instead of using the absolute orientations of each of the tags as input to the regressor, relative orientations between the tags is computed. This is done by using a tag-perspective transformation, where the perspective of one tag with respect to each of the tags around it is determined by multiplying the homogeneous matrix of the other tag with the inverse of its own homogeneous matrix. The translation and rotation relative to the two tags can now be extracted from the resulting 4x4 homogeneous transformation matrix without the effects of the larger movements of the arm.

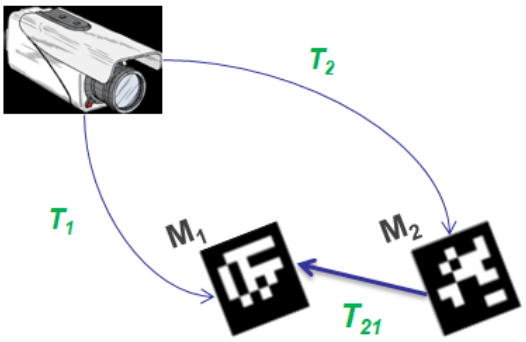


Figure 2.4: Using the tag-perspective transformation (T_{21}) to bring marker M_2 in the perspective of marker M_1 using the observed T_1 and T_2 camera to marker transformations.

The tag-perspective transformation from the above diagram (Figure 2.4) can be computed using:

$$T_{21} = T_2^{-1} * T_1 \tag{2.1}$$

where:

T_{21} = Tag-perspective transformation of marker $M2$ to marker $M1$
T_2 = Camera to marker $M2$ perspective transformation
T_1 = Camera to marker $M1$ perspective transformation

The anterior side of the forearm was restricted to move within the view of the camera, but this did not prevent the tags from not being detected. The AprilTags were then modified from the 36h11 coding (36 bits with a hamming distance of 11) to that of the 25h11 family (25 bits with a hamming distance of 11). This increased the dimensions of the black and white chequered boxes, albeit producing lesser number of tags (seven tags). The tags thus placed on the forearm reduced from 10 to 6 semi-symmetrically placed tags on the hand. This did increase the detection rate, however due to the small size (a length and breadth of 1.5cm) of the tags, not all the tags were correctly detected. Neither were they always detected.

(a) A sample of the 36h11 AprilTags family used previously

(b) A sample of the 25h11 AprilTags family

(c) A sample of the ArUco tag

Figure 2.5: Comparison of the AprilTags families used and the ArUco used to test for jitter in the calculated orientation of the tags. The 25h11 was chosen over the 36h11 family due to its larger and thus more discernable lexicode prints and its high variation

The relative tag orientation did serve its purpose on obtaining good predictions while the arm was moving within the camera's field of view. It also performed slightly better on the data collected when the hands were strapped to the set-up frame. However, the following drawbacks exist upon using such markers in OMG, even when tested with ArUco Markers [28], a similar kind of marker-based detection system commonly used in Augmented Reality:

1. Missed detections and incorrect (false) detections caused by:
   a) Small size of the markers and thus their lexicographic code when stuck to the forearm
   b) Tags falling out of view of the camera
   c) Tags viewed affine
   d) Glares
   e) Motion blur during fast movement
2. Jitter (imprecision, in Figure 2.6) in the orientation caused by:
   a) Small size of the markers when stuck to the forearm
   b) Curved surface of the tags due to the curvature of the arm
   c) Tags viewed affine
3. Tag-perspective transformation does not hold when the camera is placed on the forearm since the imaging plane keeps moving as the camera moves along with the deforming muscles.

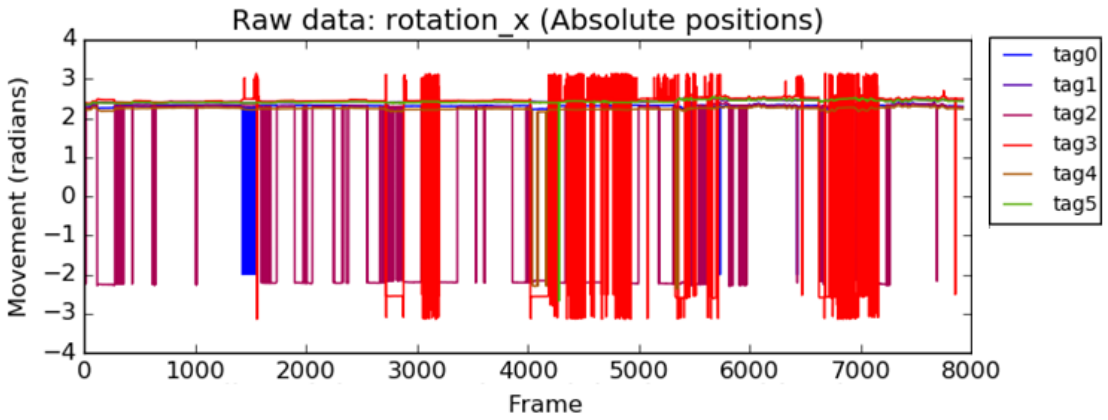
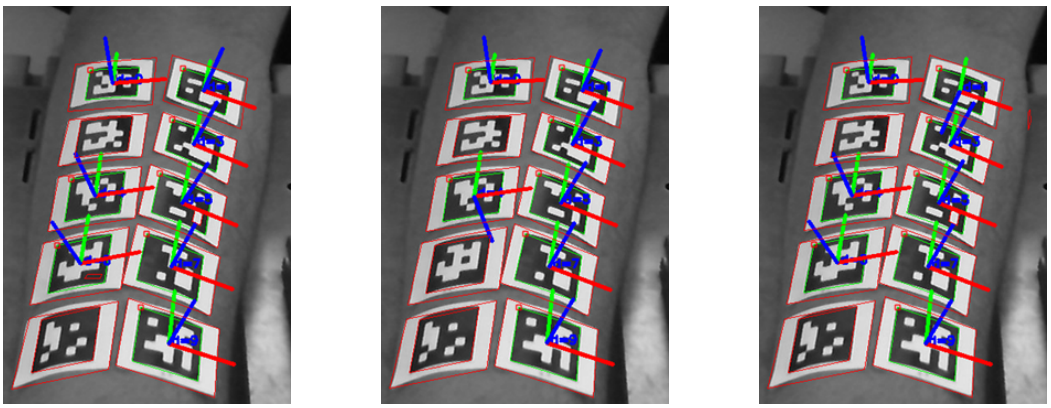


Figure 2.6: Jitter in the (absolute) rotation along one of the axes (x-axis) for each of the 6 tags of the 36h11 AprilTags family. Similar noise was also observed along the other three axes.

The incorrect detections can be reduced by making sure that the tags are designed to be more distinguishable from the tags produced along with it thus making the chances of confusing a tag for another less. Upon using thicker paper to give each tag a more rigid flat surface when stuck at one point on the forearm, the reduction in the jitter is

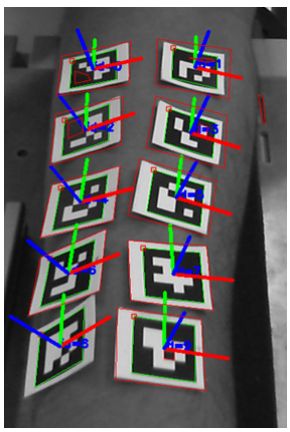
prominent. However, the possibility of the tags facing away from the camera is high once the muscles or the arm move. These two solutions are not sufficient to completely eradicate the jitter. Feeding missing and noisy data to a machine learning algorithm can decrease its efficiency. Thus, a call to determine a feature which relies less on precision of the data itself is called for in this thesis.

In the following figures (Figures 2.7, 2.8 and 2.9), the x-axis is represented by red, the y- by green and the z- by blue. The figures illustrate the various approaches used to test for when the system starts to become unstable. Although most of the flips seem to be 90°or 180°around one axis, the values observed also consisted of flips of various angles. This could mainly arise from the small size of the tags used (prominent in Figure 2.9) and also due to its curvature and how affine they appear to the camera hence leading to uncertainty in the exact perceived orientation with respect ot the camera.

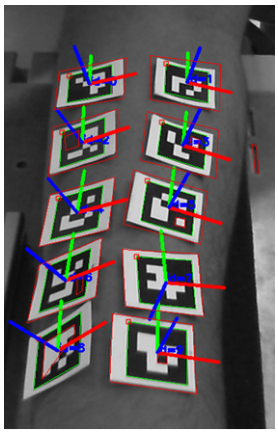


(a) Tags on forearm: Frame 1 (b) Tags on forearm: Frame 2 (c) Tags on forearm: Frame 3

Figure 2.7: Using the ArUco to test for flips in the orientation of the tags when stuck to the hand. The flips occur in the third tag from the top on the left column at Frame 2 (2.7b) and in the first tag on the right column at Frame 3 (2.7c). One an also notice the bottom tag on the left column of all the frames and the tag above it in the second frame not being detected. The tags were of size 2cm in length and breadth

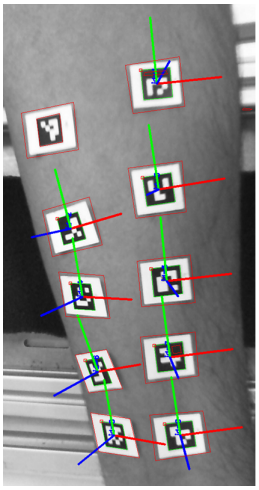


(a) Flat tags on forearm: Frame 1

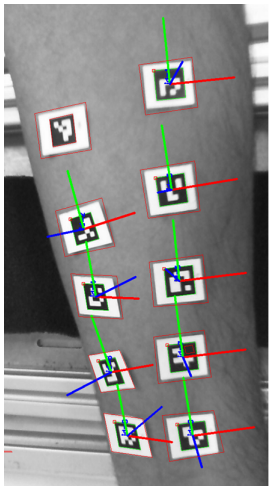


(b) Flat tags on forearm: Frame 2

Figure 2.8: Using the ArUco to test for flips in the orientation of flat tags when stuck to the hand at a point using a thicker paper. The flip occurs in the 4th tag on the right column at Frame 2 (2.8b). The tags were of size 2cm in length and breadth



(a) Small tags on forearm: Frame 1



(b) Small tags on forearm: Frame 2

Figure 2.9: Using smaller ArUco to test for flips in the orientation when stuck to the hand. The flips occur in the third tag on the left column and in the third tag of the of the second column due to imprecision in the assertion of the orientation; both in Frame 2 (2.9b). The first tag on the left column is not detected in both the frames. The tags were of size 1cm in length and breadth

3 Theoretical Background

Theoretical Background

3.1 Hand and forearm anatomy

The information discussed in this section is collected from [29] [30] [31] [32] [33] [34] and has been hand picked so that only those pertaining to the thesis shall be covered in brief.

3.1.1 Hand

The hand and forearm have more than 30 individual muscles that work together to carry out complex movements. Those located within the hand, the intrinsic muscles are responsible for the fine motor functions of the hand. The fingers are connected to the small muscles via tendons. The thenar muscles are three short muscles located at the base of the thumb and help in the fine movements of the thumb. Each of the four lumbricals in the hand are associated with a finger for its movement. They link the extensor tendons to the flexor tendons and thus help in the flexion at the metacarpophalangeal (MCP) or knuckle joint and extension at the interphalangeal (IP) or finger joint of each finger.

The movements of the fingers followed in this thesis are:

- (MCP) Flexion: Moving the base of the finger towards the palm.
- (MCP) Extension: Moving the base of the fingers away from the palm.
- (IP) Flexion: Moving the last two segments of the finger towards the base of the fingers.
- (IP) Extension: Moving the last two segments of the finger away from the base of the fingers.
- (Thumb) Abduction: The thumb is carried forwards away from the palm. Occurs at right angles to the palm with range of about 80°.
- (Thumb) Adduction: The thumb is moved back towards the palm. Occurs at right angles to the palm with range of about 80°.
- (Thumb) Opposition: Movement in which the distal pad of the thumb is brought against the distal pad of any of the remaining digits.

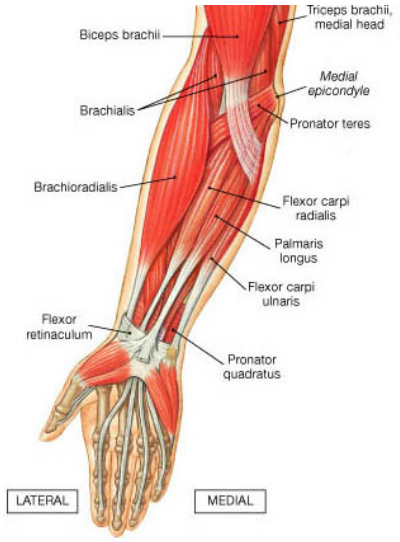
The opposition of the thumb initially occurs by its simultaneous flexion and abduction at the capometacarpal joint through the stimulation of the lexors pollicis longus and

brevis and abductor pollicis longus. The rotation is directed medially due to the posterior oblique ligament becoming taut. During this, opponens pollicis contracts to produce an active rotation of the metacarpal. The third elementary movement is the adduction at the carpometacarpal joint produced by adductor pollicis which bring the metacarpal back towards the plane of the palm of the hand. Movements of the thumb at the MCP joint contribute significantly to the overall movement of opposition. The carpometacarpal joint and the MCP joint flex and abduct simultaneously. The simultaneous movments of flexion and abduction of the proximal phalanx bring about a degree of axial rotation at this joint. The pad of the thumb faces posteromedially consequently.

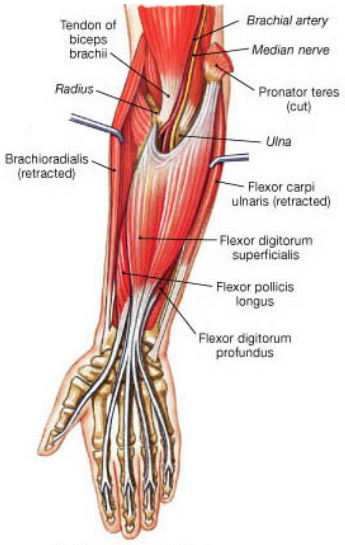
3.1.2 Forearm

Most of the muscles that move the wrist, hand, and fingers are located in the forearm. These muscles extend from the humerus, ulna and radius and insert into the carpals, metacarpals, and phalanges via long tendons. The muscles on the anterior side of the forearm, such as the flexor carpi radialis and flexor digitorum superficialis, form the flexor group that flexes the hand each of the phalanges and at the wrist. For this reason, the camera will be facing the anterior side of the forearm and further discussion shall be focused on this region.

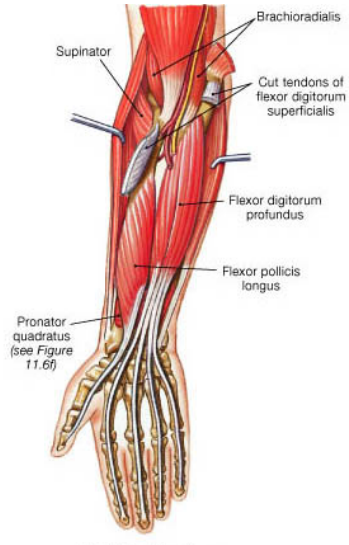
The anterior forearm is comprised of four superficial, one intermediate and three deep muscles. The superficial group (pronator teres, flexor carpi radialis, plamaris longus and flexor carpi ulnaris) arises mostly from a common flexor tendon that attaches to the anterior part of the medial epicondyle of the humerus. The intermediate muscle (flexor digitorum superficialis) arises from this common tendon and along the anterior surface of the ulna and radius and is also supplied by the median nerve. The deep group (flexor pollicis longus, flexor digitorum profundus and pronator quadratus) is supplied mostly by the anterior interosseous nerve, a branch of the median. The rest are supplied by the ulnar nerve. The tendons of the flexor carpi radialis, palmaris longus, and flexor carpi ulnaris are readily palpable. The palmaris longus is often absent.



