ISTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

GESTURE IMITATION LEARNING
IN HUMAN-ROBOT INTERACTION

M.Sc. THESIS

Itauma Isong ITAUMA

Department of Computer Engineering

Computer Engineering Programme

26 JANUARY 2012

ISTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

GESTURE IMITATION LEARNING IN HUMAN-ROBOT INTERACTION

M.Sc. THESIS

Itauma Isong ITAUMA
(504071541)

Department of Computer Engineering

Computer Engineering Programme

Thesis Supervisor: Assist.Prof.Dr. Hatice KOSE-BAGCI

26 JANUARY 2012

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TAKLİT YOLU İLE HAREKET ÖĞRENME
INSAN ROBOT ETKİLEŞİMİ

YÜKSEK LİSANS TEZİ

Itauma Isong ITAUMA
(504071541)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Assist.Prof.Dr. Hatice KOSE-BAGCI

26 OCAK 2012

Itauma Isong ITAUMA, a M.Sc. student of ITU Institute of Science and Technology 504071541 successfully defended the thesis entitled **“GESTURE IMITATION LEARNING IN HUMAN-ROBOT INTERACTION”**, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assist.Prof.Dr. Hatice KOSE-BAGCI** Istanbul Technical University	....
Jury Members :	**Assist.Prof.Dr. Sanem Sarıel Talay** Istanbul Technical University	....
	Assist.Prof.Dr. Gülay Öke Istanbul Technical University	....
		....

Date of Submission : **19 December 2011**
Date of Defense : **26 January 2012**

FOREWORD

I would not have been able to complete this work without the encouragement and support of many people. My first greatest gratitude and appreciation goes to my advisor, Assist.Prof.Dr. Hatice KÖSE-BAĞCI, for her direction, dedication and invaluable advice during this study.

I would like to thank my social robotics group members for thier helpful and constructive feedback over the course of this thesis and Assist.Prof.Dr. Sanem Sarıel Talay for her constant check in providing me with guidance and assistance in my studies.

I would like to extend gratitude to my fellow colleagues and friends Volkan Dağlı, Evangelos Sariyanidi and Cihan Tek for their invaluable academic and technical guidance. I would also like to to specially thank Rabia Yorganci, Hasan Kivrak, Evren Daglarli, Berat Denizdurduran, Meryem Önel, Hauwa Amal, and all my other friends and colleaques who have collaborated with me in my studies.

Also, this thesis would not have been completed without the support from my wife Dr. Omosalewa Itauma (nee Akinyemi).

And finally, I extend my special and profound thanks to my mum, Madam Idongesit Okon Udoh Affiah and other members of family for their love and support.

26 January 2012

Itauma Isong ITAUMA
Electrical Engineer (B.Eng)

TABLE OF CONTENTS

Page

ABBREVIATIONS

SL	**:** Sign Language
OPENNI	**:** Open Natural Interface
NI	**:** Natural Interaction
NN	**:** Neural Network
App	**:** Appendix
RGB	**:** Red Green Blue
RGBD	**:** Red Green Blue Depth
ESS	**:** Error sum-of-squares
HSV	**:** Hue, Saturation, Value
HMM	**:** Hidden Markov Model
SVM	**:** Support Vector Machine
RSR	**:** RshoulderRoll (Right shoulder Roll)
RSP	**:** RshoulderPitch (Right shoulder Pitch)
RER	**:** RelbowRoll (Right elbow Roll)
REY	**:** RelbowYaw (Right elbow Yaw)
RH	**:** Rhand (Right hand)
LSR	**:** LshoulderRoll (Left shoulder Roll)
LSP	**:** LshoulderPitch (Left shoulder Pitch)
LER	**:** LelbowRoll (Left elbow Roll)
LEY	**:** LelbowYaw (Left elbow Yaw)
LH	**:** Lhand (Left hand)
ITU	**:** Istanbul Technical University

LIST OF TABLES

LIST OF FIGURES

GESTURE IMITATION LEARNING
IN HUMAN-ROBOT INTERACTION

SUMMARY

This is an on-going study and part of a project which aims to assist in teaching Sign Language (SL) to hearing-impaired children by means of non-verbal communication and imitation-based interaction games between a humanoid robot and a child. In this paper, the problem is geared towards a robot learning to imitate basic upper torso gestures (SL signs) using different machine learning techniques. RGBD sensor (Microsoft Kinect) is employed to track the skeletal model of humans and create a training set. A novel method called Decision Based Rule is proposed. Additionally, linear regression models are compared to find which learning technique has a higher accuracy on gesture prediction. The learning technique with the highest accuracy is then used to simulate an imitation system where the Nao Robot imitates these learned gestures as observed by the users. Decision Based Rule had a 96% accuracy in prediction.

Futher more, this study also proposes an interactive game between a NAO H25 humanoid robot and preschool children based on Sign Language. Currently the demo is in Turkish Sign Language (TSL) but it will be extended to ASL, too. Since the children do not know how to read and write, and are not familiar with sign language, we prepared a short story including special words where the robot realized the specially selected word with sign language as well as pronouncing the word verbally.

After recognizing every special word with sign language the robot waited for response from children, where the children were asked to show colour flashcards with the illustration of the word. If the flashcard and the word match the robot pronounces the word verbally and continues to tell the story. At the end of the story the robot realizes the words one by one with sign language in a random order and asks the children to put the sticker of the relevant flashcard on their play cards which include the story with illustrations of the flashcards.

We also carried the game to internet and tablet pc environments. The aim is to evaluate the children's sign language learning ability from a robot, in different embodiments and make the system available to children disregarding the cost of the robot, transportation and knowhow issues.

This study started from developing a robust interface for rotation invariant Gesture Recognition. It involves the view-based detection and recognition of static hand gestures by using a single camera. Several image processing techniques are used to detect the hand region successfully. After the hand region is successfully detected geometric descriptors and Fourier descriptors are extracted. The gesture is classified using neural network.

The main contribution of this study is the features used for classification, a hybrid feature set consisting of Fourier Descriptors and a set of Geometric Descriptors which are introduced in this study. A color image segmentation algorithm is implemented to detect and segment the hand to create different feature sets. The proposed gesture recognition model has been used to control an autonomous mobile robot. The method has also been tested on different hand shapes and the result has been discussed.

With the availability of RGBD camera like kinect, the job on image processing in determining a good feature for gesture classification became easier. This study also presents human motion imitation using the calibrated skeletal view derived from openNI **(open Natural Interface)** connected to the Kinect camera. This study provides a system for computing the joint angles based on the kinematics of the skeletal view relative to the Kinect camera and these values are passed to a Nao robot simulated environment. In this study Choregraphe is used to simulate the Nao robot. Based on Nao's degree of freedom and kinematic constraints, estimated joint angles (recognized getures) are simulated to reflect a sense of imitation.

TAKLİT YOLU İLE HAREKET ÖĞRENME
INSAN ROBOT ETKİLEŞİMİ

ÖZET

Bu çalışma, insansı robot ve çocuk arasında sözlü olmayan iletişim ve taklit tabanlı etkileşim oyunları aracılığı ile işitme engelli çocukların İşaret Dili öğrenimine yardımcı olmayı amaçlayan devam eden bir çalışmadır. Bu çalışma farklı makine öğrenme teknikleri kullanarak 5 temel hareketi (jest) veya İşaret Dili hareketlerini taklit ederek öğrenme ile ilgilenmektedir. İnsan iskelet modelini izlemek ve bir eğitim seti oluşturmak için RGBD sensör (Microsoft Kinect) kullanılmıştır. Kural tabanlı hareket tanıma olarak adlandırılan yeni bir yöntem önerilmiştir. Ayrıca, hareketi tanıma da hangi öğrenme metodlarının daha doğru sonuç ürettiğini bulmak için lineer regresyon modelleri karşılaştırılmıştır. Kullanıcılardan alınan 5 farklı hareket Nao Robot tarafından en yüksek doğruluk oranına sahip öğrenme tekniği kullanılarak taklit edilmiştir. Kural tabanlı hareket tanıma yaklaşımı %96 doğruluk oranına sahiptir.

İşaret dili çalışmalarını gerçek robot üzerinde gerçekleştirdikten sonra robot ve insan arasında işaret diline dayalı çocuklara işaret dilinin temel kavramlarını öğretmek için interaktif bir oyun geliştirildi. Oyun Türk İşaret Dili (TSL) için tasarlandı ama Amerikan İşaret Dili (ASL) içinde gerçekleştirilmesi düşünülmektedir. Okuma yazma bilmeyen ve işaret diline aşina olmayan çocuklar için oyun içerisinde robotun kelimeleri sözlü olarak da söylediği ve işaret dili ile algılayibildiği özel olarak şeçilmiş işaret dili kelimeleri kullanıldı. Robot Türkçe bir hikaye anlatıyor. Hikaye içerisinde bazı kelimeleri işaret dili ile anlatıyor ve bekliyor. Çocuk bu kelimeleri tanıyıp uygun resimli kartı gösteriyor. Robot üzerindeki resim tanıma programı sayesinde bunu tanıyor ve eğer doğru resim gösterilmişse kelimenin ismini söylüyor ve hikaye devam ediyor. Doğru resim gösterilmemişse kullandığımız kurguya göre bekliyor ışıklarıyla ya da hareketleriyle bunun yanlış olduğunu ifade ediyor ve çoçuğu bir daha denemesi için teşvik ediyor. Hikayenin sonunda robot işaret dili kelimelerini rastgele olarak teker teker gerçekleyerek çocuktan oyun kartları içerisinden ilgili kelimenin resmini gösteren bilgi kartını ilgili yere yapıştırmasını istiyor. Bu oyunu tablet bilgisayar üzerine videoya dayalı bir şekilde taşıdık. Burdaki amacımız çocukların farklı cisimlerdeki robotlar tarafından işaret dili öğrenme yeteneğini değerlendirmek ve sistemi robot maliyeti göz önüne alınmaksızın, ulaşım ve teknik bilgi konularında çocuklara uygun hale getirebilmektir.

Bu çalışma değişmeyen rotasyon ile hareket tanıma için sağlam bir arayüz geliştirmekle başladı. Çalışmada tek bir kamera kullanarak statik el hareketlerini görünüme dayalı algılama ve tanıma gerçekleştirildi. El bölgesini tespit etmek için çeşitli görüntü işleme teknikleri başarıyla uygulandı. El bölgesi başarılı bir şekilde tespit edildikten sonra geometrik tanımlayıcılar ve Fourier tanımlayıcıları çıkarıldı. El hareketi yapay sinir ağları kullanılarak sınıflandırıldı.

Bu çalışmanın ana katkısı, sınıflandırma için kullanılan , fourier tanımlayıcılar ve geometrik tanımlayıcı dizilerinden oluşan melez özelliklerdir. Farklı özellik setleri oluşturmak için el tespit ve segmentasyonu için renkli görüntü bölütleme algoritması uygulanmıştır. Önerilen hareket tanıma modeli kendi kendini yöneten otonom bir mobil robotu kontrol etmek için kullanılmıştır. Yöntem farklı el şekilleri üzerinde test edilmiş ve sonuçlar tartışılmıştır.

Hareket sınıflandırması için iyi özelliği belirleyen görüntü işleme RGBD özelliğe sahip Microsoft Kinect kamerası kullanılmasıyla birlikte daha kolay hale geldi. Bu çalışma aynı zamanda insan hareketlerini, kinect kamerası için üretilen açık kaynak kodlu openNI (Open Natural Interface) kütüphanesinin sağladığı kalibre edilmiş iskelet görünümünü kullanarak taklit etmektedir. Ayrıca kinect kamerası ile elde edilen iskelet görünümü kinematiğine dayalı olarak eklem açılarının hesaplanmasını ve hesaplanan değerlerin similasyon ortamında Nao Robot üzerinde gerçeklenmesini içermektedir. Nao Rabot similasyonu için Choregraphe kullanılmıştır. Nao robotun serbestlik dereceleri ve kinematik kısıtlamaları temel alınarak hesaplanan eklem açıları (tanınmış hareketler) taklit duygusu yansıtacak şekilde simule edilmiştir.

Halen sürmekte olan bu çalışma, duyma özürlü çocuklara insansı robot ve çocuk arasında sözsüz iletişim ve emitasyon tabanlı etkileşim oyunları vasıtasıyla işaret dilini öğretmekte yardımcı olmayı hedeflemektedir. Bu çalışmada yönelinen problem, bir robotun farklı makine öğrenmesi teknikleri kullanarak 5 temel el hareketi veya işaret dili işaretlerini taklit etmeyi öğrenmesidir. RGDB sensörü (Microsoft Kinect) insanların iskelet modelini takip etmekte ve bir öğrenme kümesi oluşturmakta kullanılmıştır. Karar Tabanlı Kural isimli yeni bir metod önerilmiştir. Buna ek olarak, işaret dili tahmininde hangi öğrenme tekniğinin daha yüksek kesinliğe sahip olduğunun belirlenebilmesi için doğrusal regresyon modelleri karşılaştırılmıştır. En yüksek kesinliğe sahip öğrenme tekniği daha sonra kullanıcılar tarafından gözlemlenen Nao Robot'un 5 farklı işaret dilini taklit ettiği sistemi simule etmekte kullanıldı. Karar Tabanlı Kural method %96 kesinlik değerine sahiptir.

Bunların yanında, bu çalışmada ayrıca bir NAO H25 insansı robot ile okul öncesi çocuğun işaret dili tabanlı, etkileşimli bir oyun önerilmiştir. Şu anda demo Türk İşaret dilindedir ancak ASL için de genişletilecektir. Çocuklar okuma yazma bilmediği ve işaret diline alışkın olmadığı için, robotun işaret dili ile tanıdığı ve sözlü olarak telaffuz ettiği özel seçilmiş kelimeleri içeren kısa bir hikaye hazırladık.

Her özel kelimeyi bir işaret dili ile tanıdıktan sonra, robot çocuktan yanıt bekler. Çocuktan kelimenin ilustrasyonu ile renkli okuma fişini göstermesi istenmiştir. Eğer okuma fişi ile kelime eşleşirse robot kelimeyi sözlü olarak telaffuz eder ve hikayeyi anlatmaya devam eder. Hikayenin sonunda, robot kelimeleri rastgele olarak tek tek işaret dilinde anlar ve çocuktan, okuma fişlerinin ilüstrasyonları ile hikayeyi içeren oyun kartları üzerindeki ilgili okuma fişine etiket koymasını ister.

Oyunu ayrıca internet eve tablet kişisel bilgisayarlara da koyduk. Amaç çocukların işaret dili öğrenme yeteneğini farklı şekillerde ölçmek ve sistemi robotun masrafı, taşıması ve teknik bilgisi gibi konuları gözardı ederek çocuklara uygun hale getirmektir.

Bu çalışma, işaret dili sabitlerinin rotasyonu için güvenilir bir arayüz geliştirmekten başlamıştır. Görüntü tabanlı saptama ve sabit el hareketlerinin tek bir kamera kullanarak tanınmasını içerir. El bölgesinin başarı ile saptanması için pek çok görüntü işleme tekniği kullanılmıştır. El bölgesi başarılı bir şekilde saptandıktan sonar geometric tanımlayıcılar ve Fourier tanımlayıcılar çıkartılır. İşaretler yapay sinir ağları kullanılarak sınıflandırılır.

Bu çalışmanın ana katkısı sınıflandırma için kullanılan özelliklerdir. Bu çalışmada tanıtılan özellikler, Fourier tanımlayıcıları içeren hybrid bir özellik kümesi ve geometric tanımlayıcıları içeren bir kümedir. Farklı özellik kümeleri yaratmak için eli tesbit edip bölmelemekte bir renkli görüntü bölmeleme algoritması uygulanmıştır. Sunulan işaret tanıma modeli otonom, taşınabilir bir robotu kontrol etmek için kullanılmıştır. Bu method ayrıca farklı el şekilleri üzerinde test edilmiş ve sonuçlar tartışılmıştır.

Kinect gibi bir RGBD kameranın kullanılabilirliği ile işaret sınıflandırma için iyi özellikler belirleme daha kolay hale gelmiştir. Bu çalışma ayrıca Kinect kameraya bağlı openNI’dan türetilen ayarlanabilir iskelet görünüsü kullanarak insan hareketleri imitasyonunu sunmaktadır. Bu çalışma Kinect kamera ile ilgili iskelet görüntüsünün devimbilimsellerine dayalı ortak açıların hesaplanması için bir sistem sağlar. Bu çalışmada, Nao robotu simule etmesi için Choregraphe kullanılmıştır. Nao’nun özgürlük derecesine devimbilimsel kısıtlamalarına dayanarak tahmin edilen ortak açılar, imitasyon hissinin yansıtılmasını simule etmektedir.

1. INTRODUCTION

In recent years, research has progressed steadily in regards to the use of computers in recognizing or visualizing sign languages. Sign Language (SL) is a complete, complex language that employs signs made by hand motion combined with facial expressions and postures of the body.

This study started from developing a robust interface for rotation invariant Gesture Recognition. It involves the view-based detection and recognition of static hand gestures by using a single camera. Several image processing techniques are used to detect the hand region successfully. After the hand region is successfully detected, geometric descriptors and Fourier descriptors are extracted. With the availability of RGBD camera like Kinect, the job on image processing in determining a good feature for gesture classification became easier.

This paper presents human motion imitation using the calibrated skeletal view derived from OpenNI **(open Natural Interface)** [1] connected to the Kinect camera. This study provides a system for computing the joint angles based on the kinematics of the skeletal view relative to the Kinect camera and these values are passed to a Nao robot simulated environment. In this study and in [2], Choregraphe is used to simulate the Nao robot. Based on the Nao's degree of freedom and kinematic constraints, estimated joint angles (recognized gestures) are simulated to reflect a sense of imitation. A complete description of the Nao robot as shown in Figure 1.1 is available in [3]. This study contributes to the development of different learning techniques that recognize different human gestures.

The previous work is explained in Section 1.2. The implementation and the methods for data preparation are detailed in Chapter 2.1. In this section, two different data sets are created- a training set, and a test set. Next, we performed offline supervised learning (Chapter 3.1) to compute parameters of different models in the literature. A novel method called Decision Based Rule is also proposed. Finally, the computed

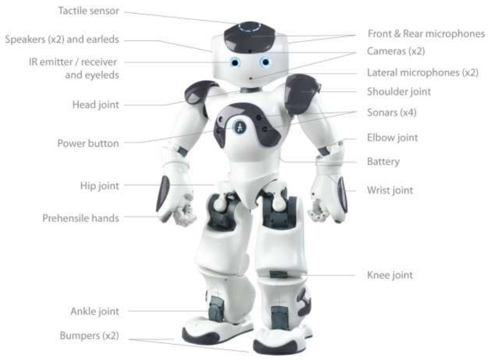


Figure 1.1: Nao description [3].

parameters of the different models are tested on the test set to determine which learning technique has a higher accuracy and less error in prediction as shown in Chapter 4. Conclusion and future work are presented in Chapter 5.

1.1 Purpose of Thesis

The purpose of this study is to design a system from which a humanoid robot can imitate upper body gestures with the aim of using it to teach sign languages to people (especially children) through interactive games.

1.2 Background

The purpose of this study is to design a system from which a humanoid robot can imitate upper body gestures with the aim of using it to teach sign languages to people (especially children) through interactive games. In our previous study [4], the robot is able to express a word in Sign Language (SL) among a set of chosen words using hand movements, body and face gestures. On having comprehended the word, the child will give relevant feedback to the robot.

This paper is part of a novel study which proposes an interaction game [5], [6] based on Sign Language, between children and a humanoid robot (currently Nao H25 with fingers). The aim of this game is to assist sign language tutoring especially for preschool children [7], [8], [9]. During the game, the robot tells a simple and short child story (Figure 1.2) and within the story uses some special words in sign language. The game is based on interaction, sign language interpretation (gesture implementation and recognition) and turn-taking.