(a) Anterior superficial muscles



(b) Anterior intermediate muscles



(c) Anterior deep muscles

Figure 3.1: Anterior compartment of the forearm [35]

Increased pressure in the anterior compartment of the forearm due to injury to the brachial artery near the elbow can prevent normal blood flow to the compartment and thus can cause ischemic damage to the deep flexors. This results in muscle scarring, with flexion deformity of the wrist and fingers (Volkmann's ischemic contracture).

3.2 Image processing

The region that is fed in as input to the Convolutional Neural Network (CNN) is the anterior section of forearm. The forearm however is a rather featureless region in the visual spectrum. This poses a challenge to visually determine changes in the forearm when its muscles are stimulated. By placing artificial markers on the hand such as AprilTag or ArUco markers, this issue can be tackled but with the assumption that the tags are always visible to the camera and can be detected precisely and without false detections. As discussed in Section 2.3, there can be certain drawbacks when relying on such data. Alternative approaches to detecting the deformations on the hand have thus been studied in this thesis and shall be discussed in this section.

3.2.1 Natural features

Taking a step beyond the visual spectrum of the electromagnetic radiation into near infra-red (NIR) wavelengths, one can observe the absorption of the NIR waves by the superficial veins on the anterior side of the forearm [36]. This is done by modifying a web-camera physically into an infra-red camera and processing the image using contrast stretching [37]. The image (Figure 3.2a) is then sent as input to the CNN after filtering and segmenting (Figure 3.2b).

This method works better to a certain extent with controlled NIR [38] through the use of NIR Light Emitting Diodes (LEDs). However, since this thesis intends to find a simple yet reliable solution to obtaining features on the hand, additional devices (such as LEDs and a good diffusion material to suppress the light patterns caused by the LED array) besides the camera are not used. This means that the NIR can be overpowered by stronger visible radiation from the surrounding environment, especially when manifested as glares. The use of artificial fiducials are thus also studied as a candidate for features to be extracted.

3.2.2 Artificial fiducials

Hand drawn grids (See Figure 3.2c) on the forearm of target hand can act as an artificial fiducial to observe the deformation of the muscles on the skin. The higher the density of the grids, the more the chances of the CNN to extract the features are. The extraction of the region of interest of such hand-drawn grids can be a bit complex as discussed in the following section (Section 3.2.3).

Thus, using a sticker stuck to the forearm with hand-drawn grids on top of it (as in Figure 3.3a) is also studied. The sticker's paper needs to be slightly stretchable in order to emulate the deformation of the parts of the forearm it is attached to. This however,

upon extraction results in the deformational changes of the boundaries of the sticker to be more perceivable than the grids drawn within it.

Sticking a paper to the forearm without grids drawn within it as in Figure 3.3c is easier to detect (Section 3.2.3) since the strong edges of the grids can sometimes be mistaken for the edges of the sticker itself especially when viewed at affine angles. It also serves as a test to see if the grids drawn on the sticker play a major role when compared to the paper's contours itself.

3.2.3 Feature extraction

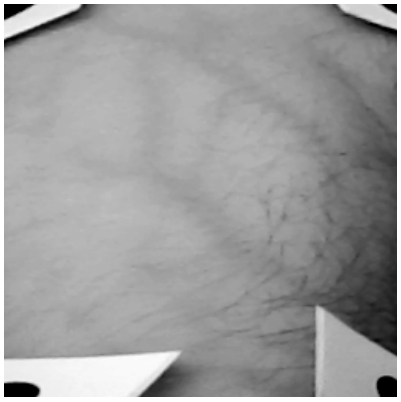
The forearm is rather featureless in the visible region of the electromagnetic spectrum and thus has no reliable visual landmarks that can be used as the bounding region for the images that will be fed into the CNN. The camera faces the anterior side of the forearm. However, since most of the length of the forearm needs to be studied for the variation in deformation of its underlying muscles, the web-camera also records the background environment. Feeding this to the CNN can be less of a problem if the machine learns from enough images that the background is irrelevant to the intended classification. Since the data collected for training is not practically large enough for the machine to learn the difference between foreground and background by itself, the simplest solution is to remove the latter from the frame before feeding it to the machine learner.

While using the natural features and the hand-drawn grids on the bare forearm, the region of interest (ROI) is marked using physical markers stuck onto the forearm within the view of the camera. These physical markers are rectangular pieces of white paper with a black circle in the centre (for maximum contrast) of each paper. Four such markers are placed on the four corners of the ROI which are warped as the corners of the CNN's rectangular input image using image processing. The markers are chosen by manually clicking on the black circles of the greyscale image of the first frame. The centres of each of the circles are computed and these would then be the source of the predicted centres for the upcoming frames of the respective circles; hence semi-automating the process.

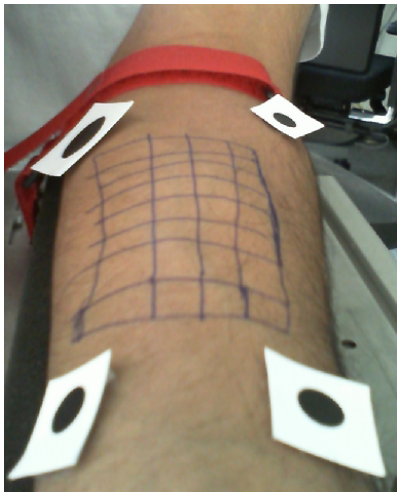
Although the ROI computed from the circles' centres can be accurate, the robustness drops when a frame fails to record. If the camera is placed at a height, the circles become smaller and the muscle movements of the forearm displace the centre of the circle from the last recorded frame; meaning that the centre now lies outside the circle. The flood fill algorithm used to recalculate the new centre of the black circle now fills regions apart from it. This springs up a new centre and thus a change in the input image due to falsely determined ROI. Placing the camera just above the surface of the arm also results in a similar issue as the physically placed markers are now viewed at

an affine angle. Blob detection relies on a high true positive rate for all four markers which is not feasible in real-time.

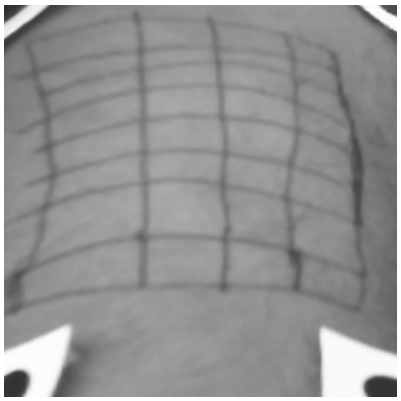
(a) Natural features: NIR input image of veins viewed in greyscale



(b) Natural features: NIR output image in greyscale with veins selected as the ROI



(c) Artitficial fiducials: RGB input image of hand-drawn grids viewed



(d) Artitficial fiducials: Output image in greyscale with hand-drawn grids selected as the ROI

Figure 3.2: Input (640x480 pixels, cropped and rotated here in 3.2a and 3.2c) and their respective output (256x256 pixels in 3.2b and 3.2d) images using the four physical markers (black circle on a white rectangular paper) as the corners for the ROI

This lead to the use of stickers on the forearm which emulate the deformation of the forearm regions it is stuck to. The sticker needs to be visually distinct from the skin in order for it to be extracted as the ROI. Various colour spaces are used to segment human skin from an image [39] [40] [41]. Recent studies have revealed that using the log-opponent chromaticity (LO) space can help in the segmentation of the skin [42] [43].

The LO colour space mimics the human visual system by not perceiving a few colour combinations together. It makes use of the Red, Green and Blue (RGB) colour space recorded by the camera and converts it into log opponents [44] [43] using:

$$L(x) = \frac{1.0}{\ln 2} * \ln(x+1) \tag{3.1}$$

$$I = \frac{L(R) + L(G) + L(B)}{3} \approx \frac{L(G)}{3} \tag{3.2}$$

$$R_g = L(R) - L(G) \tag{3.3}$$

$$B_y = L(B) - \frac{L(G) + L(R)}{2} \tag{3.4}$$

where:

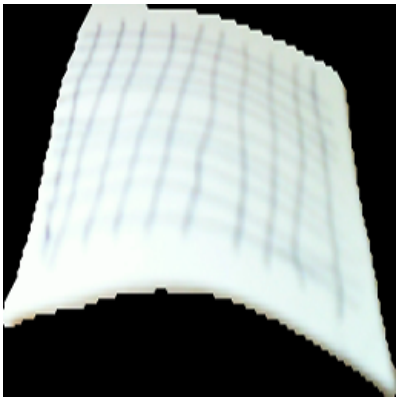
- ln = Natural logarithm
- R = Red colour channel
- G = Green colour channel
- B = Blue colour channel
- I = Intensity channel
- R_g = Red-Green opponent channel
- B_y = Blue-Yellow opponent channel

The log transformation makes the R_g and B_y values independent of illumination level and can be viewed as a simple translation in the chromatic distribution. The numerator in equation 3.1 is 1.0 for images of type float, or 255 for integer type images. In equation 3.2 the green channel can be assumed to represent intensity since sometimes the red and blue channels from cameras have poor spatial resolution. This assumption is acknowledged in the code in order to reduce computation.

Once the sticker can be distinguished better after filtering (using filter sizes larger than the grid's thickness so that the grid lines are not perceived as the sticker's edges) and improving the histogram, an adaptive threshold can be run on the image in order to draw the thresholded contours of the image. Morphological operations such as opening and closing are performed to connect broken contours and disconnect different contours. The contour detection by OpenCV, *findContours*, finds closed contours thus calling for the sticker's contours [45] [46] to be closed and distinct. The box bounding

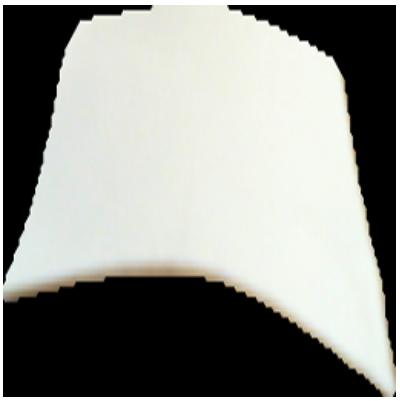
the contour of the sticker is then chosen as the region of interest and with the pixels lying outside the mask of the contour set to a pixel value of 0 (black).

(a) Input image of hand-drawn grids on a sticker stuck to the forearm



(b) NIR output image in greyscale with veins selected as the ROI

(c) Input image of a sticker stuck to the forearm



(d) Output image in greyscale with hand-drawn grids selected as the ROI

Figure 3.3: Input (640x480 pixels, cropped and rotated in 3.3a and 3.3c) and their respective output (256x256 pixels in 3.3b and 3.3d) images taken from a normal RGB web-camera and using the bounding box of the sticker stuck to the forearm as the ROI

3.3 Machine learning

Machine learning is used in order to predict the finger poses based on the information collected from the forearm. The machine learning algorithm chosen for this set of input images is the Convolutional Neural Network (CNN). Unlike in the approach carried out previously [1], [2] and [3], the machine learns to classify rather than regress the data. As stated by Tom M. Mittchel, *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."* [47], the task T is the classification between the finger poses, the experience E is the tuning of the model based on the prediction error P incurred upon predicting the pose from data fed to the machine during its training. The machine uses a supervised learning approach where each of the images recorded are labelled to the respective finger pose the subject is expected to follow.

3.3.1 Convolutional Neural Network

Inspired by the organization of the visual cortex of animals, the Convolutional Neural Network (CNN) is a type of feed-forward artificial neural network in which the response of an individual neuron to stimuli within its receptive field can be approximated mathematically by a convolution operation.

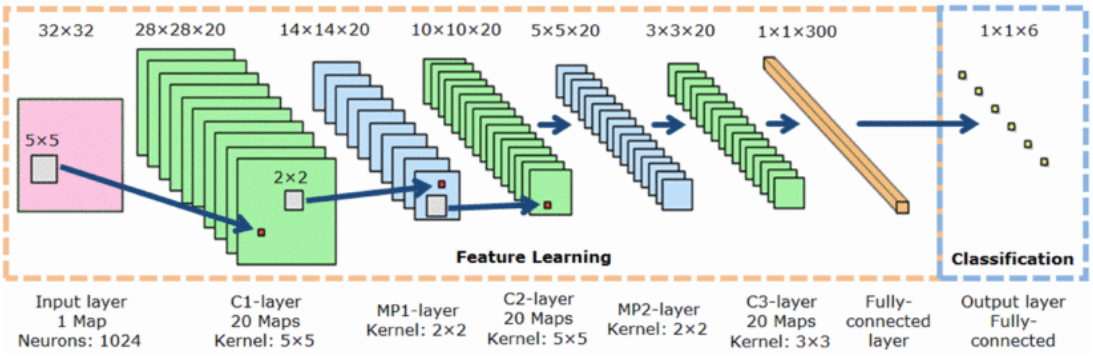


Figure 3.4: A typical Max-pooling Convolutional Neural Network (MPCNN) used for classification [48]

A typical CNN is modelled with at least one convolutional layer and a fully connected layer. Sub-sampling techniques such as max-pooling of the image after a convolutional layer is said to have its role in increasing the speed of learning when there are many such convolutional layers and also in improving robustness against noise and small changes in the data [49] [50] [51] [52]. Thus, adding max-pooling layers is also a trend

in networks with at least two convolutional layers. Some have recently found ways to obviate this by placing more convolutional layers and changing the strides of the kernels [53]. The more the number of such layers, the deeper the network is. Other architectures also employ a time delay neural network in which images or data can be analysed based on temporal information [54] [55]. Since the variation between the forearm's deformation during each of the finger poses is not too distinct and not too much, this thesis uses pure classification based on training the end poses rather than the intermediate (transition between the rest and the target) poses.

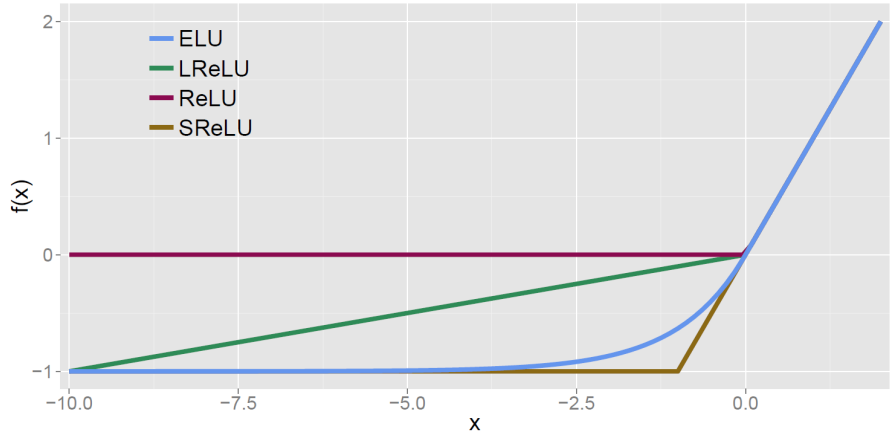


Figure 3.5: Comparison of some activation functions [56] such as the Exponential Linear Unit (ELU) with $\alpha = 1.0$, Rectified Linear Unit (ReLU), leaky ReLU (LReLU) with $\alpha = 0.1$ and shifted ReLUs (SReLUs)

The model learns by taking the images (usually in a predefined batch size) as input and then passing it to the first convolutional layer. In this layer, the machine tries to determine appropriate weights for defined amount the convolutional kernels. Due to the receptive field of the convolutional network, the learnt filters produce the strongest response to a spatially local input pattern. The activation function used for this is the Exponential Linear Unit (ELU). The ELU is formulated as:

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha\left(\exp(x) - 1\right), & \text{if } x \leq 0, \alpha > 0 \end{cases} \tag{3.5}$$

Here, α is the hyperparameter which controls the ELU saturation for negative net inputs. The mean of the activation is closer to zero due to the negative values allowed by ELU, thus leading to a faster learning by bringing the gradient closer to the natural gradient [56].

The resulting images are then passed over to the next layer. If the next layer is a sub-sampling layer, each image is reduced in size based on the kernel used to sub-sample the image. In the case of max-pooling, this would be to take the maximum pixel value within the overlaying kernel, thus reducing the image size (by 50% for a 2-D image and a max-pooling kernel size of 2). The fully connected layer then takes all the activations with a bias to find the predictions for the labels of the supervised input data. The predictions are trained using a loss function which attempts to minimise the loss in prediction. This is done over a number of set or conditioned iterations using optimization techniques such as Gradient decent.

4 Methods and Materials

Methods and Materials

4.1 Experimental set-up

4.1.1 Hardware

The computer used for recording the data during the experiment was a Windows 7 (64-bits) PC with 6 GB of Random-Access Memory (RAM) and a 2.80 GHz Intel® Xeon® processor. The images are processed on a Linux based system (64-bits) with 12 GB of RAM and a 2.80 GHz Intel® Xeon® processor. The graphics processing unit (GPU) used to run the CNN is a GeForce GTX TITAN X with a memory of 12 GB, run on a Linux based system with a similar processor but with 125 GB of RAM.

An elastic band is used to strap the camera onto the forearm. The camera used is the same as in [2] and [3] - a Microsoft® LifeCam HD-3000 web-camera with a focus at 3mm-1.5m, a resolution of 1280x720 pixels and an achievable frame rate of 30 FPS. But the images are recorded at a lower resolution of 640x480 pixels and a higher frame rate (25 FPS instead of 15 FPS). The frame rate however is set higher to allow for reduction in motion blur and also to counter any interference in frequency caused by the indoor artificial lighting. Artificial lighting (from LED tube lights) is used in these experiments to achieve uniformity in the experiments and to avoid changes in illumination caused by natural sunlight over the time of recording the various subjects. Motion blur was also physically suppressed to a certain extent by using a VELCRO® band around the forearm and the camera to prevent upwards vertical movement; with the assumption that when applied practically, the camera would be inbuilt or mounted to the base of the AHP.

4.1.2 Software

Since the graphical user interface (GUI) is the same as that used in [2] and [3], the system used to run the C# based GUI was the Windows 7 system (described in Section 4.1.1). The GUI sends User Datagram Protocol (UDP) signals to a 3-D hand model whose finger poses the subject is asked to follow. The video recorded via the GUI is saved as images along with data regarding the corresponding finger poses.

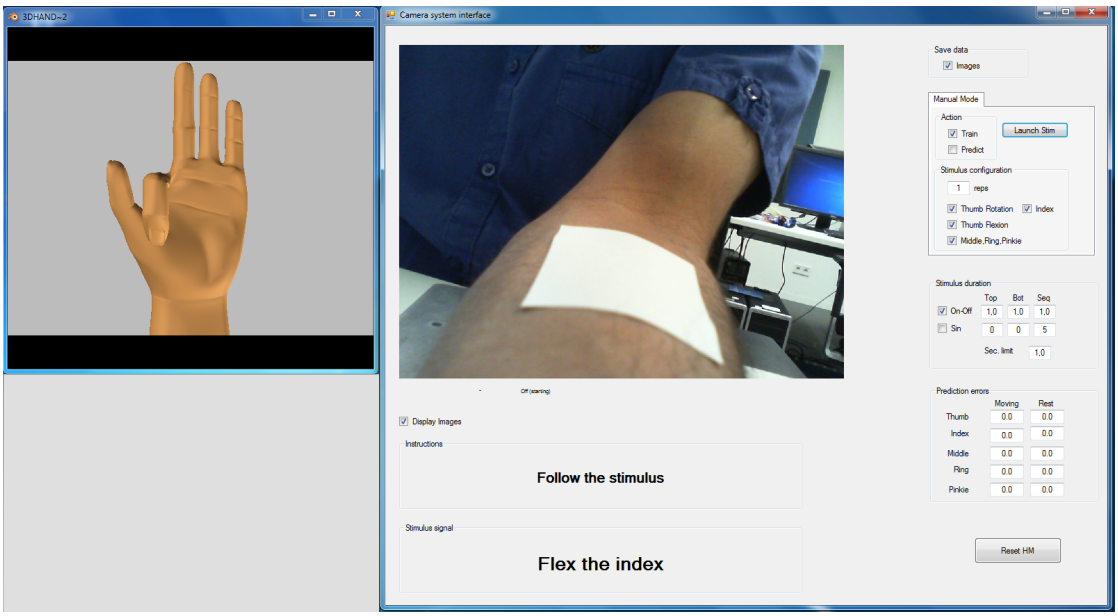


Figure 4.1: The GUI (right) used to record the forearm as the subject follows the stimuli displayed by the 3-D hand model (the window on the left)

The reason that the images are processed and trained on a Linux based system is because the language used to process and train the images is Python 2.7, a language used by the TensorFlow™ software library [57], which is at the moment more easily set up on Linux. The images that are fed into the CNN are processed to obtain the ROI using OpenCV [58] as the main library.

4.2 Image pre-processing and extraction

As this thesis seeks to find the most reliable features that can be extracted from images of a forearm, the image processing varies based on the candidates used as sources for the features. The candidate source of features selected from Sections 3.2.1 and 3.2.2 is the plain white sticker stuck to the forearm. Besides this candidate, the remaining shall be discussed in brief. The following flowchart (Figure 4.2) gives an overview of the off-line process used to extract the ROI which is then fed into the CNN for training and testing.

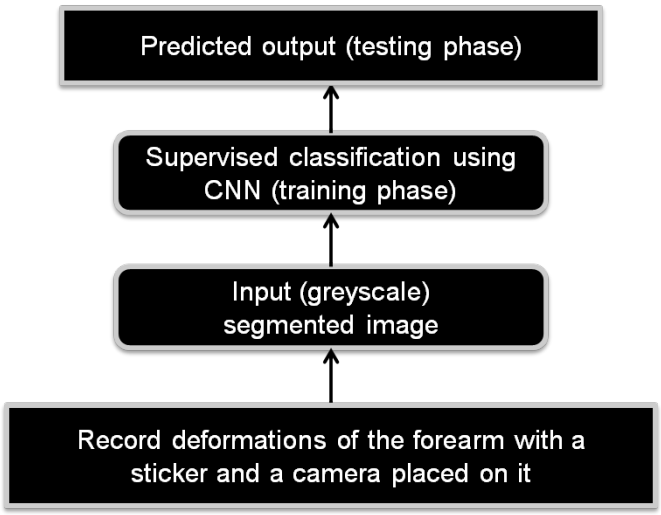


Figure 4.2: Overall flowchart of Optical Myography using Convolutional Neural Networks to estimate finger poses off-line

4.2.1 With physical markers placed on the forearm

As described in Sections 3.2.1 and 3.2.3, the centre of the circles within each four physical markers are used as corners of the final ROI. The subject's arm was strapped onto the frame with the camera mounted onto the frame. The process used to extract the ROI is carried out by manually selecting the four circles in the first frame which then computes centre for each circle used as coordinates for their respective ROI corners. The computed centres then become the estimate of the centres of the succeeding frame's circles, thereby automating the process for the remaining frames. The image is cropped to the bounds of the ROI and is resized to a size of 256x256 pixels. The CNN used is the same as that described in Section 4.3. The segmented images (containing either subcutaneous veins within the ROI or the hand-drawn grids) fed into the network is resized to 130x130 pixels.

4.2.2 With a sticker placed on the forearm

When a sticker is placed on the forearm for the extraction of features, the ROI can be the sticker itself. Since the chosen source (reasoned in Sections 3.2.2 and 3.2.3) is a plain white sticker and not a sticker with grids drawn on it, the following section will revolve around using the former, although the process is the same except for the larger kernel sizes used to blur out the grid's lines (as explained in Section 3.2.3). The 2-D image is first downscaled to half the size for speed and ease of segmentation. In

order to segment the sticker, the RGB colour space is transformed into a three channel LO chromatic space. In order to cope with intensity issues such as glares, the intensity channel (I channel) is subtracted from the R_g channel and this one new channel is used as the base for the rest of the image processing upon normalization. The R_g channel is used since the segmentation of the paper is a lot easier.

A fixed ROI is set for the first frame in order to remove unnecessary background captured due to the camera's zoom. A median filter smoothens the image before an adaptive threshold is used to segment the sticker and other boundaries. The image is then resized to its original size. Morphological opening (vertical and horizontal line masks) and closing (square shaped mask) are used in cases where the sticker's boundary merges with the forearm's or the sleeve's due to its placement. The contours of the image are then extracted and are enclosed within a rectangular bounding box. The contour of the sticker is filtered by placing the following conditions on the bounding box:

- The dimensions of the contour exceed 50 (in case the sticker is small or stuck far) or 100 pixels
- The sticker lies completely within the ROI and within a safe distance from its boundaries
- The sticker's position and size does not change drastically
- The contour does not have more than 10 sides. (To account for the curvature of the forearm the sticker takes on when stuck onto it.)

Once the estimated contour and bounding box of the sticker is verified by the user in the first frame, a mask of the contour segments only the sticker and sets the regions outside the contour to a pixel value of 0 (black). The resulting image is then resized to a shape of 256x256 pixels. The RGB channel is restored to display the segmented sticker and to compute the normalized cross-correlation (NCC) during the succeeding frames.

The bounding box of the sticker of each frame is then used as the new ROI for the succeeding frames. The threshold used for the NCC between the newly segmented image and that of the preceding frame is set high (above 76%) to ensure that the sticker is always selected correctly; hence automating the process after the first frame has been manually validated. The process runs through the images captured and saves the final segmented image (256x256 pixels, RGB image) at a rate of around 23 FPS without a GPU.

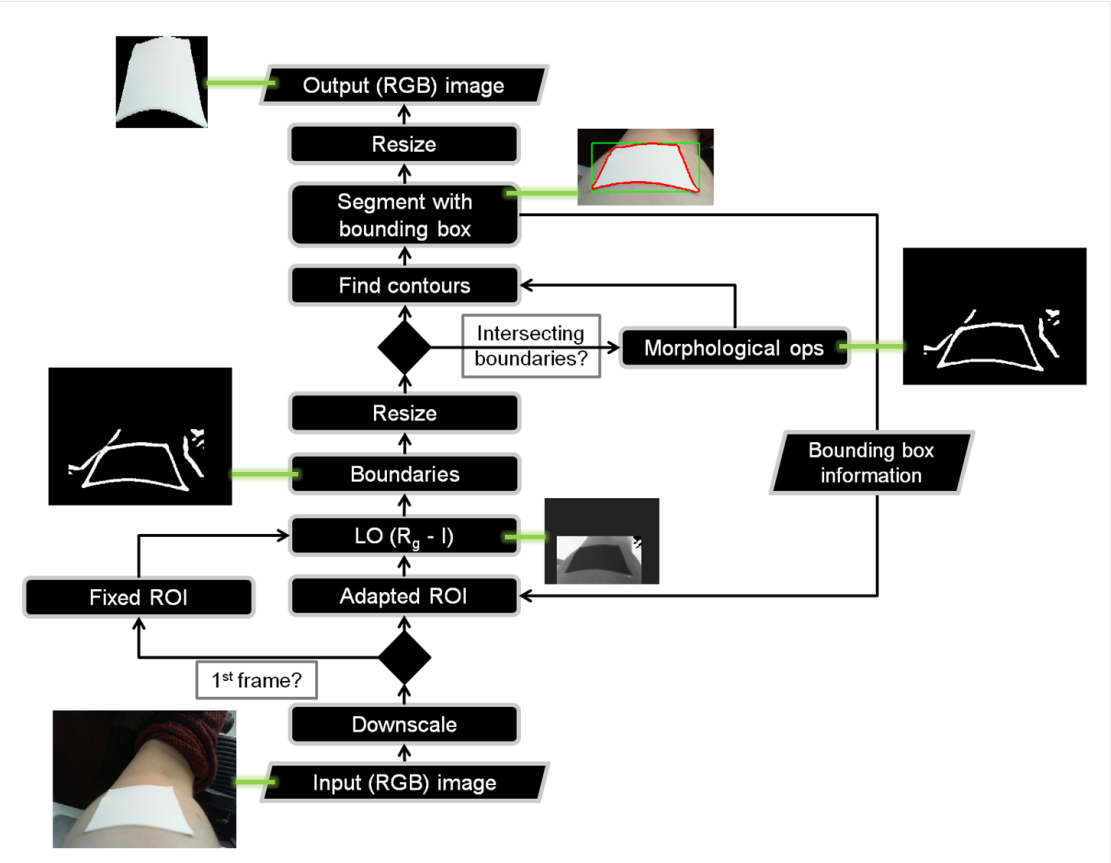


Figure 4.3: Flowchart depicting the segmentation of the sticker on the forearm

4.3 The network

The final segmented image is now ready to be sent as input to the machine learning network. The network is a simple CNN with 2 convolutional layers with ELU activation and a fully connected layer. Each convolutional layer consists of 16 filters. The filters in the first layer have a relatively large size of 11x11 pixels to account for positional changes of the sticker on the forearm. The second layer's filters are of size 5x5 pixels to obtain dependable features. The images sent as input to the CNN are greyscale images of size 130x130 pixels and are fed in batches of 70 images each. The image dimensions and the batch size are restricted by the capacity of the GPU's memory used during the training and testing. Hence, larger batch sizes could not be tested. Image sizes are also restricted to be smaller than the down-scaled image to preserve information.

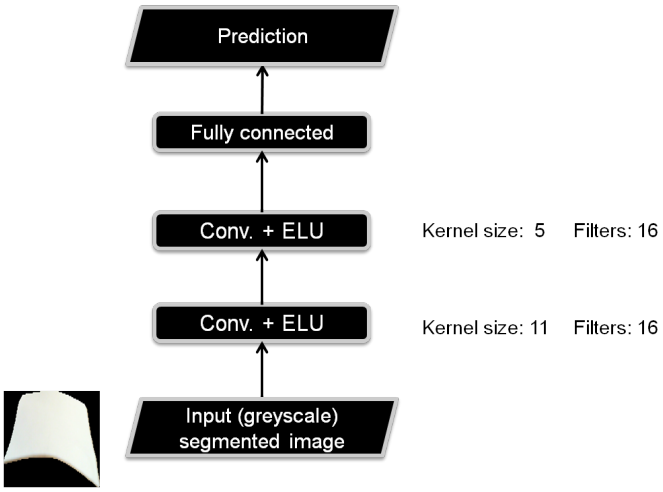


Figure 4.4: The CNN network used to classify 5 poses of the fingers

4.3.1 Training

The GUI used to record the images also saves the corresponding UDP (used as labels for the supervised learning) signals sent to the 3-D hand model. However, this does not include the specific Rest pose's UDP signals. In order to deduce the Rest pose for both the training and testing for a 5 pose classification, the timestamps where no UDP signals are sent to the 3-D hand model are considered as the resting pose.