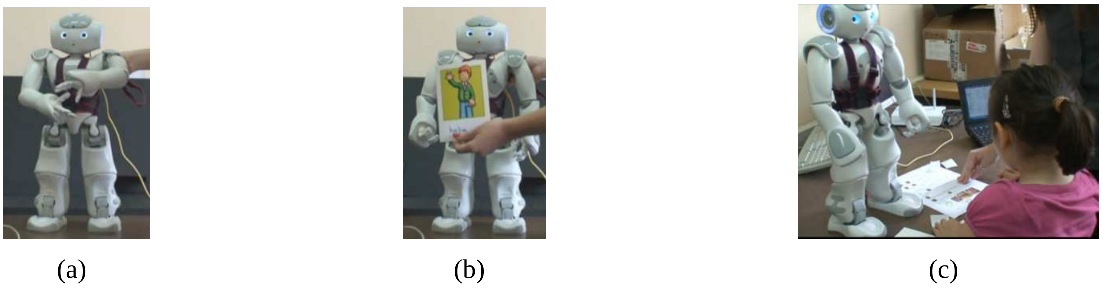
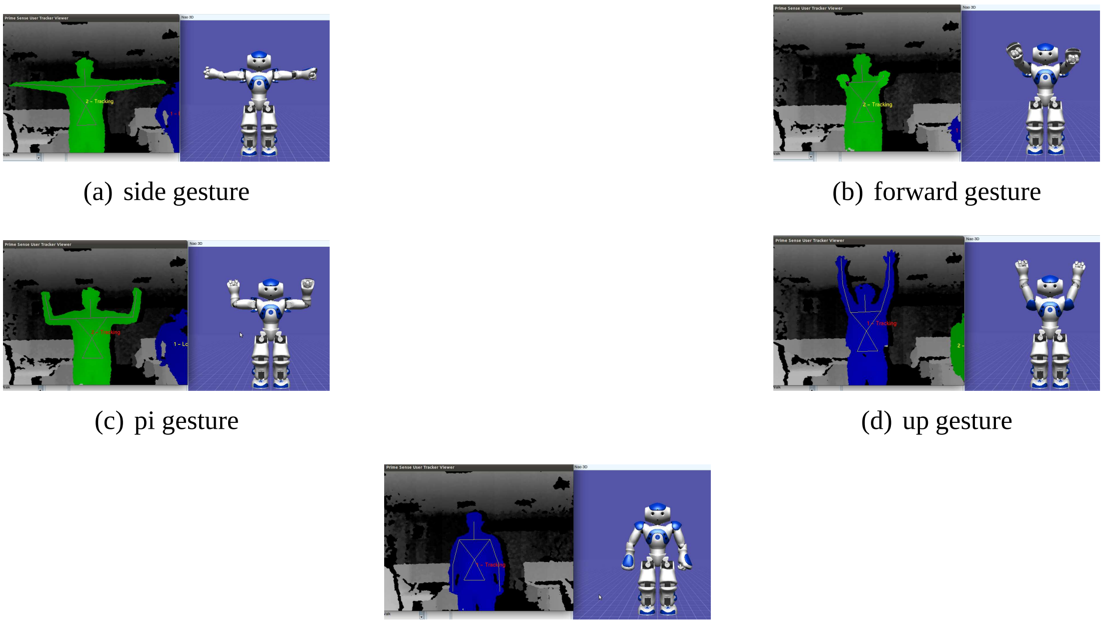


Figure 1.2: Screen shots from the game [10]
(a) Robot performing a sign, (b) Child shows the colored card of the word "dad" to robot and (c) Child completes the playcard using stickers of the colored cards.

Further more, this study will emphasize how the robot is taught to imitate human gestures. The problem is simplified by teaching the robot five basic gestures or SL signs as shown in Figure 1.3 using different machine learning techniques. The robot learns to imitate the human using the learned model. The RGBD sensor is used to compute joint angles from the skeletal model of the human using an approach based on Pythagoras theorem [11].

Our work is inspired by previous studies which use three-dimensional (3D) Cartesian coordinates (XYZ position). In [12], a system that recognizes gestures using 3D trajectories consisting of a reduced set of key-points was proposed. This was extracted from their novel adaptive curvature function. In [13], the authors described trajectory learning from multiple demonstrations with a 3D dimensional model of the human hand for pick and place operations. In [14], the authors proposed the Maximum Margin algorithm that solves imitation problems by learning linear mappings from features to cost functions in a planning domain. Also, [14] demonstrated that imitation learning of long horizon and goal-directed behavior can be naturally formulated as a structured prediction problem over a space of policies. In [15], the authors discuss that imitation learning is reduced to a regression problem. In addition, [15] demonstrated the validity of their approach by learning to map motion capture data from human actors to a humanoid robot, and the composition of several regression models yields qualitatively better imitation results than using a single, more complex regression model. Hidden Markov Models (HMMs) can be considered an advanced modeling scheme used in gesture recognition which have been employed in similar works [16], [17].



(a) side gesture

(b) forward gesture

(c) pi gesture

(d) up gesture

(e) down gesture

Figure 1.3: Five basic gestures.

Different techniques in solving the problem of gesture recognition has been widely investigated. Hand gesture recognition can be very roughly considered as the successful accomplishment of three large sets: Hand region segmentation, extraction of descriptors and gesture classification. The generic flowchart is shown in Figure 1.4.

1.2.1 Hand region segmentation

In this section the methods which have been used for hand region segmentation in literatures is discussed. This study focuses on the View-based methods. Wrist-cropping techniques also exist, such as in the study of Wah Ng and Ranganath [27].

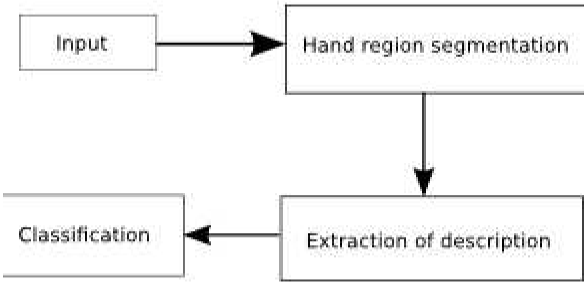


Figure 1.4: Flowchart of the generic approach used in literature for solving hand gesture recognition problem.

1.2.1.1 Color based segmentation

Information has been widely used due to the distinctive color of the hand. In the popular study of Brand, Oliver and Pentland [28] titled "A Bayesian Computer Vision System for Modeling Human Interactions", color information has been used. As its name suggests, this work is a more comprehensive study involving the modeling of human attitude. Another study involving hand segmentation based on color information is the study of Kjeldsen and Kender [29], which dedicated only for solving the detection of the skin color and not recognition of hand gesture. A comprehensive study for color based detection is the study of J. Yang, A. Waibel [30], which actually involves face detection but can operate equally well for hand gesture recognition. One other publication regarding gesture recognition in which hand detection problem is solved via color information is the paper of Chan Wah Ng, Surendra Ranganath [27]. This paper involves a real-time application development using hand gesture recognition. Color segmentation has been widely used with different techniques and in several color spaces, such as HSV, YCbCr which are much more convenient forms for color-based segmentation.

1.2.1.2 Gray level based segmentation

Another popular approach is the usage of intensity levels in detection of hand region. Eickeler, Kosmala and Rigoll worked for real-time hand gesture recognition in their study titled "Hidden Markov model based continuous online gesture recognition" [31]. The gray level detection takes advantage of the hand shape information, but usually it is not sufficient, therefore enhanced methods such as neural network based classification are necessary. Intensity-level information is mostly combined with other information such as motion and color.

1.2.1.3 Motion based segmentation

Motion provides very valuable information in the accomplishment of the goal mentioned in this paper. Filtered conveniently, motion gives a huge clue about gesture.

A sample study involving pure motion is the study of Kohler [32] in which success under unconstrained conditions was aimed.

1.2.1.4 Combination of different information

It is obvious that human hand has a very unique color compared to many objects and scenes in real life. Therefore color information is not discarded in most studies. But color by itself is mostly insufficient for robust segmentation. Combining information from different sources is mostly the key in success of such applications. Different combinations of color, motion and shape have been widely used. For one, is the study of Crowley [33] in which tracking of human body parts is discussed, where gray-level and color data are combined. In the popular paper of Triesch and Malsburg [25] color and motion data are combined. The combination of color and motion is really widely used.

1.2.2 Feature extraction

There are many candidates in literature for hand feature extraction. Fingertips have been used in many papers (Davis and Shah [34] , Nolker and Ritter [35]). Geometric properties are also widely used. One instance is the popular work of Starner [36] regarding the interpretation of American Sign Language. Another work is the study of Brand, Oliver and Pentland [28], in which several region properties have been used, such as elongation, eccentricity, centroid, mass etc. A famous paper is the paper of Freeman [37], which is based on orientation histogram usage for hand gesture recognition. However orientation histograms are not very robust in this case, so they are not used widely.

Fourier descriptors are also very popular and very handy as they provide very robust information, data invariant to translation, rotation and scaling [38]. The drawback of this method is that it is totally 2D dependent, as many of the descriptors mentioned here. Fourier descriptors have been used in the following studies: “Real-time gesture recognition system and application” [27],“Hand gesture recognition using a real-time tracking method and hidden Markov models” [39]. These are only two sample studies, Fourier descriptors have been subject of countless hand gesture recognition studies.

1.2.3 Classification

Nearly in every study involving hand gesture recognition mentioned so far, machine learning based classification is used instead of rule based classification. A very rare example to a study in which machine learning is not used, is the study of Triesch and Malsburg [25] in which elastic graph matching has been preferred for classification. The focus of this work has been researching, comparing and testing several segmentation techniques and various features, and also comparing classification methods. A backpropagation network has been used and it has been the only classification algorithm tested, no comparison between methods has been made.

However in literature Hidden Markov Models, various neural networks and the combination of these two have been widely used, compared and tested.