Intermediate finger poses are also removed by setting a threshold for the UDP signals (ranging from 0 to 1.0) at 0.995. Thus, all signals above the threshold are set as 1 (maximum) and all those below are reset to 0, i.e. intermediate. Delay in the reaction time of the subject was also accounted for by removing the frames which occur within the first 20 of those after each switch of the stimulus (pose). This can vary based on the frame rate, but since the frame rate is fixed at 25 FPS, about nearly a second of reaction time is sufficient to be considered here (800 ms or 0.8 seconds).

The data sent in for training are shuffled in a pseudorandom process so that each of the stimuli is uniformly distributed through out the training set. This is ensured to be even by trimming off all stimuli to the number of occurrences of the least. The training set is then split into batches such that each batch contains 70 images and its corresponding 70 labels. Each batch is then sent through a CNN with two convolutional layers of 16 filters each; the first having filters of size 11 and the second, 5. The first layer intends to capture positional changes of the sticker since these are some of the differences one can visually observe between the various poses. The convolutions are activated by ELU and the fully connected layers realise the relation between all the

parameters (kernel weights) and reduces the final layer to a prediction level. The final layer is thus of size 70x5; 70 images in the batch and 5 prediction values for each of the finger poses.

The predictions are estimated by optimizing the parameters of the CNN using gradient decent which minimizes the loss over 19 iterations. The initial learning rate is 0.001 and is then decreased by 5% every succeeding iteration. The loss function is that of the mean of the sparse softmax cross-entropy of the output of the final fully connected layer (the logits). The training time when using one of the GPUs mentioned in Section 4.1.1 is around 40 minutes (for 9 repetitions).

4.3.2 Testing

The data sent in as the test set is not shuffled but sent directly to the CNN. The learned network is derived from a stored check-point saved after the training iterations are complete. The CNN model must hence be the same as that used during the training. The batch size in which the images are sent during the test is also the same as that during the training. The accuracy is calculated by finding the argmax of the logits and comparing it to the true labels. The time taken to test using the one of the GPUs mentioned in Section 4.1.1 is less than 10 seconds (for 1 repetition).

5 Experiments

Experiments

5.1 Workflow of the experiment

The experiments are carried out in compliance with the World Medical Association's *Declaration of Helsinki*, regarding the ethical principles for medical research involving human subjects, last version as approved at the 59th WMA General Assembly, Seoul, October 2008. The hardware used are CE approved and clinically safe to use. The modality is tested as mentioned in this section.

5.1.1 Participants

The participants chosen for the experiment were people with all their fingers intact. Since this thesis is meant to explore the various options for feature extraction for the OMG, amputees are not involved in initial test of the modality. The experiment was conducted on 10 subjects, each of whom agreed to the following conditions:

- The participation is voluntary
- They are not facing issues (e.g. injuries, wounds, pain, swell etc.) in moving their forearm and fingers
- They are not under any medication for the forearm and the hand
- They are not under the recovery period or have just recovered from an injury to the forearm and hand
- They are not allergic to the sticker, the VELCRO® band, the elastic band or the alcoholic disinfectant used to clean the elastic band used on every subject.
- The data is recorded anonymously and may be analysed and published along with the results of the analysis
- They may withdraw at any time during the experiment

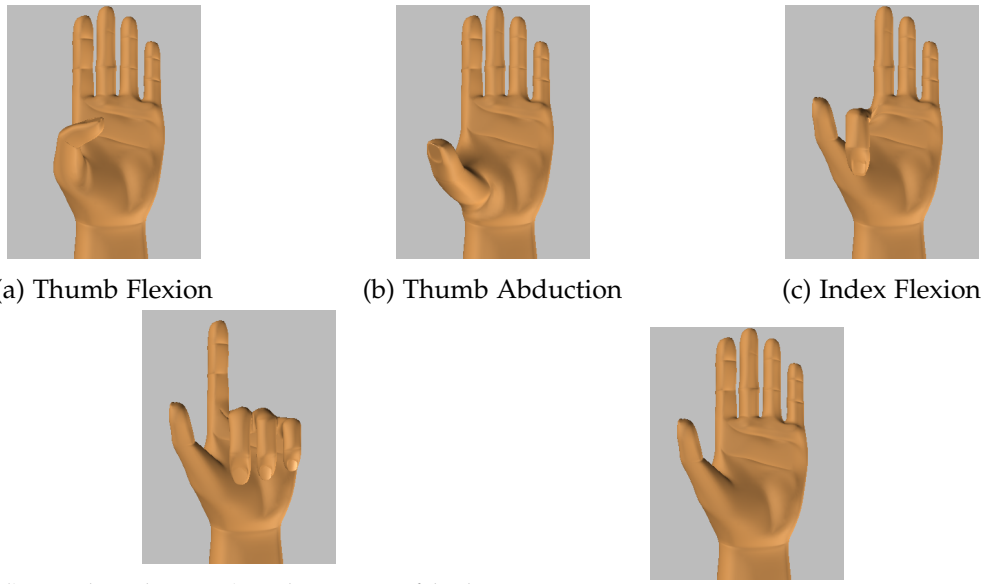
Each subject was described the procedure of the 10 minute experiment and was asked to use around 80% of their maximum force while following the stimuli. The camera was strapped on to the subject as if it were fixed to the base of an AHP as show in the figure below (Figure 5.1).

Figure 5.1: The experimental set-up where a subject is following the stimuli on the screen with a sticker stuck onto the left forearm and a camera strapped on to the same arm to capture its deformations

This is set-up on the left arm of all the subjects. There were two with the dominant hand being their left, while the others were right handed. The average circumference of the participants' left forearm measured 10 cm from the elbow is 23.58 ± 3.59. The average age of the participants (three female and 7 male) who volunteered as subjects is 26.2 ± 3.65. The sticker fastened to the hand was manually cut to fit within the forearm of each individual. Once prepared for the experiment, the subject was asked to place the forearm (freely) on the same set-up frame as that conducted previously using the tags. The subjects could have been asked to leave their arm move freely, although it can lead to a less uniform data acquisition. The subject was then allowed to see both the GUI used for recording and the 3-D hand model to be followed. The subject was given no other cues except a round about length of the experiment (around 10 minutes).

5.1.2 Data acquisition

The subjects were shown five poses for the finger of the left hand. The poses were Thumb Flexion, Thumb Abduction, Index Flexion, Combo Flexion and Rest (Figure 5.2). The thumb abduction can also be named thumb rotation for easy understanding as done in the previous OMG study [1], [2] and [3].



(a) Thumb Flexion

(b) Thumb Abduction

(c) Index Flexion

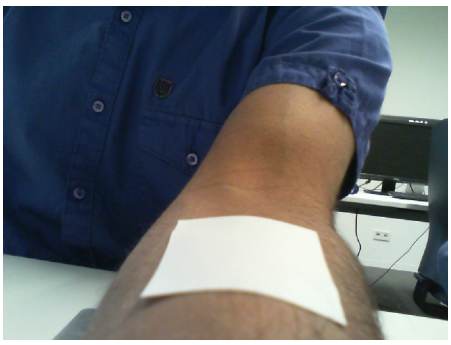
(d) Combo Flexion (combination of little, ring and middle fingers)

(e) Rest

Figure 5.2: The 3-D hand model used to stimulate the subject to follow the five displayed finger poses

Four of the poses (5.2a, 5.2b, 5.2c and 5.2d) had a time of 6 seconds to be followed and the Rest pose (5.2e) occurred in between each of them for 3 seconds. One repetition thus consisted of each of the four poses with the Rest pose acting as the intermediate pose between each of the four. Ten such repetitions were made throughout the experiment without a pause. The data was recorded off-line and was later processed.

For subjects whose sticker's boundaries touched others' when viewed in the image, morphological operations such as opening and closing were used to enable the *findContours* function of OpenCV to identify the sticker's contours as distinct. The sticker detected in the first frame is prompted for a manual check. The remaining frames are verified using Normalized Cross-Correlation (NCC) based on the previous frame's segmented sticker. The cropped and resized images were then saved in RGB Portable Network Graphics (PNG) format and then used for the training and testing off-line.



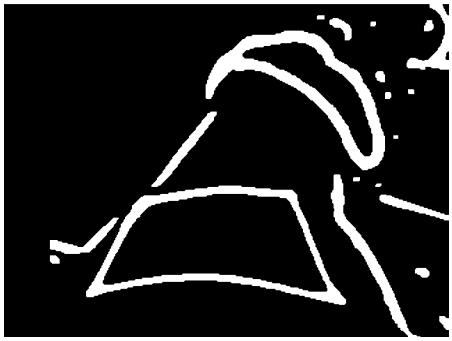
(a) Input image of subject whose sticker's boundary does not appear to touch any other's

(b) Input image of subject whose sticker's boundary appears to touch the arm's

(c) Extracted boundaries of the image

(d) Extracted boundaries of the image with the top-left corner of the sticker's boundary in contact with the forearm's

(e) The extracted boundaries do not undergo further processing



(f) The extracted boundaries after morphological operations where the top-left corner no longer touches the forearm's boundary

Figure 5.3: Morphological operations are not used (left column) unless the subject's sticker's boundary is in contact with its surrounding's (right column). Both the subjects' images displayed here are from the first frame

5.1.3 System training and classification

The saved cropped segmented images were resized (to 130x130 pixels in greyscale; for a batch size of 70 images) based on the GPU's memory capacity. The machine was asked to learn to classify all five poses in one session unlike when using the regressor where only one stimulus was shown to the machine for each training and testing session. The tests were run as a leave-one-repetition-out cross-validation, meaning that one among the ten repetitions will be used for testing while the other nine are sent in for training. The code was hence run ten times, each time taking a different repetition for testing and training freshly on the remaining nine. This provides an unbiased result for the tested classification. A confusion matrix is formed out of each of the test runs, leaving ten confusion matrices at the end of every subject's leave-one-repetition-out cross-validation test procedure.

(a) Thumb Flexion (b) Thumb Abduction (c) Index Flexion (d) Combo Flexion (e) Rest

Figure 5.4: Average of each of the segmented stimuli of one of the subjects sent in (after conversion to greyscale) during training

The above figure (Figure 5.4) shows the average of each of the segmented stimulus from a training batch. The images shown are in RGB but are fed into the machine in greyscale since colour doesn't play a major role in the classification. What does affect it is the positional change and deformational difference between the stimuli. Under the assumption that these slight positional changes are reproducible, as one can see, the CNN is fed these images with no further processing. When a stimulus's average are scattered, the increase in error is reflected in the predicted output.

6 Results

Results

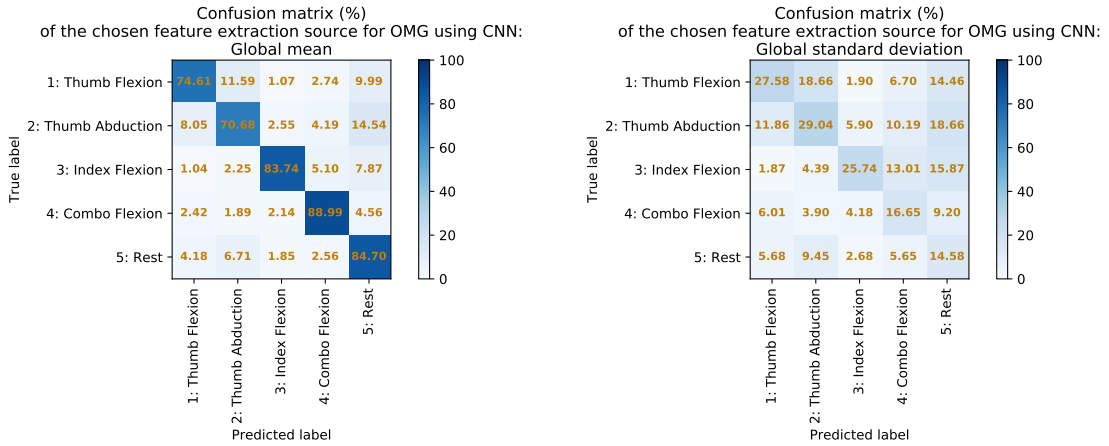
6.1 Computation of the results for the studied feature extraction sources

Although the chosen candidate used for feature extraction is a plain paper stuck to the forearm (Section 6.2), other candidates (Section 6.3) shall also be used as a comparison for the finger pose estimation. In the following two sections, a normalized confusion matrix will be used to infer the results of the candidates. The matrix is formed by taking each the subject's (or in Section 6.3, trial's) average of the mean, standard deviation, median and range of prediction results conducted tested over each repetition (taking a new repetition as test and training freshly over the remaining 9 repetitions of the five finger poses; 10 times). In other words, the intra-subject (using a leave-one-repetition-out cross-validation) mean, standard deviation, median and range was averaged over all subjects (inter-subject) to yield a global mean, standard deviation, median and range of all the subjects (or trials) respectively.

The normalization is computed once within each separate test of the CNN such that the absolute number of stimulus is not used but rather the relative amount of stimulus occurrences during the test. The mean and standard deviation cannot be taken as the best value over the repetitions since the overall (intra-subject) sample space is small (10 repetitions yielding 10 separate tests). Thus the median and range (maximum value minus the minimum over all 10 repetitions for each subject) depict an evaluation closer to that of the true performance of the model. This can be exemplified by taking the mean over one repetition where a subject fails to replicate the finger pose (and thus the deformation of the forearm) leading to a prediction accuracy of 0.10 while the remaining nine repetitions are quite accurate at 0.90; thus brining the mean down to 0.82 while the median (0.90) is robust against this one outlier (0.10) in the small sample space of 10 samples. The final results are displayed in percentage (after normalization).

6.2 Chosen feature extraction source

The results presented in this section are based on the experiments described in Chapter 5. The chosen source of input to the CNN for finger intent estimation is a plain white sticker stuck to the forearm with the camera strapped on to the same arm to record its deformation as the forearm does when moving the fingers. The global statistics are drawn out of those from each of the 10 participating subjects.

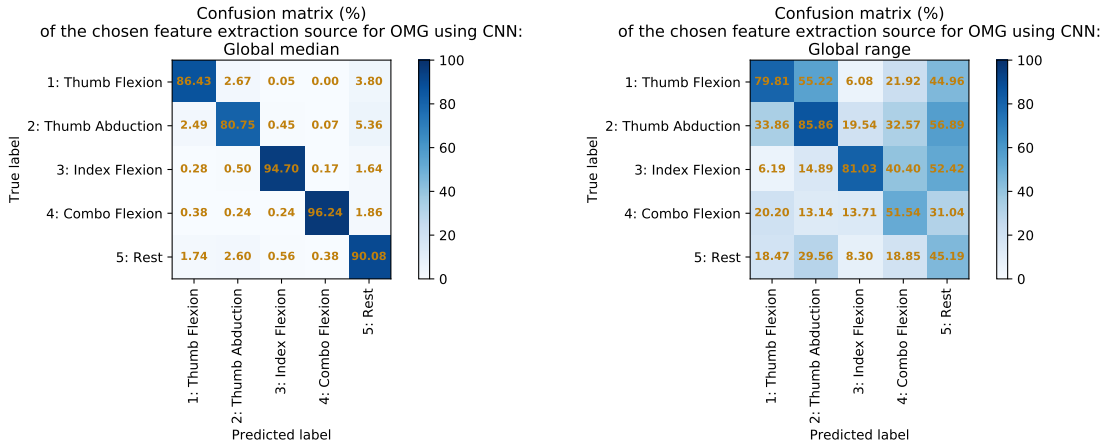


(a) Global mean when using a plain sticker on the forearm

(b) Global standard deviation when using a plain sticker on the forearm

Figure 6.1: Confusion matrix (in percentage) of the global mean and standard deviation of OMG using the chosen feature source computed by taking the inter-subject average of the intra-subject leave-one-repetition-out cross-validation's mean and standard deviation respectively

The distinction between the mean (Figure 6.1a) and median (Figure 6.2a) is obvious since some of the predictions in one or two repetitions were not predicted correctly although the performance in the remaining repetitions was quite good. Both the thumb poses however are not so well predicted as compared to the combo and index finger poses. The thumbs seem to get miss-classified either between each other or as rest. This can be due to the fact the muscles which move the thumbs are deep-muscles and its deformation is not as visible on the surface as with other fingers'. The standard deviations (Figure 6.1b) are low hence concluding a stable system.



(a) Global median when using a plain sticker on the forearm

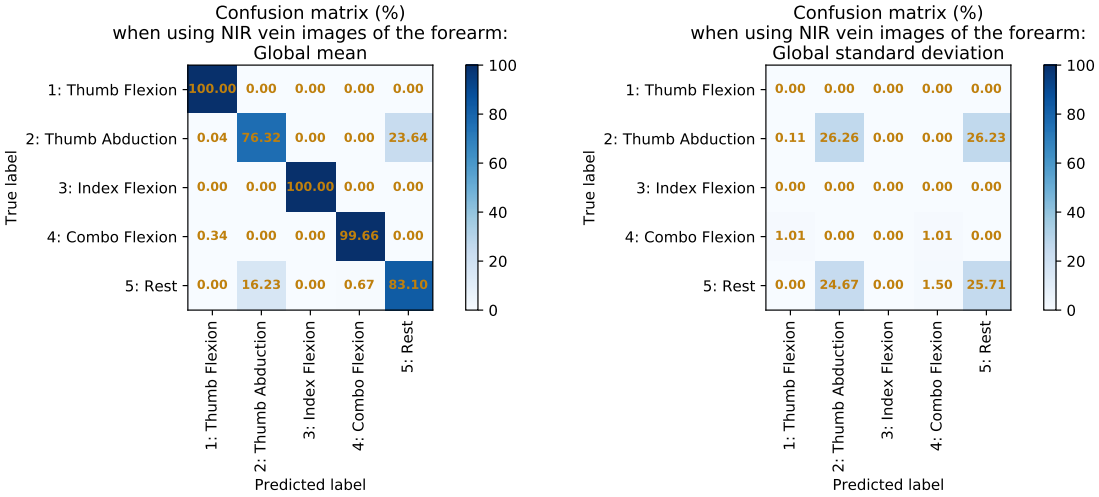
(b) Global range when using a plain sticker on the forearm

Figure 6.2: Confusion matrix (in percentage) of the global median and range of OMG using the feature source computed by taking the inter-subject average of the intra-subject leave-one-repetition-out cross-validation's median and range respectively

6.3 Other feature extraction candidates studied for the CNN

Since the thesis is intended to find a reliable way to detect features, the results of the other methods experimented shall also be discussed in this section. However, since the plain sticker's advantages outweigh the remaining candidate sources of feature extraction, the experiments carried out in this section are carried out in two separate sessions or trials. The global statistics are thus drawn from these two trials (except in Section 6.3.2 where only one trial's result is displayed). The following cases in this section are carried out by a subject who is considered to be used to the experiments and the results may thus seem better than on the 10 subjects in the final experiment who were not as used to the experimental workflow.

6.3.1 Natural features: NIR vein images

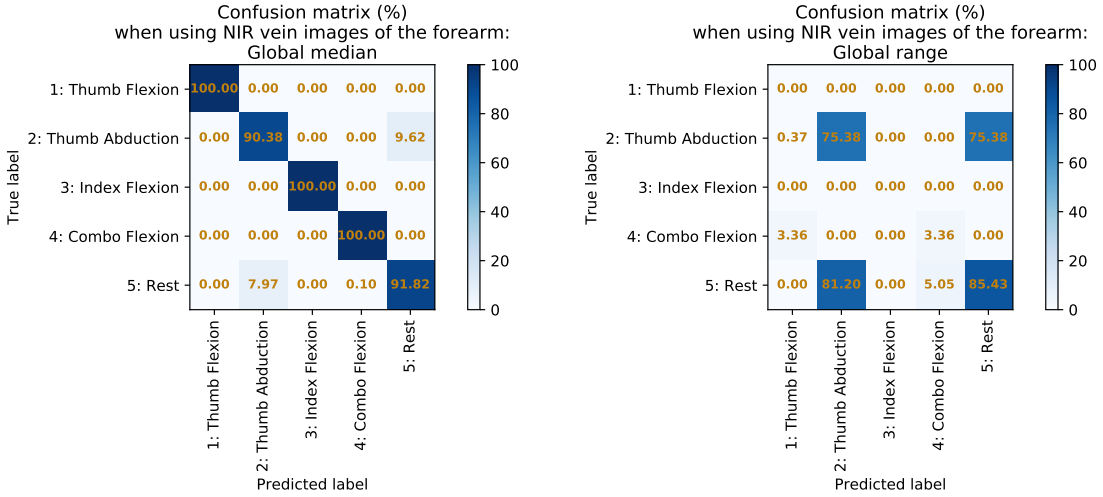


(a) Global mean when using NIR vein images of the forearm

(b) Global standard deviation when using NIR vein images of the forearm

Figure 6.3: Confusion matrix (in percentage) of the global mean and standard deviation of OMG using NIR vein images of the forearm computed by taking the inter-subject average of the intra-subject leave-one-repetition-out cross-validation's mean and standard deviation respectively

The Near-infrared (NIR) images of the veins performed really well (Figure 6.4a). The thumb abduction and rest are still slightly confused for one another. The reason for not proceeding with NIR images mainly stems from the drawbacks of depending on the NIR rays to be consistently as good as the visible illumination such as glares or the thickness of adipose tissue in various subjects [59] [60]. The other reason is the use of physical markers. These two reasons (as elaborated in Sections 3.2.1 and 3.2.3) can be overcome by using a better NIR camera (which can counter one of the objectives of searching for a cost-effective solution) and by extracting the veins from a fixed location on the forearm. The latter requires extracting landmarks from the forearm which are reliable and does not change over time nor hamper the rate of feeding the trained CNN for a real-time application.



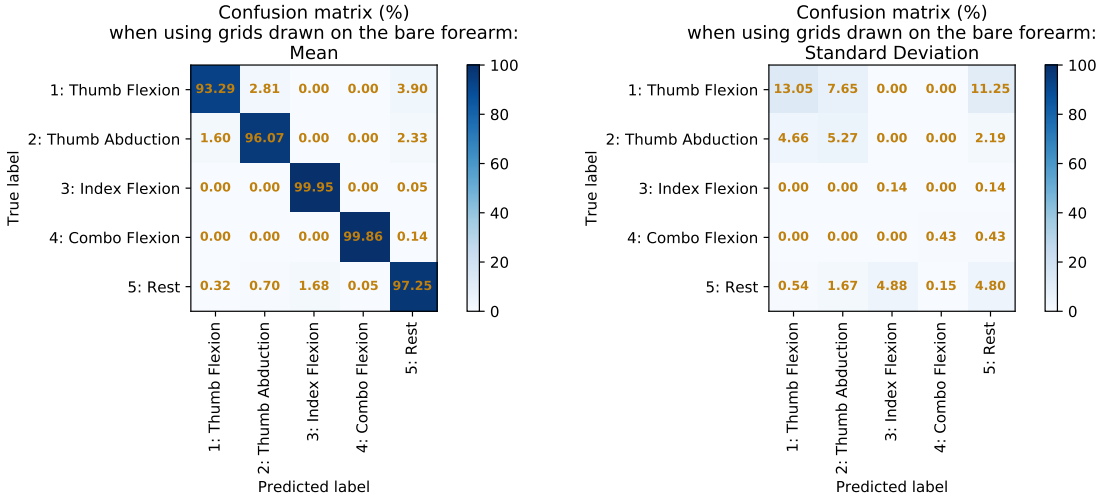
(a) Global median when using a NIR vein images of the forearm

(b) Global range when using a NIR vein images of the forearm

Figure 6.4: Confusion matrix (in percentage) of the global median and range of OMG using NIR vein images of the forearm computed by taking the inter-subject average of the intra-subject leave-one-repetition-out cross-validation's median and range respectively

6.3.2 Artificial fiducials: Hand drawn grids on bare forearm

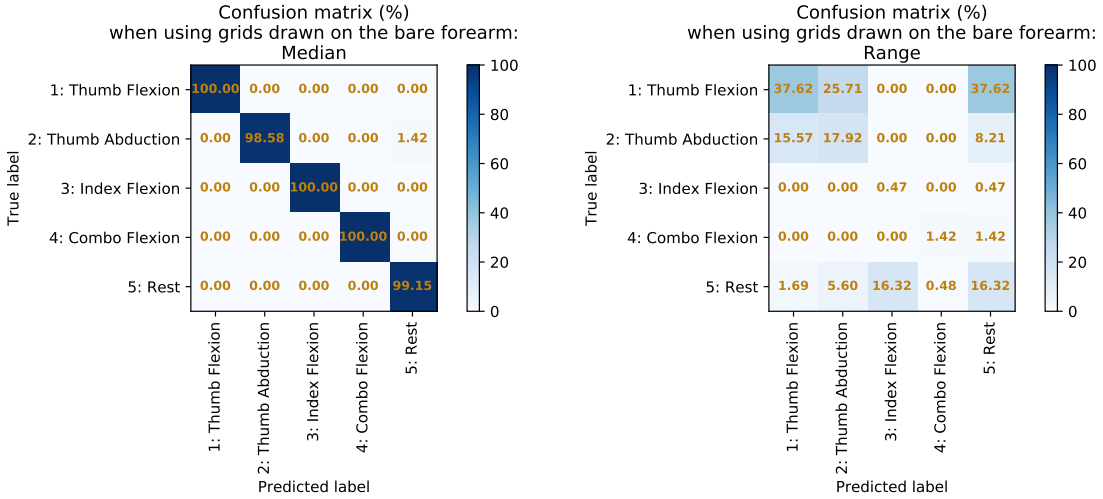
Hand-drawn grids on the bare forearm perform better than using the comparatively weak natural features of the veins obtained from a low-cost NIR camera. If a biocompatible ink or marker can be used without its disintegration over time or the avoidance of physical markers as landmarks to extract the ROI can be achieved (as reasoned in Section 3.2.3), this method can be one of the reliable sources of feature extraction.



(a) Mean when using hand-drawn grids on the bare forearm

(b) Standard deviation when using hand-drawn grids on the bare forearm

Figure 6.5: Confusion matrix (in percentage) of the mean and standard deviation of OMG using hand-drawn grids on the bare forearm computed by taking the intra-subject leave-one-repetition-out cross-validation's mean and standard deviation respectively

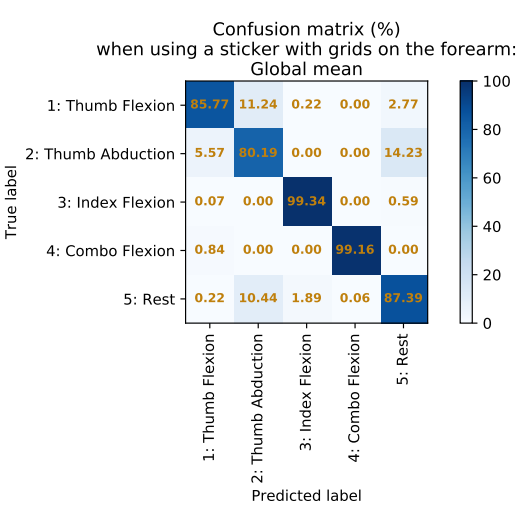


(a) Median when using hand-drawn grids on the bare forearm

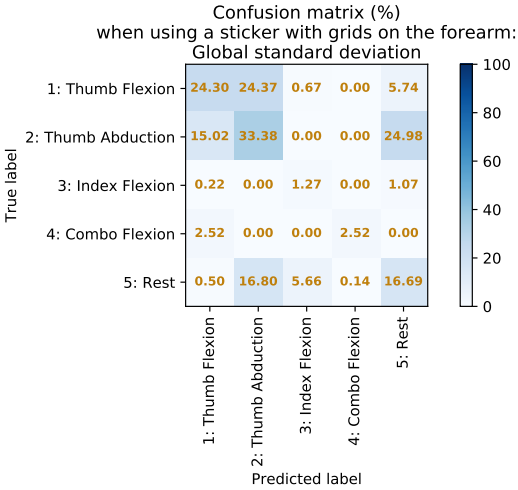
(b) Range when using hand-drawn grids on the bare forearm

Figure 6.6: Confusion matrix (in percentage) of the median and range of OMG using hand-drawn grids on the bare forearm computed by taking the intra-subject leave-one-repetition-out cross-validation's median and range respectively

6.3.3 Artificial fiducials: Hand drawn grids on sticker

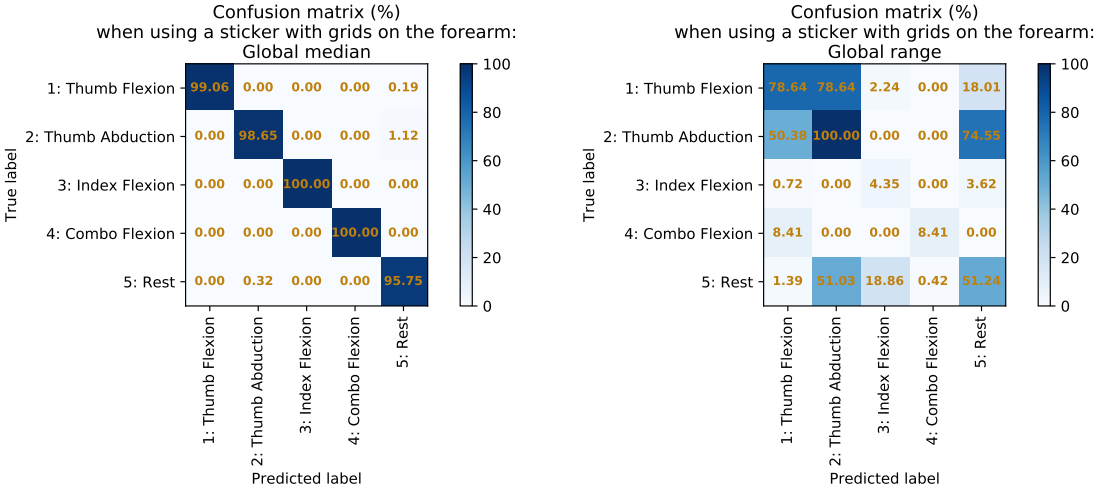


(a) Global mean when using a sticker with hand-drawn grids stuck to the forearm



(b) Global standard deviation when using a sticker with hand-drawn grids stuck to the forearm

Figure 6.7: Confusion matrix (in percentage) of the global mean and standard deviation of OMG using a sticker with hand-drawn grids stuck to the forearm computed by taking the inter-subject average of the intra-subject leave-one-repetition-out cross-validation's mean and standard deviation respectively



(a) Global median when using a sticker with hand-drawn grids stuck to the forearm

(b) Global range when using a sticker with hand-drawn grids stuck to the forearm

Figure 6.8: Confusion matrix (in percentage) of the global median and range of OMG using a sticker with hand-drawn grids stuck to the forearm computed by taking the inter-subject average of the intra-subject leave-one-repetition-out cross-validation's median and range respectively

The thumb poses and rest are usually confused when using this candidate source as can be inferred from 6.7b and 6.8b. The reasons described in Section 3.2.3 support the move from the hand-drawn grids on a sticker to using a plain sticker as the chosen candidate for the final experiments.