2. IMPLEMENTATION

Several works in the literature make use of supervised learning such as [15] and [18]. In supervised learning, a feature vector and a target label $x_1, x_2, x_3, \ldots, x_n - > Y$ are assumed to be given. For example, the feature vectors are different computed arm joint angles and the target label is the desired gesture. Machine learning is carried out on past experience to create a hypothesis that fits the features to the labels. The goal is to choose a function among a family of functions $f(x) = Y$ that allows us to predict gestures Y based on new feature data x. Ten different discriminative arm joint angles ('RShoulderRoll', 'RShoulderPitch', 'RElbowRoll', 'RElbowYaw', 'RHand', 'LShoulderRoll', 'LShoulderPitch', 'LElbowRoll', 'LElbowYaw', 'LHand') are used for the feature set. In supervised learning, given a dataset:

$$DataSet_{m,n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} & - > & Y1 \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} & - > & Y2 \\ \vdots & \vdots & \ddots & \vdots & & \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} & - > & Ym \end{pmatrix}$$

Given any x_m or future vector x, it should predict the target label Y. $f(x_m) = Y$.

2.1 Data Preparation

2.1.1 Feature generation procedure

A platform known as OpenNI [1] (Open Natural Interaction) provided by Prime Sense [19] is used to interface with the Kinect sensor unit. It offers a good solution as it has already been used to track a person and his joints in 3D space. NiUserTracker sample code is used as a base for our implementation. The depth-sensor of Kinect is used to gather depth information that enables OpenNI to gather the xyz coordinate system of the scene. Using an RGB camera instead would be computationally expensive. The

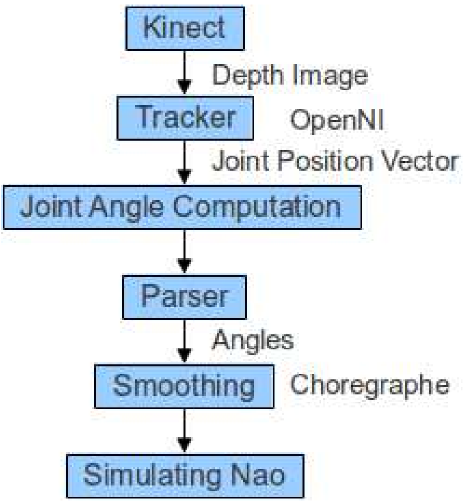


Figure 2.1: One-to-One human-robot imitation system.

depth information is used for segmentation and 3D scene recognition for tracking the calibrated human body.

Basic mathematical geometry of 3D vectors is used in the computation of joint angles of the different joint poses. A one-to-one mathematical model for human motion imitation using the calibrated skeletal view derived from OpenNI connected to the RGBD sensor was developed [20]. Using OpenNI, the human body is calibrated and a skeletal model is segmented and tracked. The computed joint angles are passed to Choregraphe [21], which is a graphical environment developed by Aldebaran Robotics for simulating the Nao robot based on joint angle constraint. We are thereby able to map the human joints to the corresponding joint of the Nao robot to simulate an imitation system (Figure 2.1).

2.1.2 Joint angle computation

- Get Skeleton Joint Positions rightshoulderJoint, rightelbowJoint, righthandJoint in terms of (x,y,z) plane.
- Build the directional vectors between the joints as shown in Figure 2.2.
- Compute roll,pitch,yaw rotation matrix by projecting the vectors to the xy/yz/xz-plane or axis.
- Compute the vector in dependencies to the previous joints (elbow depends on shoulder and shoulder to torso).

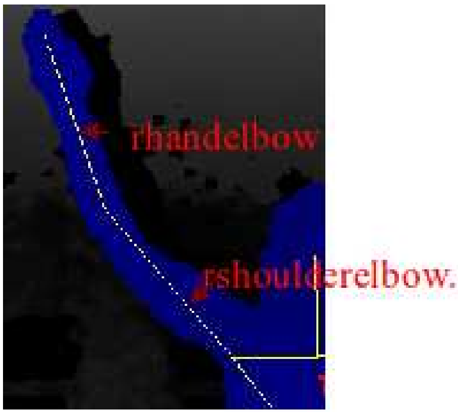


Figure 2.2: Representation of shoulder-elbow vector and hand-elbow vector.

The dot product of the directional vectors (righthandelbow and shoulderelbow) between the shoulder and elbow and the elbow to hand is computed in order to calculate the angles for the RightElbowRoll and LeftEblowRoll.

2.1.3 Angle between 3D vectors [22]

The dot product is used for computing the angle: $\cos\theta$ is equal to dot product of two vectors. The formula for the angle θ between two vectors is:

$$\cos\theta = \frac{f \cdot g}{\|f\| \cdot \|g\|} \tag{2.1}$$

Given two vectors: $f = \left(x_f, y_f, z_f\right)^T$ and $g = (x_g, y_g, z_g)^T$

The angles that separates the two vectors are computed below:

1. Lengths are:

$$|f|^2 = \left(x_f, y_f, z_f\right)^T \cdot \left(x_f, y_f, z_f\right)^T = x_f^2 + y_f^2 + z_f^2, \tag{2.2a}$$

$$|g|^2 = (x_g, y_g, z_g)^T \cdot (x_g, y_g, z_g)^T = x_g^2 + y_g^2 + z_g^2, \tag{2.2b}$$

2. The normalized vectors are:

$$f_u = \left(x_f, y_f, z_f\right)^T / \sqrt{x_f^2 + y_f^2 + z_f^2}, \tag{2.3a}$$

$$g_u = (x_g, y_g, z_g)^T / \sqrt{x_g^2 + y_g^2 + z_g^2}, \tag{2.3b}$$

3. The dot product is:

$$f_u \cdot g_u = \left(x_f, y_f, z_f\right)^T \cdot \left(x_g, y_g, z_g\right)^T / \left(\sqrt{x_f^2 + y_f^2 + z_f^2}\sqrt{x_g^2 + y_g^2 + z_g^2}\right), \tag{2.4a}$$

$$= \left(x_f \times x_g + y_f \times y_g + z_f \times z_g\right) / \left(\sqrt{x_f^2 + y_f^2 + z_f^2}\sqrt{x_g^2 + y_g^2 + z_g^2}\right), \tag{2.4b}$$

4. The angle is:

$$\cos\theta = \left(x_f \times x_g + y_f \times y_g + z_f \times z_g\right) / \left(\sqrt{x_f^2 + y_f^2 + z_f^2}\sqrt{x_g^2 + y_g^2 + z_g^2}\right), \tag{2.5a}$$

$$\theta = \arccos\left(x_f \times x_g + y_f \times y_g + z_f \times z_g\right) / \left(\sqrt{x_f^2 + y_f^2 + z_f^2}\sqrt{x_g^2 + y_g^2 + z_g^2}\right), \tag{2.5b}$$

We show that it is possible to use the RGBD camera (Kinect) to implement an imitation system. The computation was based on points in a 3D Cartesian coordinate system.

A survey was carried out on several users as shown in Table 2.1. This is the result of our previous work which from observation, the simulated Nao robot imitated actions considered unsafe to apply on a real Nao robot. The problem was due to noise in the environment and the misalignment of the segmented skeletal image from which the human pose is computed.

2.1.4 Good arm joint feature

Based on the one-to-one robot control shown in the previous section and the observations from different users, we noticed that there are manipulations that can not be implemented on the real Nao robot due to its degree of freedom and singularities. So we decided to make use of different machine learning techniques to create a system in which the Nao robot learns the observed human gesture and performs the right imitation. We implemented this learning system using Decision Based Rule [17], linear regression learning techniques [18], [15] for offline learning of joint feature parameters for robot control which will be explained in the next section. Figure 2.3 shows the system design for Joint Angle computation.

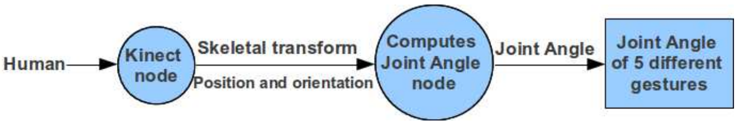


Figure 2.3: Data preparation system design.

Table 2.1: One-to-One imitation demo survey.

Name	Sex	Demo	Comments
User1	female	[20]	Very good, slow, some computation errors.
User2	male	[20]	Successfully imitates though a few computation errors.
User3	male	[20]	Slow due to the speed of the PC. The left arm imitates accurately but the right arm does not imitate all gestures. In general, works successfully.
User4	male	[20]	There are some gestures that can not be performed on the right arm due to wrong computation.
User5	male	[20]	The left arm imitates accurately but there are certain manipulations in which the right arm gives a wrong imitation. This study can be used in games.
User6	male	[20]	A nice project. Responds to movement with ease. The right arm was not responding well compared to the left.
User7	male	[20]	Good initiative and successful project. It worked really fine on all gestures tested.
User8	male	[20]	A very good study. Perhaps a learning method could be adapted.

Arm gestures are the primary component of sign language communication. Hence, a good system that accurately classifies the needed features of the arm for sign language gesture recognition is neccesary. The use of joint angles base on the orientation of the arms reduces the challenge faced in Human Robot Interaction research, in respect to tracking and segmentation of human arms using different computer vision techniques. Several proposed solutions addressing this issue constrain the users actions one way or the other thereby limiting the degree of communication. These constraints vary from making the user wear markers or position his arms at a particular distance relative to the camera. However, while performing sign language gestures, most of these constraints affects the users sign space. Thereby the gestures performed by the users are not in the natural sequence. The use of machine learning techniques helps to remove this constraints such that the users can express the sign language as natural as possible.

[26] proposed a new method for recognising primitive movements using Bayesian classifiers which can be applied in complex motion analysis. Sign language recognition

Table 2.2: Dataset of joint angle gestures of different users.

RSR	RSP	RER	REY	RH	LSR	LSP	LER	LEY	LH	Behavior
87	6	165	92	7	-86	7	159	-87	3	side
18	17	165	0	8	-9	10	171	0	2	forward
76	0	98	9	86	-78	-1	97	-12	85	piUp
27	81	155	-12	89	-33	81	151	9	92	Up
22	-85	161	-2	-90	-22	-82	150	5	-89	down
84	-11	168	79	4	-80	-21	160	-80	4	side
22	0	159	3	34	-11	0	156	7	41	forward
72	2	98	4	77	-87	3	95	1	77	piUp
24	83	169	6	89	-22	93	169	-3	86	Up
23	-88	161	-5	-88	-18	-87	160	5	-85	down
87	4	163	79	-4	-88	-7	174	-78	-4	side
13	5	169	0	0	-21	3	165	0	0	forward
67	-3	94	19	75	-72	-6	97	-22	76	piUp
32	84	168	11	90	-39	80	163	-13	86	Up
22	-88	153	3	-81	-21	-89	153	0	-83	down
51	-7	40	-40	-28	-2	-46	37	-4	53	side
5	-30	64	0	0	12	-48	37	-19	48	forward
5	-30	64	0	0	12	-48	37	-19	48	piUp
18	92	162	-6	88	-27	86	164	-10	89	Up
15	-71	161	2	-78	-10	-75	158	-3	-76	down

is a complex problem, which requires a divide-and-conquer approach. Complex sign recognition can be considered as recognition of a sequence of primitive movements. It is, however, usually difficult to recognise primitive movements from raw images. This is mainly because getting motion information from raw images usually involves target detection and visual tracking that are also complex problems in computer vision.

3. SUPERVISED LEARNING

Imitation learning can be seen as a subset of Supervised Learning. In Supervised Learning the system is presented with labeled training data and learns an approximation to the function which produced the data.

Learning capabilities are essential for successful integration of robots in human-robot domains, in order to learn from human demonstrations and facilitate natural interaction with people.

3.1 Learning Techniques

3.1.1 Model representation

The task is to predict the correct gesture based on user joint angles of the different features. This problem is tackled as both a regression whereby we predict real-valued output and a classification whereby we predict discrete-valued output.

Table 3.1 and Table 3.2 show the experimental statistics carried out in this study in generating both the training set and test set.

- *Notations used:*
  - n = number of features
  - m = number of training examples
  - x's = input features (joint angles)
  - y's = output gesture (side= 1, forward = 2, piUp = 3, up = 4, and down = 5)
  - (x,y) = one training example
  - $(x^{(i)}, y^{(i)}) = i^{th}$ training example

Table 3.1: Training set.

Number of Users	10
Number of features (n)	10
Number of examples (m)	50
Number of behaviors	5

Table 3.2: Test set.

Number of Users	9
Number of features (n)	10
Number of examples (m)	45
Number of behaviors	5

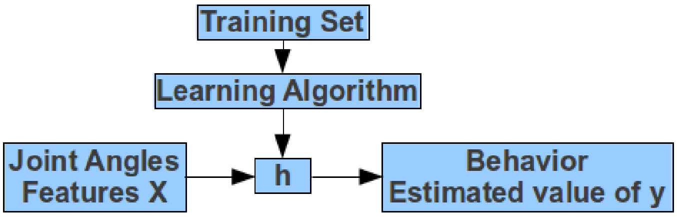


Figure 3.1: Behavior prediction model.

In designing a learning algorithm, we need to decide how we represent the hypothesis. Figure 3.1 shows a general representation of deciding on a hypothesis that can predict a behavior from the set of joint angle features. The hypothesis (h), maps x's to y's.

3.1.2 Data visualization

Before determining the hypothesis, we first visualized the data to see the sparsity. Figure 3.2 shows that our dataset cannot be separated based on the side and piUp gestures as shown in Figure 1.3(a) and Figure 1.3(c) by a straight line through the plot. Therefore, a straight forward application of logistic regression would not perform well on this dataset since logistic regression only finds a linear decision boundary. We decided to check the data set for features that are linearly dependent.

Figure 3.3 shows that our dataset can be separated after reducing the feature set to ensure that all features are not linearly dependent. Therefore, a straightforward application of linear regression and logistic regression is expected to perform well on this dataset. The next section shows our attempt using linear regression.

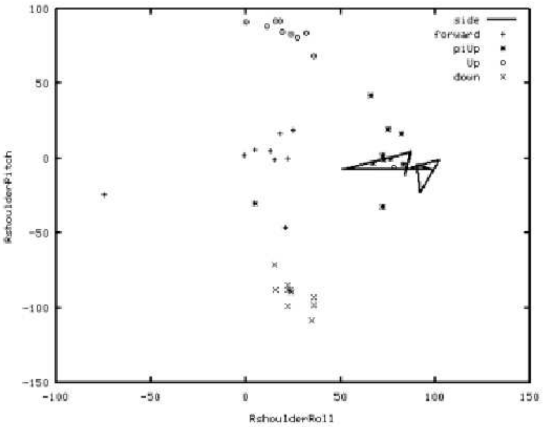


Figure 3.2: Plot of training data with poor sparsity.

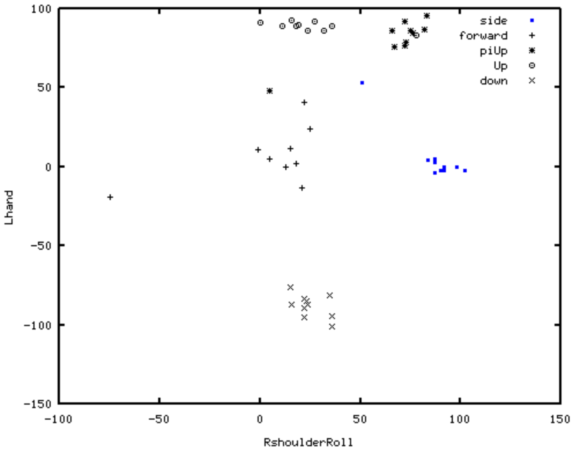


Figure 3.3: Plot of training data with good sparsity.

3.1.3 Linear regression analysis (LRA)

In linear regression, the idea is to choose $\theta's$ so that $h_\theta(X)$ is close to y for our training examples (x,y). This can be solved as an optimization problem:

$$\min_{\theta_0,\theta_1} \Sigma_{i=1}^{m}(h_\theta(X^{(i)})-y^{(i)})^2 \qquad \textbf{(3.1)}$$

Find the value of θ_0, θ_1 that makes the equation minimized. This is a squared error function used for regression problems. In this study, we have a set of 10 features. The form of hypothesis is:

$$h_\theta(X) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_{10} X_{10} \qquad \textbf{(3.2)}$$

For convenience of notation, define $X_0 = 1$

$$X = \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_1 0 \end{bmatrix} \in \mathbb{R}^{n+1},\ \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_1 0 \end{bmatrix} \in \mathbb{R}^{n+1}$$

$h_\theta(X) = \theta^T X$

This is known as *Multivariate linear regression*. We have multiple features in which we try to predict the value y. In this study, both Gradient Descent and Normal Equation learning models (Figure 3.4) are used and compared.

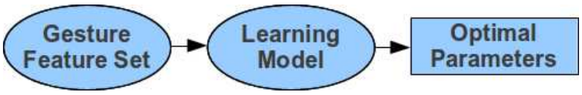


Figure 3.4: Offline supervised learning.

3.1.3.1 LRA with gradient descent

For $n >= 1$ the gradient descent algorithm is:

$repeat\{$

$\theta_i := \theta_i - \alpha \frac{\partial}{\partial \theta_i} J(\theta)$

$\}$ (simultaneously update θ_i for every $i = 0, \ldots, n$)

where $\frac{\partial}{\partial \theta_i} J(\theta) = \frac{1}{m} \Sigma_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) X_i^{(i)}$

The Gradient descent is used for minimizing the cost function $J(\theta)$, and α is the learning rate. Using the data set as represented in Table 3.1, Figure 3.5 shows 10 training examples.

$$X = \begin{array}{cccccccccc} 87 & 6 & 165 & 92 & 7 & -86 & 7 & 159 & -87 & 3 \\ 18 & 17 & 165 & 0 & 8 & -9 & 10 & 171 & 0 & 2 \\ 76 & 0 & 98 & 9 & 86 & -78 & -1 & 97 & -12 & 85 \\ 27 & 81 & 155 & -12 & 89 & -33 & 81 & 151 & 9 & 92 \\ 22 & -85 & 161 & -2 & -90 & -22 & -82 & 150 & 5 & -89 \\ 84 & -11 & 168 & 79 & 4 & -80 & -21 & 160 & -80 & 4 \\ 22 & 0 & 159 & 3 & 34 & -11 & 0 & 156 & 7 & 41 \\ 72 & 2 & 98 & 4 & 77 & -87 & 3 & 95 & 1 & 77 \\ 24 & 83 & 169 & 6 & 89 & -22 & 93 & 169 & -3 & 86 \\ 23 & -88 & 161 & -5 & -88 & -18 & -87 & 160 & 5 & -85 \end{array}, Y = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

Figure 3.5: 10 Training Examples.

Figure 3.6 shows the convergence of gradient descent with the best learning rate found. With a small learning rate, the gradient descent takes a very long time to converge to

Table 3.3: LR with gradient descent optimized value for θ.

LR parameters (θ)
1.902 -0.109 -0.106 0.062 -0.268 -0.055 0.033 -0.052 0.046 0.320 -0.109

Table 3.4: A sample instance and the predicted gesture.

X	Y
[-13 42 170 -6 30 40 16 3 -19 31]	2.237

the optimal value. Conversely, with a large learning rate, the gradient descent might not converge or might even diverge!

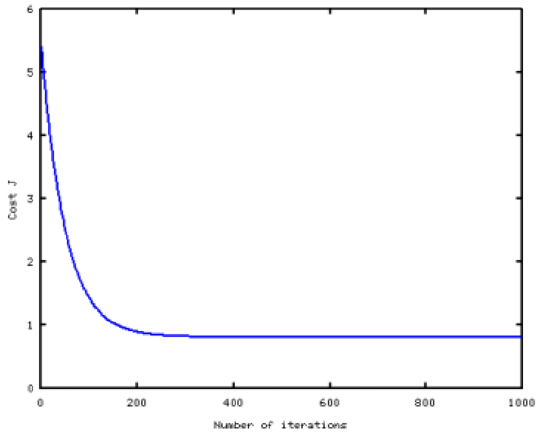


Figure 3.6: Convergence of gradient descent with an appropriate learning rate.