6.4 Comparison to previous approaches

Since this thesis mainly focusses on classification of finger poses, the results can be fairly compared only to other methods that worked on classification. The results can thus not be compared to the regression used in the proof of concept of the Optical Myography using AprilTags. The results have also been adapted to mimic as close as possible to the finger poses studied in this thesis, as shall be described during the introduction to the modalities' studies.

G. Naik et al. proposed two sEMG configurations in [61] which they considered the most optimal. This was done using a model based approach using independent component analysis (ICA) and Icasso clustering. The tests were carried out on five transradial amputees with 11 finger poses of around 5 to 7 repetitions using a leave-

one-repetition-out cross-validation. Since the experiment consisted of neither a rest pose nor a combo flexion, the two shall be transformed by taking the average of the little, ring, middle, pointer and thumb extension, and the little, ring and middle finger flexion as the rest pose and combo flexion respectively. The remaining poses were the same as in this thesis.

Four transradially amputated subjects participated in the study on FMG conducted by E. Cho et al. in [24]. Each subjected performed five trials with 11 different grip gestures; in which four were used for training and the remaining trial for testing using an inter-trial cross-validation approach to produce the average result of the five different tests. The grips were classified using linear discriminant analysis (LDA). Since the 11 grip patterns weren't as close to the ones executed in this thesis, the key grip performed in their study was assumed similar to the force applied by the thumb (under the assumption that this force dominates those by the remaining fingers) during the thumb flexion in OMG's. The mouse grip was almost the same as the thumb abduction, the precision open is assumed similar to the index flexion and the finger point and relaxed hand were the same as the combo flexion and rest position respectively. The final confusion matrix drawn from the four subjects' was that of their median as used by the OMG method.

Using a KNN-Classifier, S. Sikdar et al. used B-mode ultrasound images (as described in Section 2.2) to classify individual finger poses to evaluate SMG [25] using a leave-one-repetition-out cross-validation. The combo flexion was derived by taking the average of the performance by the little, ring and middle finger flexion, while the remaining were the same as in this thesis. The rest and the thumb abduction could not be gathered for the comparison (marked as N/A or not available in the comparison plots) since their study did not involve any similar pose.

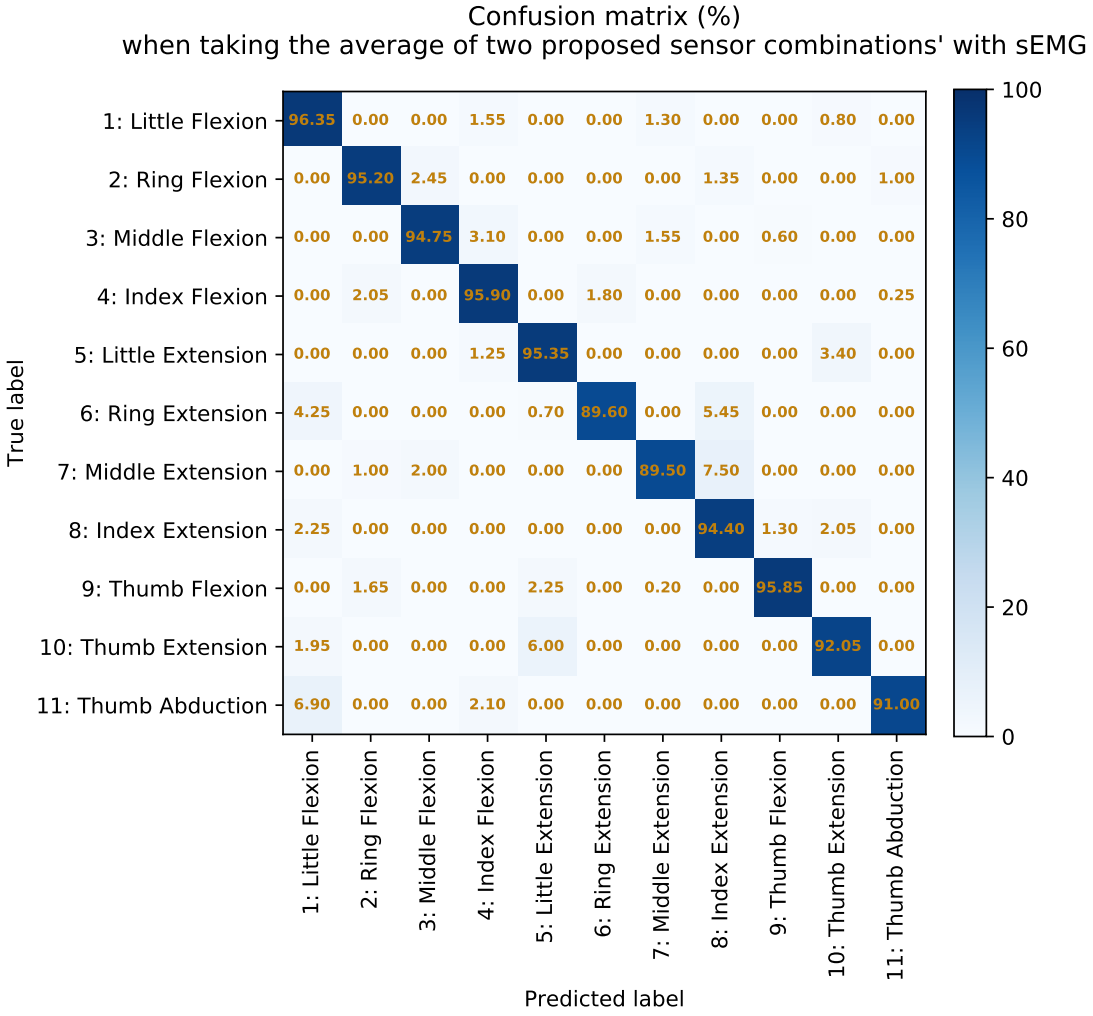


Figure 6.9: sEMG confusion matrix (in percentage) adapted from [61] by taking the average of the two proposed sensor combinations'

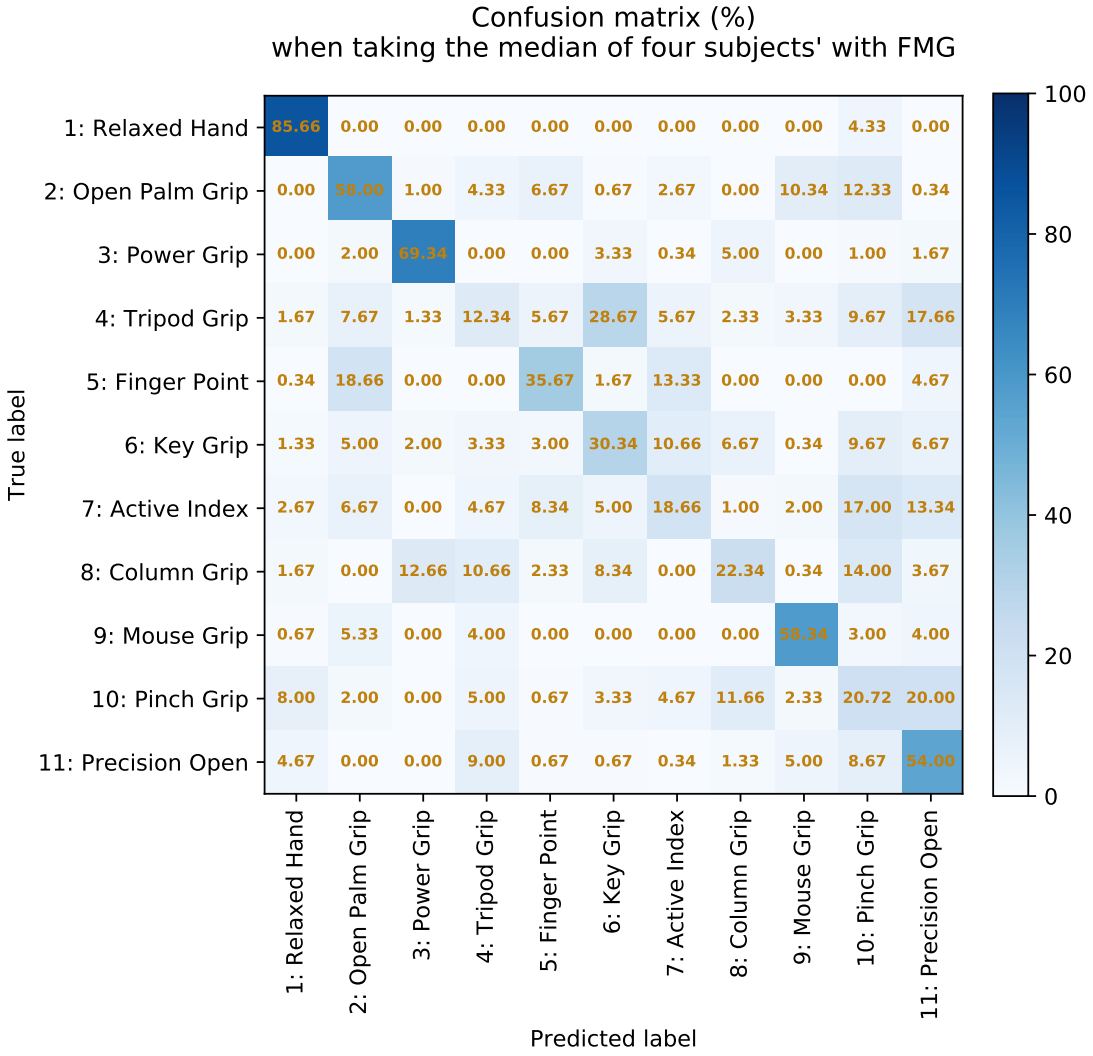


Figure 6.10: FMG confusion matrix (in percentage) adapted from [24] by taking the median of the four subjects'

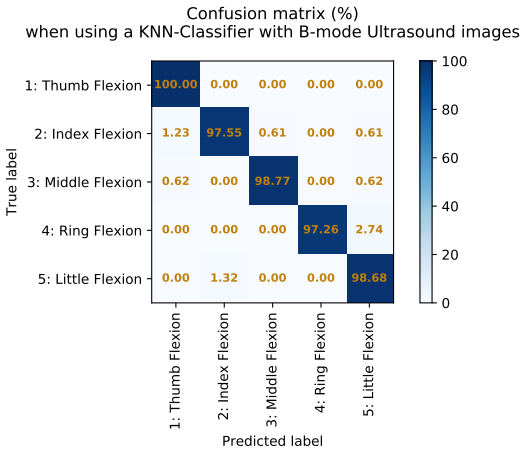


Figure 6.11: SMG confusion matrix (in percentage) adapted from [25]

In order to compare the modalities together, their accuracy, precision, sensitivity and specificity are calculated based on their final computed confusion matrix (in percentage).

The formulae used to compute these measures are:

$$Accuracy = \frac{TPs + TNs}{TPs + TNs + FNs + FPs} * 100 \tag{6.1}$$

$$Precision = \frac{TPs}{TPs + FPs} * 100 \tag{6.2}$$

$$Sensitivity = \frac{TPs}{TPs + FNs} * 100 \tag{6.3}$$

$$Specificity = \frac{TNs}{TNs + FPs} * 100 \tag{6.4}$$

The following plots (Figures 6.12 to 6.19) are plotted using the confusion matrices of the reference modalities (Figures 6.9, 6.10 and 6.11 referred to as sEMG, FMG and SMG respectively) and the chosen feature extraction source's (Figure 6.2a referred to as OMG_CNN) of OMG using CNN. From the modality-wise plots below (Figures 6.12 to 6.15), one can see that OMG using CNN is almost as accurate as the standard sEMG. The overall precision is not as high mainly due to the confusion between the rest, and the weak deformations tracked during the two thumb poses. All the methods portray high specificity meaning that the false-positives are low. The sensitivity is quite good compared to the other modalities. The pose-wise comparisons (Figures 6.16 to 6.19) reveal the accuracy of all the detected poses to be just as good as the preceding modalities. The high reliability (Figure 6.17) can be maintained during a more practical use if the camera is fixed sturdily, for instance, to the base of an AHP .