The gradient descent was run until convergence to find the final values of θ. Next, the value of θ was used to predict the behavior of users from the computed joint angle features as shown in Table 3.3 and Table 3.4. Out of 50 gestures from 10 users the system had a 70% accuracy.

3.1.3.2 LRA with normal equation

It gives a better way to solve for the parameter θ analytically rather than solving it iteratively using Gradient Descent. In our experiment, we have $m = 50$ training examples. In order to implement this Normal Equation Method, an extra column $X_0 = 1$ is added as shown in Table 3.5. Then we created a matrix of all features and called it X. We did the same for label y which we want to predict.

Table 3.5: Feature set X (joint angles).

X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	87	6	165	92	7	-86	7	159	-87	3
1	18	17	165	0	8	-9	10	171	0	2
1	76	0	98	9	86	-78	-1	97	-12	85
1	27	81	155	-12	89	-33	81	151	9	92
1	22	-85	161	-2	-90	-22	-82	150	5	-89
1	84	-11	168	79	4	-80	-21	160	-80	4
1	22	0	159	3	34	-11	0	156	7	41
1	72	2	98	4	77	-87	3	95	1	77
1	24	83	169	6	89	-22	93	169	-3	86

The normal equation is:

$$\theta = (X^T X)^{-1} X^T y \tag{3.3}$$

which gives the value of θ that minimizes the cost function. Using Normal Equation, feature scaling is not needed. Out of 50 gestures from 10 users, the system had a 72% accuracy.

3.1.3.3 Comparison of methods

So long as the number of features is not too large, the Normal Equation gives us a great alternative method to solve for the parameters θ. Hence, so long as the number of features $n < 10^3$, Normal Equation is preferable. Table 3.6 shows the comparison of the two learning models.

Table 3.6: Comparison of the different linear regression methods.

Gradient Descent	Normal Equation
Need to choose α Needs many iterations Works well even when n is large	No need to choose α Do not need to iterate Need to compute $(X^T X)^{-1}$ of order $n \times n$ Slow if n is very large

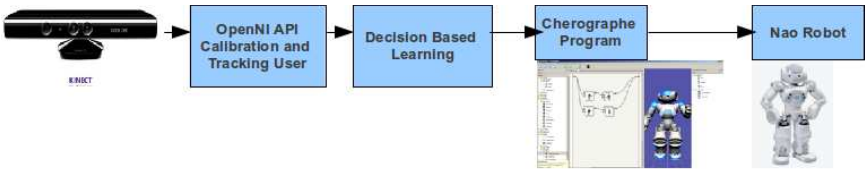


Figure 3.7: Decision-Based learning gesture detection pipeline.

3.2 Online Imitation System

Online learning from data sequences is a challenging aspect of machine learning. In this study, we developed a novel idea called *Decision Based Rule (DBR) Online Learning* which learns from streams of generated pose (position and orientation) of a human hand joint as shown in Figure 3.7. Online learning allows learning from a continuous stream of data (Figure 3.8). Decision is based on the current instance of data. It can adapt to changing user poses (i.e, if $p(y|x,\theta)$ changes over time). For recogntion, we need obserprocessingvations, or data.

The first robotics work to address imitation was focused on assembly task-learning from observation. Typically, a series of arm trajectories of a human, performing object moving/stacking tasks, were recorded either using a manipulandum, with the advantage of measuring directly the joint torques or using video images. Data were analyzed to remove inconsistencies and extract key features of movement. An industrial non-human-like robotic arm would then be trained to reproduce the trajectory which maximizes the data key features. These efforts constitute a significant body of research in robotics, and contribute to data segmentation and understanding.

3.2.1 Decision-Based human-robot imitation learning method

- Given four problems:

  Hands Up, Down, Forward and Sideways as shown in Figure 3.9

- Goal: We derived a cost function that takes into consideration the space of manipulation.

  (costHandsUp × HandsUp) + (costHandsDown × HandsDown) + (costHandsForward × HandsForward) + (costHandsSideways × HandsSideways),

- Problem Constraint:

  (positionHandUp × HandsUp) + (positionHandDown × HandsDown) + (positionHandForward × HandsForward) + (positionHandSideways × HandsSideways) < T,

  Where T is a position threshold in the x, y, z coordinate dependent on the Shoulder and Elbow position of the user.

  Behavior Decision depends on the difference between the Shoulder and Elbow joint position and direction vector relative to the Kinect (RGBD) camera origin.

- Simulation Decision Constraint for behavior selection algorithm:

```
{

if ((DiffBetRShoAndRElbowPosZ > T1 && DiffBetRShoAndRElbowPosZ < T2) &&
(DiffBetRShoAndRElbowPosX > T3 && DiffBetRShoAndRElbowPosX < T4) && directionVector > 0)
SetBehavior = HandsUp;

else if ((DiffBetRShoAndRElbowPosZ > T5 && DiffBetRShoAndRElbowPosZ < T6) &&
(DiffBetRShoAndRElbowPosY > T7 && DiffBetRShoAndRElbowPosY < T8))
SetBehavior = HandsSideways;

else if ((DiffBetRShoAndRElbowPosX > T9 && DiffBetRShoAndRElbowPosX < T10) &&
(DiffBetRShoAndRElbowPosY > T11 && DiffBetRShoAndRElbowPosY < T12) )
SetBehavior = HandsForward;

else if ((DiffBetRShoAndRElbowPosZ > T13 && DiffBetRShoAndRElbowPosZ < T14) &&
(DiffBetRShoAndRElbowPosX > T15 && DiffBetRShoAndRElbowPosX < T16) && directionVector < 0)
SetBehavior = HandsDown;

}
```

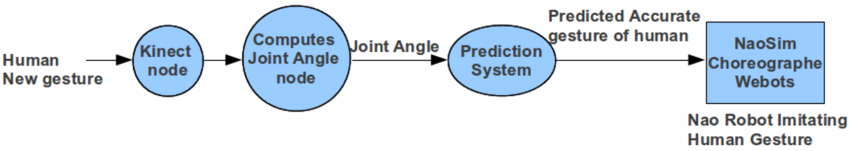
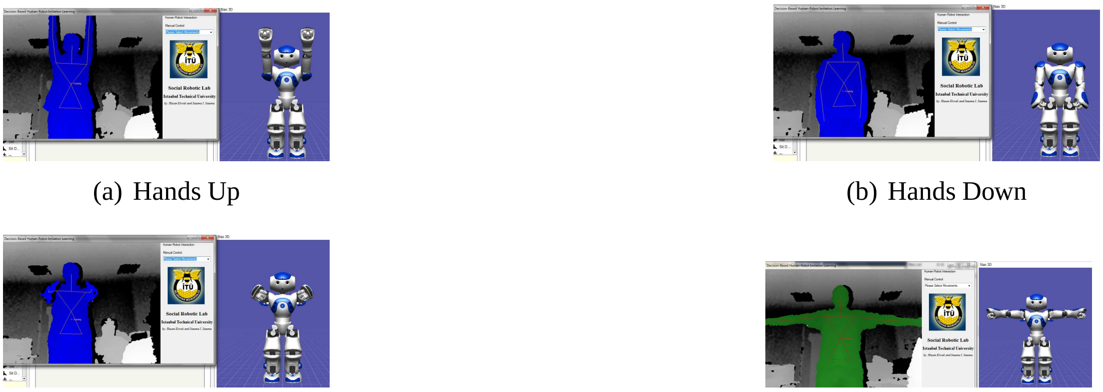


Figure 3.8: Online imitation system.



(a) Hands Up

(b) Hands Down

(c) Hands Forward

(d) Hands Sideways

Figure 3.9: Four basic gestures.

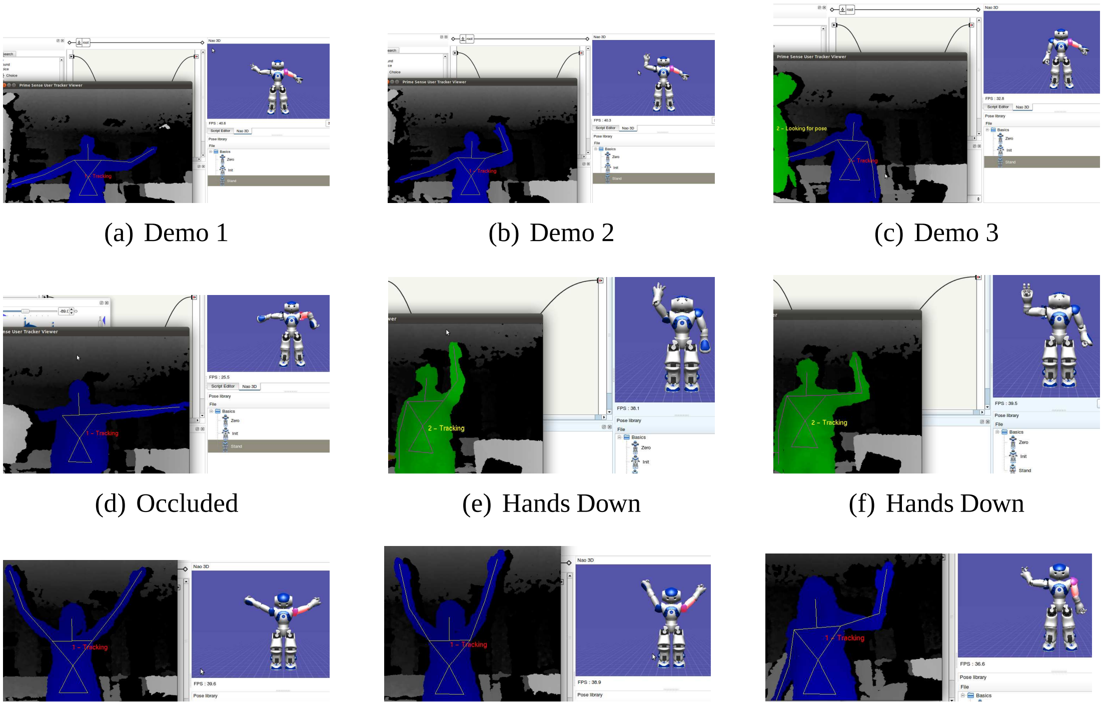
4. EXPERIMENTAL RESULTS

In this section, we present the performance of the different methods experimented in this study. The performance was tested on several participants to detect its level of accuracy. We present the results of the one-to-one real-time imitation system, the results of the linear regression normal equation (NE) learning method and our decision based rule (DBR) method.

The learning model was trained to distinguish between side, forward, pi, up and down gestures. Table 4.1 shows a confusion matrix that summarizes the results of linear regression normal equation algorithm. 50 behaviors were tested from 10 users - 10 side, 10 forward, 10 pi, 10 up and 10 down gestures. In this confusion matrix, the model predicted two forward gesture, of the ten actual side gestures, and of the ten forward gestures, it predicted that one was a up gesture and four were pi gestures. All correct predictions are located in the diagonal of the table, so it is easier to visually inspect the table for errors, as these are represented by any non-zero values outside the diagonal.

Table 4.1: ConfusionMatrix. Using LRA with normal equation.

		Predicted gesture				
		side	forward	pi	up	down
Actual gesture	side	8	2	0	0	0
	forward	0	5	4	1	0
	pi	0	2	8	0	0
	up	0	0	3	7	0
	down	0	0	0	2	8



(a) Demo 1 (b) Demo 2 (c) Demo 3

(d) Occluded (e) Hands Down (f) Hands Down

(g) Hands Up1 (h) Hands Up2 (i) Wave Gesture

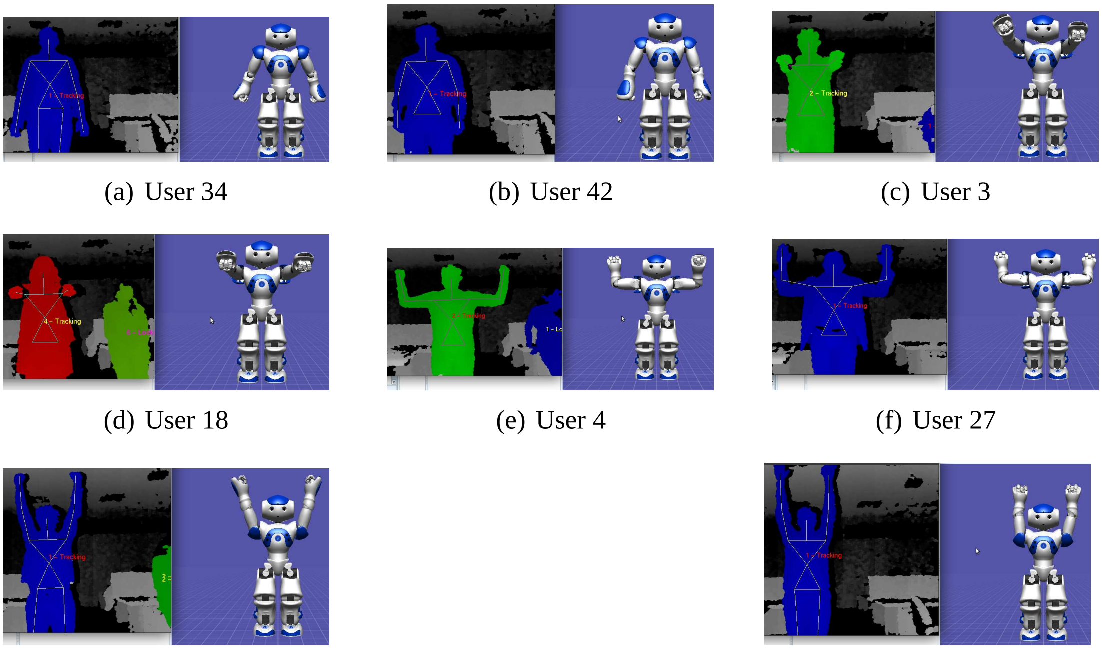
Figure 4.1: One-to-One imitation result.

4.0.2 Numerical Simulation Result

A screen shot from the one-to-one simulation system demo presentation can be seen in Figure 4.1.

A screen shot from the supervised learning system demo presentation can be seen in Figure 4.2. For Gradient Descent, out of 50 behaviors from 10 users, the system had a 70% accuracy. Similarly, for NE, out of 50 behaviors from 10 users, the system had a 72% accuracy.

A screen shot from the Decision-Based Rule system demo presentation can be seen in Figure 4.3. For DBR [23], after testing on 10 people with 5 different gestures, given a total of 50 gestures, 48 cases were correctly detected. Hence a 96% accuracy, which can be improved by getting a better threshold for the DBR Algorithm.

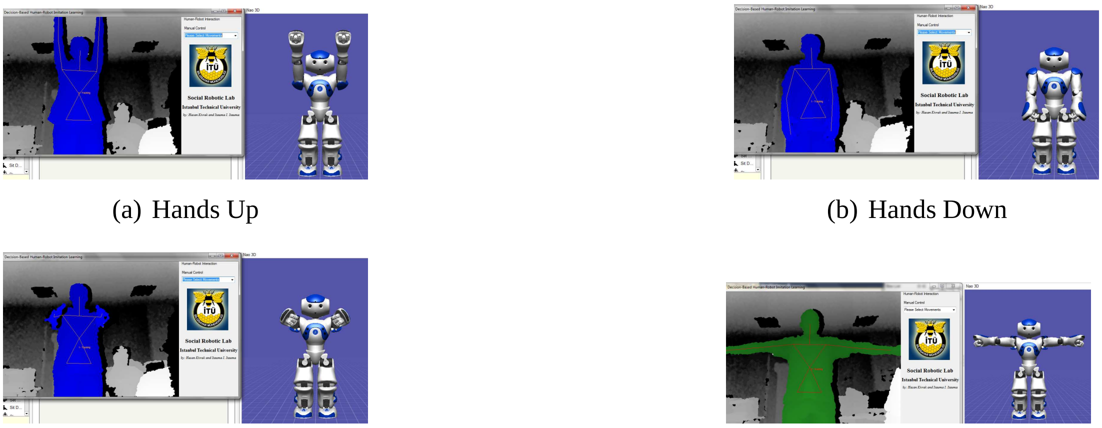


(a) User 34 (b) User 42 (c) User 3

(d) User 18 (e) User 4 (f) User 27

(g) User 10 (h) User 33

Figure 4.2: Supervised learning using normal equation imitation result.



(a) Hands Up (b) Hands Down

(c) Hands Forward (d) Hands Sideways

Figure 4.3: Decision based rule imitation result.

5. CONCLUSION AND FUTURE WORK

In this paper, we present the results of the different learning techniques used for imitation. Out of 50 behaviors from 10 users, the best result for linear regression was 72% accuracy and that of Decision Based Learning was 96%. Improving the feature set will lead to a better and faster imitation model. Decision Based Learning, a new imitation method for the humanoid robot was proposed.

The different methods in the literature are expensive in terms of computational cost, memory consumption and gesture recognition ability. [18] for example, is computationally expensive, unlike our approach using DBR. Our approach refers to the position of the X-, Y-, and Z-axis of the arm in 3D space in making a decision of the gesture.

Several avenues of future research would increase system usability. So far, our model is user dependent, in which the proposed Decision Based Rule algorithm threshold values have to be adjusted for each user. We propose combining several regression models to see if this will improve the imitation accuracy.

This study involved using RGB-D cameras like Kinect for Human motion capturing which is used for Robot Imitation and American Sign Language tutoring. The contribution of this study involved building different learning techniques that recognized different Human Gestures.

More studies are needed in order to obtain better objective and detailed data. Sometimes creating new features determine better model and we propose to try several polynomial regression methods. We also propose an experimental environment in ROS. Applying other classification methods than linear regression and logistic regression to learn a better hypothesis for imitation such as Neural Networks, Support Vector Machines will be implemented. We are working towards making the system more flexible and dependent on user’s state since the goal is towards a better social interaction between human and robot.

I have always thought of how Robotics can be used in enhancing Agricultural capability building with teleoperated Robots. With the resent drought in the Horn of Africa, it shows how little effort is made in Agricultural research. Having Robots explore the environment autonomously or teleoperated with mechanism of analyzing the soil will be a novel idea in Africa. This can lead to the discovery of seeds that are tolerant to droughts and resistant to disease base on the quality of soil in different terrains. Tertiary institutions do not have to have Robots to already commence this research. The students with the use of ROS and the Robot simulated environment can already start mapping how Robots can explore different regions using the terrain data. Then rather than waiting for a natural or man-made disaster, the Robots will be able to save life by preventing famine in the world.

REFERENCES

[1] **OpenNI User Guide**, 2010. `http://www.openni.org/documentation`.

[2] **E.Pot**, **J.Monceaux**, **R.Gelin**, **and B.Maisonnier**, 2009. Choregraphe: a Graphical Tool for Humanoid Robot Programming, Robot and Human Interactive Communication, Toyama, Japan.

[3] **Gouaillier D.**, **Hugel V.** et al, 2009. "Mechatronic design of NAO humanoid". IEEE Int. Conf. on Robotics and Automation, Kobe.

[4] **Kose-Bagci, H.**, **R. Yorganci**, **and Itauma I.I.**, 2011. "Humanoid Robot Assisted Interactive Sign Language Tutoring Game", 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO) Phuket Island, Thailand. **7-11 Dec**.

[5] **Kose-Bagci, H.**, **E. Ferrari**, **K. Dautenhahn**, **D. S. Syrdal**, **and C. L. Nehaniv**, 2009. “Effects of Embodiment and Gestures on Social Interaction in Drumming Games with a Humanoid Robot“, Special issue on Robot and Human Interactive Communication, Advanced Robotics **vol.24, no.14**.