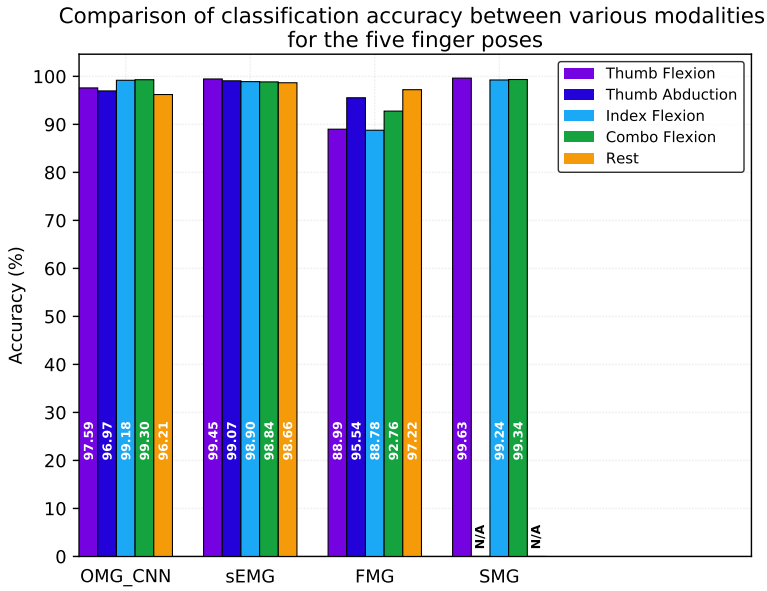


Figure 6.12: Modality-wise comparison of accuracy (in percentage) between the various finger poses

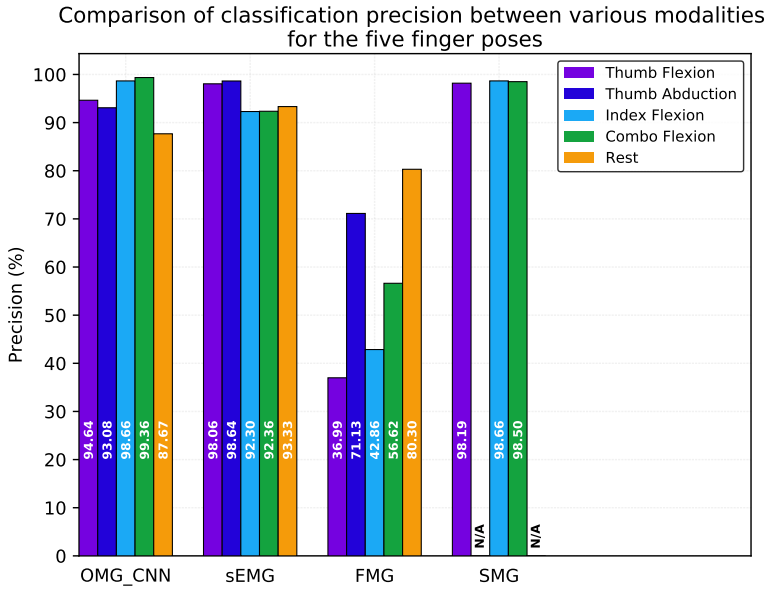


Figure 6.13: Modality-wise comparison of precision (in percentage) between the various finger poses

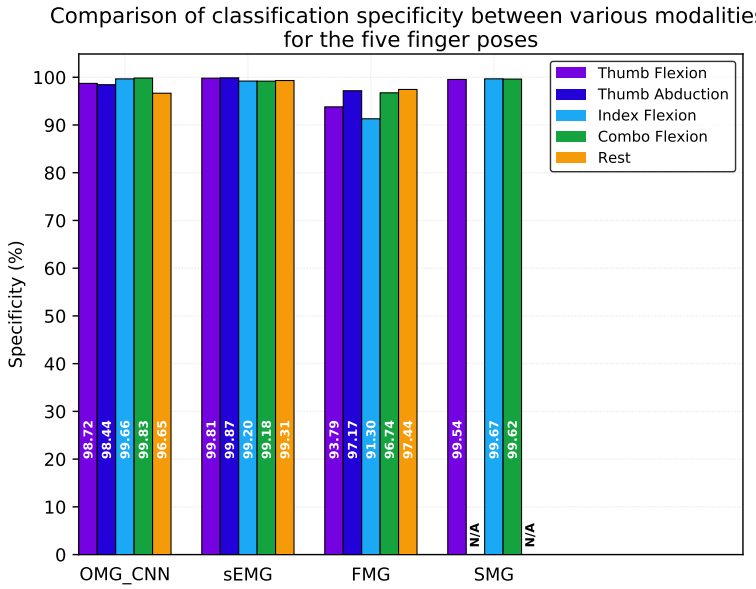


Figure 6.14: Modality-wise comparison of specificity (in percentage) between the various finger poses

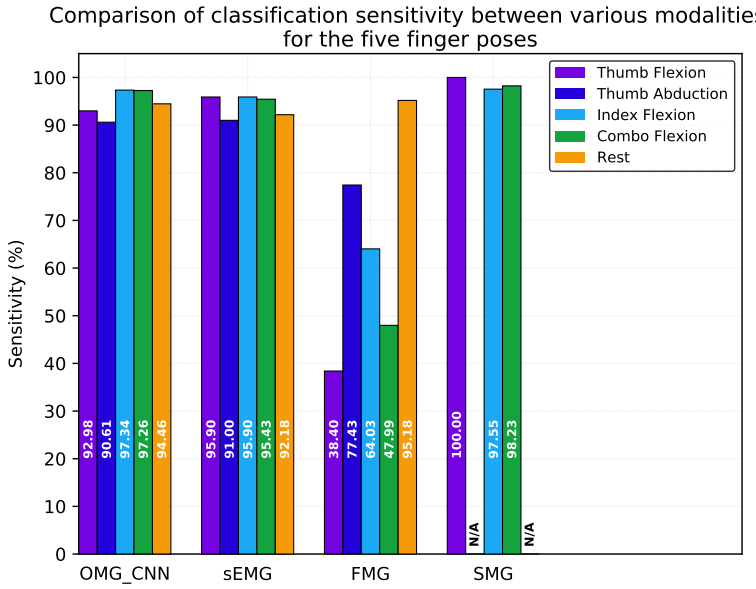


Figure 6.15: Modality-wise comparison of sensitivity (in percentage) between the various finger poses

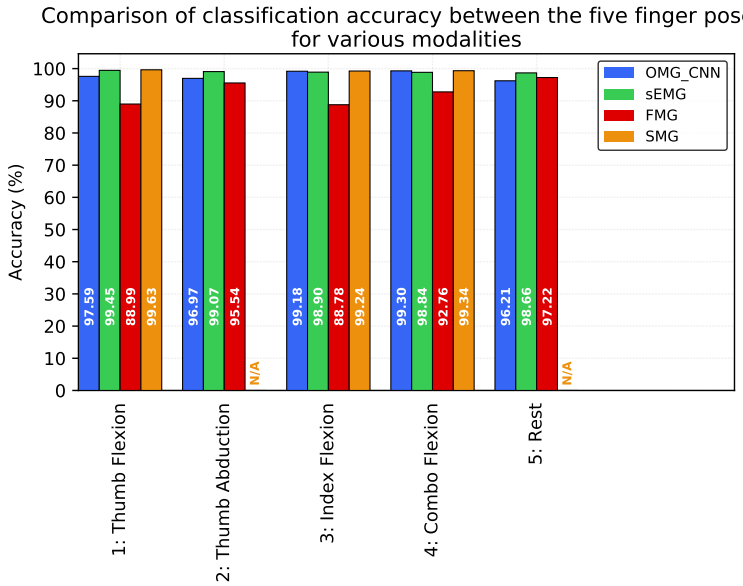


Figure 6.16: Pose-wise comparison of accuracy (in percentage) between the various modalities

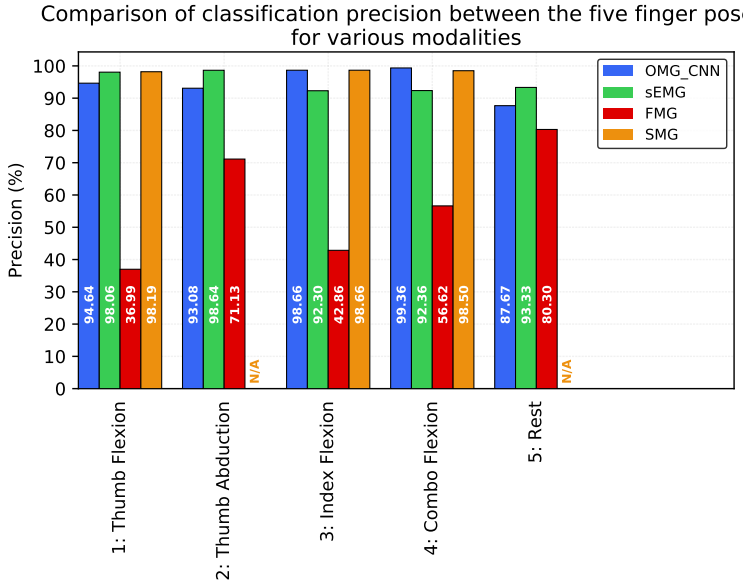


Figure 6.17: Pose-wise comparison of precision (in percentage) between the various modalities

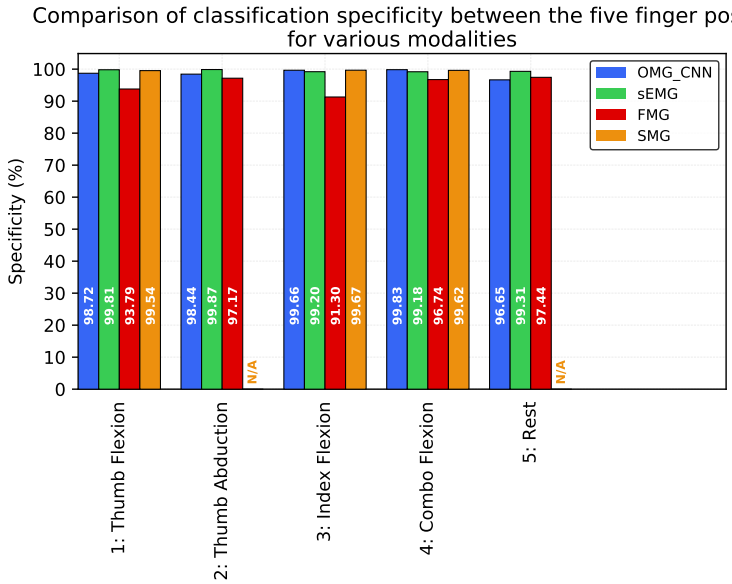


Figure 6.18: Pose-wise comparison of specificity (in percentage) between the various modalities

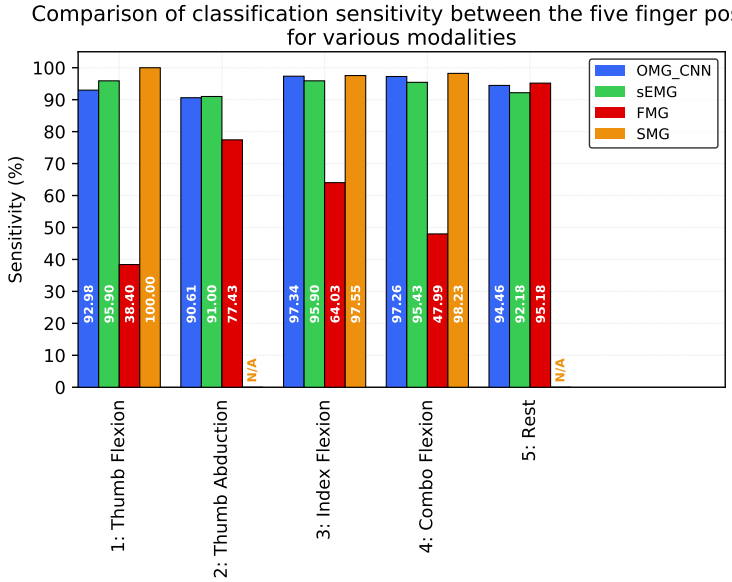


Figure 6.19: Pose-wise comparison of sensitivity (in percentage) between the various modalities

7 Conclusion and Discussion

Conclusion and Discussion

7.1 Conclusion

The search for a feature extraction source boiled down to placing a simple sticker on the forearm, which (using a prototype set-up) performed as good as other modalities using classification. Occlusion of the sticker is the Achilles heel of the system. The weak deformations during the thumb poses and its confusion with the rest pose still posed a challenge compared to the other poses. Although the other sources did perform on par with each other, they each have their own drawbacks. The thesis is also aimed to find the limitations of using normal images as input for a CNN which is trained to classify finger poses of an individual subject as opposed to using augmented reality markers along with the more quickly trainable ridge-regression.

Augmented reality markers yield high accuracy in detecting deformations on the forearm. However, this strength also turned out to be its weakness when the tags were made small enough to be placed on the forearm. The jitter caused by this uncertainty in precision even after each tag is glued to a point using a thick flat paper can cause the machine to learn on noisy data hence leading to a less reliable output. This leads to using the image itself as an input to the machine learner. Since it is complicated to train a ridge regressor on a rather unprocessed image, a CNN is chosen. Due to the web-camera's fixed focus and its resulting field of view, not just the forearm, but its surrounding environment also get treated as input to the CNN. To make the approach more practically viable by obviating a very large input dataset, an ROI in which only the features of the forearm are present is explored in this thesis.

Using NIR images of the veins on the forearm performs quite well. However, the four corners of the segmented ROI was extracted by computing the centre of a circle printed on four physical markers placed on the forearm. To increase the accuracy of the centres, the markers were stuck using a thick flat paper thus decreasing inaccuracies caused by the 3-D curvature when tracked on a 2-D image. A pause in the recording of the frames during the finger movements however can cause the estimated centre of the circle to jump outside the circle. The chances of the circles appearing too affine remain in both a flat and a marker stuck to fit the curvature of the spot on the forearm. Blob detections demand a high true positive rate for the four markers thus withdrawing itself as an alternative. Another drawback of using NIR images is the overpowering of the remaining electromagnetic spectrum reflected by the skin over the NIR rays absorbed by the veins. A controlled source of NIR light as in [37] and [38] was not an option in order to reduce the amount of hardware for a practical scenario and also to avoid the possible pattern of the NIR light array reflected from the skin causing the machine to not learn the necessary features over a shorter period of time.

The same physical markers were also used when hand-drawn grids on the forearm were used as a source of feature extraction. The drawbacks still did exist, but using hand-drawn grids paved the path to testing the remaining two sources of feature-extraction. By drawing grids directly on the surface of the forearm, one can detect the deformations on the surface of the forearm more easily. Using a biocompatible ink which does not smudge over time can be a solution to reverting to this fiducial system. Extracting only the grids on the surface of the forearm without using the physical markers poses a challenge due to changes in illumination causing the true external boundaries of the grids to be diminished by glares on the curved surface of the forearm.

This lead to the use of a sticker stuck to the forearm with grids drawn on top of it. The performance was quite close to the extracting features directly from the forearm. The image processing now required a search for the boundaries sticker rather than relying on accuracy and precision in the detection of four image corners. The ROI was hence the bounding box of the sticker's boundaries. The detection of these contours was challenged by the contours of the grids within it during image processing. Using filter sizes larger than the thickness of the grids to blur out the effects of the grids being miss-detected as the boundaries of the sticker resulted in loosing the exact contours of the sticker due to less sharper edge gradients.

Since the images are recorded in 2 dimensions, each point (intersection of two grid lines) within the sticker can be linked to corresponding points along its edges. The initial experiments conducted using just a plain sticker, without the grids drawn within, performed as good as those with the grids. The inference drawn was that the network was probably laying too much emphasis on the grids within or was getting confused on choosing the right features with the grids which acted as additional and perhaps redundant feature sources. The plain sticker also simplified and accelerated the process of segmenting the contours of the sticker whose bounding box was used as the ROI. The final experiments with the intact subject was then conducted using this method, one that was reliable and suited for a real-time application.

The CNN used performed better without max-pooling layers especially in the case of NIR images. The main reason for this is the subtle positional changes of the deformation pattern which can be lost especially when the input images are fed in at low resolutions (of 130x130 pixels). A minimum error was attained within 20 epochs using 70 images per batch when using the plain white stickers since the only probable features to be tracked were the changing shape of the edges of the sticker and the slight positional change. The network also assumes that displacements in the camera angle and position caused by the deformation at the distal end of the forearm where the camera is strapped is replicable. The performance of the prototype was however as good as the other modalities which also used classification as the basis of finger pose estimation.

7.2 Discussion

As far as real-time execution of the image processing for Optical Myography (OMG) goes, a simple web-camera strapped on to the distal side of the forearm recording the deformations of the forearm with the help of a sticker stuck to it can help in finger pose estimation using Convolutional Neural Networks (CNN) with quite good accuracy. The image segmentation alone took about 23 FPS (including writing each segmented image of 256x256 pixels) on a CPU (mentioned in Section 4.1.1); the speed can increase when sent directly to the prediction phase (with image sizes of 130x130 pixels) after the model has been trained). The testing (prediction) phase can take about 1 second to compute 70 frames' predictions on the same CPU; this can be increased by changing the batch size used for testing to 1 image per batch or on a GPU. The model was also tested aside from the main goals of the thesis with the forearm moving about the room at varying heights, angles and positions (keeping the sticker in view of the camera) proving its capabilities of a free-arm system when tested off-line. The only challenge faced here was the web-camera's response time to sudden changes in illumination especially when using the ceiling's artificial LED lighting. The system was quite stable against relatively slow illumination changes, even when the source of light was behind the forearm.

Although motion blur was suppressed to its maximum, the camera was still susceptible to movement due to its design. This pronounced quite clearly when a trial real-time run (during testing phase without a GPU) was conducted around 90 minutes after recording the data for training; after which a slight change in the camera angle caused the predictions to start wavering. However, this issue can be overcome if the camera is mounted on top of or is inbuilt to the fixed base of the AHP. The motion blur can also be avoided to a good extent this way.

Most web-cameras are designed such that the camera can be placed on a monitor and is focused on a person's face at an arm's distance from it. This lead to the web-camera (and a few other tested web-cameras) to have its focus on the upper arm (on the biceps) rather than the forearm, thus reducing the resolution of the sticker's boundaries.

Due to the accumulation of sweat, the current sticker's adhesiveness can wear off. However, using materials such as plasters used to heal wounds, but large enough to cover the required area of the forearm for deformation detection, can prove to be more adhesive, biocompatible and water resistant.

Modalities using natural features such as the NIR vein imaging has also the potent for finger pose estimation. However, a good infra-red camera which does not increase the cost of the system too much is yet to hit the market. Even if it does, a reliable and computationally efficient way of extracting the ROI should be discovered.

A regression based approach of the CNN rather than a classification, or a Convolu-

tional Long-Short Term Memory (ConvLSTM) can also be used to obtain a smoother transition of the estimation. These however rely heavily on the discernibility of the various finger poses during the transitional phase. This can tend to become more imprecise if the variations are small during transition and the number of poses increase. However, such a study can reveal its capabilities.

Since the posterior side of the forearm also has muscles pertaining to wrist movements such as extension, a second camera on the posterior side or even acting as a stereo-assist for depth imaging of the deformations can help in increasing the accuracy and the number of the estimated hand poses. Perhaps an OMG which does not use a web-camera but rather optical sources such as reflecting patterns of light or a band of NIR emitting and sensing arrays placed over the forearm to detect just the veins as input to the CNN can also surface as a source for OMG using CNN since this thesis has proved that vein deformations induced by the forearm can be used to estimate its corresponding finger locations.

List of Figures

Bibliography

[1] N. Mouriki, C. Castellini, C. Nissler, N. Navab and V. Belagiannis. 'Hand motion estimation using image based reconstruction of the forearm'. In: *Master's Thesis in Informatics: Biomedical Computing* (2015).

[2] C. Nissler, N. Mouriki and C. Castellini. 'Optical Myography: Detecting Finger Movements by Looking at the Forearm'. In: *Frontiers in Neurorobotics* 10 (2016), p. 3. ISSN: 1662-5218. DOI: 10.3389/fnbot.2016.00003.

[3] C. Nissler, N. Mouriki, C. Castellini, V. Belagiannis and N. Navab. 'OMG: Introducing optical myography as a new human machine interface for hand amputees'. In: *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. 2015, pp. 937–942. DOI: 10.1109/ICORR.2015.7281324.

[4] S. Micera, J. Carpaneto and S. Raspopovic. 'Control of hand prostheses using peripheral information'. In: *IEEE Reviews in Biomedical Engineering* 3 (2010), pp. 48–68.

[5] B. Peerdeman, D. Boere, H. Witteveen, H. Hermens, S. Stramigioli, J. Rietman, P. Veltink, S. Misra et al. 'Myoelectric forearm prostheses: State of the art from a user-centered perspective'. In: (2011).

[6] A. Fougner, Ø. Stavdahl, P. J. Kyberd, Y. G. Losier and P. A. Parker. 'Control of upper limb prostheses: terminology and proportional myoelectric control—a review'. In: *IEEE Transactions on neural systems and rehabilitation engineering* 20.5 (2012), pp. 663–677.

[7] T. Bionics. *touchbionics*. http://www.touchbionics.de/. [Online; accessed 10-April-2016]. 2016.

[8] ottobock. *ottobock*. http://www.ottobock.de/. [Online; accessed 10-April-2016]. 2016.

[9] utaharm. *utaharm*. http://www.utaharm.com/. [Online; accessed 10-April-2016]. 2016.

[10] steeper. *steeper*. http://rslsteeper.com/. [Online; accessed 10-April-2016]. 2016.

[11] T. Bionics. *touchbionics i-limb ultra*. http://touchbionics.com/resources/images/i-limb-ultra-images. [Online; accessed 17-November-2016]. 2016.

[12] Bebionic. *bebionc*. http://www.cyberpunkworld.com/terminator-this-bebionic-prosthetic-hand-is-pretty-sweet/bebionic-black-glove/. [Online; accessed 17-November-2016]. 2016.

[13] T. D'Alessio and S. Conforto. 'Extraction of the envelope from surface EMG signals'. In: *IEEE Engineering in Medicine and Biology Magazine* 20.6 (2001), pp. 55–61. ISSN: 0739-5175. DOI: 10.1109/51.982276.

[14] J.-Y. Guo, Y.-P. Zheng, L. P. Kenney, A. Bowen, D. Howard and J. J. Canderle. 'A comparative evaluation of sonomyography, electromyography, force, and wrist angle in a discrete tracking task'. In: *Ultrasound in medicine & biology* 37.6 (2011), pp. 884–891.

[15] C. Castellini, G. Passig and E. Zarka. 'Using ultrasound images of the forearm to predict finger positions'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 20.6 (2012), pp. 788–797.

[16] D. Sierra González and C. Castellini. 'A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees'. In: *Frontiers in neurorobotics* 7 (2013), p. 17.

[17] Y.-P. Zheng, M. Chan, J. Shi, X. Chen and Q.-H. Huang. 'Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis'. In: *Medical engineering & physics* 28.5 (2006), pp. 405–415.

[18] Z. G. Xiao and C. Menon. 'Towards the development of a wearable feedback system for monitoring the activities of the upper-extremities'. In: *Journal of neuroengineering and rehabilitation* 11.1 (2014), p. 1.

[19] D. Yungher and W. Craelius. 'Improving fine motor function after brain injury using gesture recognition biofeedback'. In: *Disability and Rehabilitation: Assistive Technology* 7.6 (2012), pp. 464–468.

[20] N. Celadon, S. Dosen, M. Paleari, D. Farina and P. Ariano. 'Individual finger classification from surface EMG: Influence of electrode set'. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 7284–7287.

[21] T. Hiyama, S. Sakurazawa, M. Toda, J. Akita, K. Kondo and Y. Nakamura. 'Motion estimation of five fingers using small concentric ring electrodes for measuring surface electromyography'. In: *2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE)*. IEEE. 2014, pp. 376–380.

[22] A. Gijsberts, R. Bohra, D. Sierra González, A. Werner, M. Nowak, B. Caputo, M. A. Roa and C. Castellini. 'Stable myoelectric control of a hand prosthesis using non-linear incremental learning'. In: *Frontiers in neurorobotics* 8 (2014), p. 8.

[23] M. Wininger, N.-H. Kim and W. Craelius. 'Pressure signature of forearm as predictor of grip force'. In: *Journal of rehabilitation research and development* 45.6 (2008), p. 883.

[24] E. Cho, R. Chen, L.-K. Merhi, Z. Xiao, B. Pousett and C. Menon. 'Force myography to control robotic upper extremity prostheses: a feasibility study'. In: *Frontiers in bioengineering and biotechnology* 4 (2016).

[25] S. Sikdar, H. Rangwala, E. B. Eastlake, I. A. Hunt, A. J. Nelson, J. Devanathan, A. Shin and J. J. Pancrazio. 'Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.1 (2014), pp. 69–76.

[26] N. Chen, S. Urban, C. Osendorfer, J. Bayer and P. Van Der Smagt. 'Estimating finger grip force from an image of the hand using Convolutional Neural Networks and Gaussian Processes'. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2014, pp. 3137–3142.

[27] april robotics laboratory. *AprilTags*. `https://april.eecs.umich.edu/wiki/AprilTags`. [Online; accessed 9-November-2016]. 2016.

[28] U. de Córdoba. *ArUco Markers*. `https://www.uco.es/investiga/grupos/ava/node/26`. [Online; accessed 9-November-2016]. 2016.

[29] TeachMe. *TeachMe Anterior Upper-Limb*. `http://teachmeanatomy.info/upper-limb/muscles/anterior-forearm/`. [Online; accessed 6-November-2016]. 2016.

[30] innerbody. *innerbody Muscular Arm-Hand*. `http://www.innerbody.com/anatomy/muscular/arm-hand`. [Online; accessed 6-November-2016]. 2016.

[31] innerbody. *innerbody Muscular Hand and Wrist*. `http://www.innerbody.com/image_skel13/ligm27.html`. [Online; accessed 6-November-2016]. 2016.

[32] healthline. *healthline Hand*. `http://www.healthline.com/human-body-maps/hand`. [Online; accessed 7-November-2016]. 2016.

[33] xraymachines. *xraymachines Thumb Joints*. `http://www.xraymachines.info/article/402829873/the-joints-of-the-thumb-part-i/`. [Online; accessed 7-November-2016]. 2016.

[34] dartmouth. *dartmouth Anterior Forearm*. `https://www.dartmouth.edu/~humananatomy/part_2/chapter_10.html`. [Online; accessed 6-November-2016]. 2016.

[35] Antranik. *Anterior compartment of the forearm*. http://antranik.org/wp-content/uploads/2014/04/anterior-compartment-of-the-forearm.jpg. [Online; accessed 20-November-2016]. 2016.

[36] V. P. Zharov, S. Ferguson, J. F. Eidt, P. C. Howard, L. M. Fink and M. Waner. 'Infrared imaging of subcutaneous veins'. In: *Lasers in Surgery and Medicine* 34.1 (2004), pp. 56–61.

[37] M. Mansoor, S. N. Sravani, S. Z. Naqvi, I. Badshah and M. Saleem. 'Real-time low cost infrared vein imaging system'. In: *Signal Processing Image Processing Pattern Recognition (ICSIPR), 2013 International Conference on*. IEEE. 2013, pp. 117–121. DOI: 10.1109/ICSIPR.2013.6497970.

[38] S. N. Sravani, S. Z. Naqvi, N. Sriraam, M. Mansoor, I. Badshah, M. Saleem and G. Kumaravelu. 'Portable Subcutaneous Vein Imaging System'. In: *International Journal of Biomedical and Clinical Engineering (IJBCE)* 2.2 (2013), pp. 11–22.

[39] G. P. Surampalli, J Dayanand and M Dhananjay. 'An Analysis of Skin Pixel Detection using Different Skin Color Extraction Techniques'. In: *International Journal of Computer Applications* 54.17 (2012).

[40] R. Azad and H. R. Shayegh. 'Novel and Tuneable Method for Skin Detection Based on Hybrid Color Space and Color Statistical Features'. In: *arXiv preprint arXiv:1407.6506* (2014).

[41] A. Kaur and B. Kranthi. 'Comparison between YCbCr color space and CIELab color space for skin color segmentation'. In: *IJAIS* 3.4 (2012), pp. 30–33.

[42] W. R. Tan, C. S. Chan, P. Yogarajah and J. Condell. 'A fusion approach for efficient human skin detection'. In: *IEEE Transactions on Industrial Informatics* 8.1 (2012), pp. 138–147.

[43] T. A. El-Hafeez. 'A new system for extracting and detecting skin color regions from pdf documents'. In: *International Journal on Computer Science and Engineering* 2.9 (2010), pp. 2838–2846.

[44] J. Berens and G. D. Finlayson. 'Log-opponent chromaticity coding of colour space'. In: *Pattern Recognition, 2000. Proceedings. 15th International Conference on*. Vol. 1. IEEE. 2000, pp. 206–211.

[45] itseez. *OpenCV findContours*. http://docs.opencv.org/2.4/modules/imgproc/doc/structural_analysis_and_shape_descriptors.html#findcontours. [Online; accessed 14-November-2016]. 2016.

[46] S. Suzuki et al. 'Topological structural analysis of digitized binary images by border following'. In: *Computer Vision, Graphics, and Image Processing* 30.1 (1985), pp. 32–46.

[47] T. M. Mitchell. 'Machine learning'. In: *New York* (1997).

[48] P. Bezák, Y. R. Nikitin and P. Božek. 'Robotic Grasping System Using Convolutional Neural Networks'. In: *American Journal of Mechanical Engineering* 2.7 (2014), pp. 216–218.

[49] J. Nagi, F. Ducatelle, G. A. D. Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber and L. M. Gambardella. 'Max-pooling convolutional neural networks for vision-based hand gesture recognition'. In: *Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on*. 2011, pp. 342–347. DOI: 10.1109/ICSIPA.2011.6144164.

[50] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang and G. Penn. 'Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition'. In: *2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP)*. IEEE. 2012, pp. 4277–4280.

[51] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella and J. Schmidhuber. 'Fast image scanning with deep max-pooling convolutional neural networks'. In: *arXiv preprint arXiv:1302.1700* (2013).

[52] A. Krizhevsky, I. Sutskever and G. E. Hinton. 'Imagenet classification with deep convolutional neural networks'. In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[53] J. T. Springenberg, A. Dosovitskiy, T. Brox and M. Riedmiller. 'Striving for simplicity: The all convolutional net'. In: *arXiv preprint arXiv:1412.6806* (2014).

[54] P. L. Callet, C. Viard-Gaudin and D. Barba. 'A Convolutional Neural Network Approach for Objective Video Quality Assessment'. In: *IEEE Transactions on Neural Networks* 17.5 (2006), pp. 1316–1327. ISSN: 1045-9227. DOI: 10.1109/TNN.2006.879766.

[55] V. Peddinti, D. Povey and S. Khudanpur. 'A time delay neural network architecture for efficient modeling of long temporal contexts'. In: *Proceedings of INTERSPEECH*. ISCA. 2015, pp. 2440–2444.

[56] D.-A. Clevert, T. Unterthiner and S. Hochreiter. 'Fast and accurate deep network learning by exponential linear units (elus)'. In: *arXiv preprint arXiv:1511.07289* (2015).

[57] G. B. Team. *TensorFlow*. https://www.tensorflow.org/. [Online; accessed 10-November-2016]. 2016.

[58] itseez. *OpenCV*. http://opencv.org/. [Online; accessed 10-November-2016]. 2016.

[59] K. Matsushita and E. Okada. 'Influence of adipose tissue on near infrared oxygenation monitoring in muscle'. In: *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*. Vol. 4. IEEE. 1998, pp. 1864–1867.

[60] A. E. Cerussi, A. J. Berger, F. Bevilacqua, N. Shah, D. Jakubowski, J. Butler, R. F. Holcombe and B. J. Tromberg. 'Sources of absorption and scattering contrast for near-infrared optical mammography'. In: *Academic radiology* 8.3 (2001), pp. 211–218.

[61] G. Naik, A. Al-Timemy and H. Nguyen. 'Transradial amputee gesture classification using an optimal number of sEMG sensors: an approach using ICA clustering'. In: (2015).