[6] **Kose-Bagci, H.**, **K. Dautenhahn**, **D. S. Syrdal**, **and C. L. Nehaniv**, 2010. “Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot,” Connection Science **vol. 22, no. 2, pp. 103– 134**

[7] **Kose-Bagci, H.**, **R. Yorganci**, **and H. E. Algan**, 2011 accepted. “Evaluation of the Robot Sign Language Tutor using video-based studies”, 5th European Conference on Mobile Robots (ECMR11) **7-9 Sep**

[8] **Kose-Bagci, H.**, **and R. Yorganci**, 2011 accepted. “Tale of a robot: Humanoid Robot Assisted Sign Language Tutoring”, 11th IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia (HUMANOIDS 2011) **Oct. 26-28**

[9] **Kose-Bagci, H.**, **R. Yorganci**, **H. E. Algan**, **and D. S. Syrdal**, 2011. “Evaluation of the Robot Assisted Sign Language Tutoring using video-based studies”, SORO special issue on "Measuring Human-Robot Interaction. [accepted].

[10] **SL Tutoring Game**, `http://www.youtube.com/watch?v=SB_YJqMM-x8`, IEEE ROBIO 2011 [Last accessed: 14.01.2012].

[11] **Matthias Greuter**, **Michael Rosenfelder**, **Michael Blaich**, **and Oliver Bittel**, 2011. "Obstacle and Game Element Detection with the 3D-Sensor Kinect". Research and Education in Robotics - EUROBOT 2011.

[12] **Bandera J.P.**, **R. Marfil**, **A. Bandera**, **J.A. Rodriguez**, **L. Molina-Tanco**, **and** F. Sandoval, 2009. “Fast gesture recognition based on a two-level representation”, Pattern Recognition letters.

[13] **Aleotti J.**, **and S. Caselli**, 2006. “Robust trajectory learning and approximation for robot programming by demonstration”, Robotics and Autonomous Systems.

[14] **Ratliff, N.**, **Bagnell, J.A**, **and Zinkevich, M.**, 2006. “Maximum Margin Planning”, Twenty second International Conference on Machine Learning (ICML06).

[15] **Shon A.P.**, **Grochow K.**, **and Rao R.P.N**, 2005. Robotic Imitation from Human Motion Capture using Gaussian Processes”. In Humanoid Robots, 2005 5th IEEE-RAS International Conference on, pages 129-134.

[16] **Asfour T.**, **F. Gyarfas**, **P. Azad.**, **and R. Dillemann**, 2004. “Imitation Learning of Dual-Arm Manipulation Tasks in Humanoid Robots”, Institute for Computer Science and Engineering (CSE/IAIM).

[17] **Sylvian C.**, **Florent G.**, **and Aude B.**, 2005. “Goal-Directed Imitation in a Humanoid Robot”, International Conference on Robotics and Automation.

[18] **Ratliff, N.**, **Bagnell, J.A**, **and Chestnut, J.**, 2007. “Boosting Structured Prediction for Imitation Learning”, Robotics Institute. http://repository.cmu.edu/robotics/54.

[19] **Prime Sensor^{TM} NITE 1.3 Algorithms notes**, 2010. PrimeSense Inc. http://www.primesense.com.

[20] **One-to-one Imitation system**, http://www.youtube.com/playlist?list=PLFBDA2B2B2757358B, [Last accessed: 14.01.2012].

[21] **NAO robot**, http://www.aldebaran-robotics.com/en/Discover-NAO/Software/choregraphe.html, Aldebaran Robotics.

[22] **Vector Math for 3D Computer Graphics**, http://chortle.ccsu.edu/vectorlessons/vectorIndex.html, [Last accessed: 14.01.2012].

[23] **Decision-based Imitation**, http://www.youtube.com/watch?v=qK0F3VxcZXw, [last accessed: 09.02.2012].

[24] **Jonathan C. H.**, 2011. "How to Do Gesture Recognition With Kinect Using Hidden Markov Models", http://www.creativedistraction.com/demos/gesture-recognition-kinect-with-hidden-markov-models-hmms/, [Last accessed: 16.01.2012].

[25] **Triesh J.**, **and Malsburg C.**, 2002. Classification of hand postures against complex backgrounds using elastic graph. Image and Vision Computing.

[26] **Wong S-F**, **and Cipolla R.**, 2005. Real-time Interpretation of Hand Motions using a Sparse Bayesian Classifier on Motion Gradient Orientation Images". In: The 16th British Machine Vision Conference (BMVC'05), pp. 379-388.

[27] **Wah Ng C.**, **and Ranganath S.**, 2002. Real-time gesture recognition system and application. Image and Vision Computing.

[28] **M. Oliver**, **Rosario B.**, **and Pentland P.**, 2000. A Bayesian Computer Vision System for Modeling Human Interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[29] **R. Kjeldsen**, **and J. Kender**, 1996. Finding skin in color images.

[30] **J. Yang**, **and A. Waibel**, 1995. Tracking Human Faces in Real Time. Technical Report CMU–CS–95–210, Department of Computer Science, Carnegie Mellon University.

[31] **Eickeler, S.**, **Kosmala, A.**, **and Rigoll, G.**, 1998. Hidden Markov model based continuous online gesture recognition.

[32] **Kohler, M.R.J.**, 1997. System architecture and techniques for gesture recognition inunconstrained environments.

[33] **K. Schwerdt et J. L. Crowley**, 2000. "Roboust Face Tracking using Color", 4th IEEE International Conference on Automatic Face and Gesture Recognition", Grenoble, France, **March**.

[34] **J.Davis**, **and M.Shah.**, 1994. Recognizing hand gestures.

[35] **Nölker C.**, **and Ritter H.**, 1999. Visual Recognition of Hand Postures.

[36] **Starner T.**, 1995. Visual Recognition of American Sign Language Using Hidden Markov Models.

[37] **Freeman W.**, 1995. Orientation Histograms for Hand Gesture Recognition Export.

[38] **Persoon, E., Fu**, 1977. Shape discrimination using Fourier Descriptors.

[39] **Feng-Sheng C.**, **and Fu, Huang**, 2003. Hand gesture recognition using a real-time tracking method and hidden Markov models.

[40] **Color Space**, `http://en.wikipedia.org/wiki/Lab_color_space`, Lab color space - Wikipedia, the free encyclopedi. Wikipedia. [Online] Wikimedia. [Cited: 01 2011, 21.]

[41] **Mahalanobis distance**, `http://en.wikipedia.org/wiki/Mahalanobis_distance`, Mahalanobis distance - Wikipedia, the free encyclopedia. Wikipedia. [Online] [Cited: 01 2011, 21.]

[42] **Rafael C. Gonzalez**, **and Richard E. Woods**, 1992. Digital Image Processing (2nd ed.). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

APPENDICES

APPENDIX A.1 : Training set

APPENDIX A.1

Table A.1: Training set.

RSR	RSP	RER	REY	RH	LSR	LSP	LER	LEY	LH	Behavior
87	6	165	92	7	-86	7	159	-87	3	side
18	17	165	0	8	-9	10	171	0	2	forward
76	0	98	9	86	-78	-1	97	-12	85	piUp
27	81	155	-12	89	-33	81	151	9	92	Up
22	-85	161	-2	-90	-22	-82	150	5	-89	down
84	-11	168	79	4	-80	-21	160	-80	4	side
22	0	159	3	34	-11	0	156	7	41	forward
72	2	98	4	77	-87	3	95	1	77	piUp
24	83	169	6	89	-22	93	169	-3	86	Up
23	-88	161	-5	-88	-18	-87	160	5	-85	down
87	4	163	79	-4	-88	-7	174	-78	-4	side
13	5	169	0	0	-21	3	165	0	0	forward
67	-3	94	19	75	-72	-6	97	-22	76	piUp
32	84	168	11	90	-39	80	163	-13	86	Up
22	-88	153	3	-81	-21	-89	153	0	-83	down
51	-7	40	-40	-28	-2	-46	37	-4	53	side
5	-30	64	0	0	12	-48	37	-19	48	forward
5	-30	64	0	0	12	-48	37	-19	48	piUp
18	92	162	-6	88	-27	86	164	-10	89	Up
15	-71	161	2	-78	-10	-75	158	-3	-76	down
98	-7	168	93	0	-111	-7	172	-90	0	side
5	6	173	-3	6	-11	5	174	-4	5	forward
82	17	102	-1	87	-85	16	98	5	87	piUp
16	92	162	-14	93	-35	85	159	0	93	Up
36	-98	158	7	-98	-35	-103	162	-13	-101	down
90	-4	170	86	5	-100	-9	175	-77	-2	side
-1	2	176	1	9	-2	-4	169	9	11	forward
83	-4	87	0	86	-78	-14	79	0	96	piUp
…	…	…	…	…	…	…	…	…	…	…

CURRICULUM VITAE

Name Surname: Itauma Isong, ITAUMA

Place and Date of Birth: Calabar, Nigeria, 11.02.1981

Adress: Bahçeköy Öğrenci Yurdu, Hacıosman, Sariyer, Istanbul

E-Mail: itauma@itu.edu.tr

B.Eng.: Electrical Engineering, 2004

M.Sc.: Computer Engineering, 2012

Professional Experience and Rewards:

List of Publications and Patents: Humanoid Robot Assisted Interactive Sign Language Tutoring Game

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- **Itauma I.I.**, Kose-Bagci H., 2011: Natural Robot Control Gesture Recognition Interface. *International Conference on Computing and Computer Vision (ICCCV 2011),* July 1-2, 2011, Kathmandu, Nepal. [Accepted]

- Kose-Bagci H., Yorganci R., **Itauma I.I.**, 2011: Humanoid Robot Assisted Interactive Sign Language Tutoring Game. *2011 IEEE International Conference on Robotics and Biomimetics*, pp. 2247-2249, Phuket Island, Thailand.

- **Itauma I.I.**, Kose-Bagci H., 2012: Gesture Imitation Learning for Human-Robot Interaction. *IK 2012 (Emotion and Aesthetics) conference*, March 16-23, 2012 Günne, Lake Möhne, Germany.

- **Itauma I.I.**, Kivrak H., Kose-Bagci H., 2012: Gesture Imitation Using Machine Learning Techniques. *SIU 2012, 20. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, April 18-20, 2012 Lykia World Oteli, Ölüdeniz, Fethiye, Muğla, Turkey.