

**T.C.  
BAHÇEŞEHİR ÜNİVERSİTESİ**

**HUMAN MOTION ANALYSIS AND ACTION  
RECOGNITION**

**Master Thesis**

**MURAT NURULLAHOĞLU**

**İSTANBUL, 2008**

**T.C.  
BAHÇEŞEHİR ÜNİVERSİTESİ**

**Institute of Science  
Computer Engineering Graduate Program**

**HUMAN MOTION ANALYSIS AND ACTION  
RECOGNITION**

**Master Thesis**

**Murat NURULLAHOĞLU**

**Supervisor: ASSC. PROF.DR. ADEM KARAHOCA**

**İSTANBUL, 2008**

**T.C**  
**BAHÇEŞEHİR ÜNİVERSİTESİ**  
**Graduate School in Sciences**  
**Computer Engineering Graduate Program**

Name of the thesis: HUMAN MOTION ANALYSIS AND ACTION RECOGNITION

Name/Last Name of the Student: Murat Nurullohoğlu

Date of Thesis Defense:

The thesis has been approved by the Institute of Graduate School in Sciences.

Prof. Dr. Erol Sezer  
Director

Signature

I certify that this thesis meets all the requirements as a thesis for the degree of Master of Science.

Assc. Prof. Dr. Adem Karahoca  
Program Coordinator

Signature

This is to certify that we have read this thesis and that we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Science.

Examining Committee Members

Signature

Title Name and Surname

Thesis Supervisor

Assc.Prof.Dr. Adem Karahoca

Member

Prof.Dr. Nizamettin Aydın

Member

Asst.Prof.Dr. M.Alper Tunga

## **ACKNOWLEDGEMENTS**

It is a pleasure to thank the many people who made this thesis possible, first and foremost, I would like to thank my mother, father, sister who have untiringly shown me guidance, patience and love. Without them, I never would have been able to get here.

I would like to express my deep gratitude both personally and professionally, to my advisor, Assc. Prof. Adem Karahoca, without whose guidance and encouragement the completion of this study would be impossible. I will remember every word of his valuable advice in my future studies.

I would also thank to the committee members. I would also like to express my thanks to the committee members Prof. Nizamettin Aydın and Asst. Prof. Dr. M. Alper Tunga for carefully reading the thesis manuscript and their valuable suggestions.

I would like to thank CSE faculty and research assistants for making my two years at Bahçeşehir University a great experience.

Finally, I would like to thank all my friends for their motivation.

Murat Nurullahođlu

# ABSTRACT

## HUMAN MOTION ANALYSIS AND ACTION RECOGNITION

Nurullohoğlu Murat

The Institute of Sciences, Computer Engineering Graduate Program

Supervisor: Assc. Prof.Dr. Adem Karahoca

August 2008, 62 pages

The goal of human activity recognition systems is to build a system that can automatically infer a range of predefined activities, such as running, handclapping, etc, from recorded video sequences. Such a computerized system would be of great use for a variety of applications ranging from video surveillance for security to human-machine interaction. Typically, pattern recognition is the key component of such a system, where the goal is to classify, or more specifically “*recognize*” the data, based on a priori knowledge or statistical information extracted from the patterns.

In this study, we presented a comparison of several well-known pattern recognition techniques for a human activity recognition system. Firstly, the database of short video sequences that each consisting of a set of predefined actions, i.e. running, jogging, boxing and handclapping, performed by different people at different environments are used. Then, using sophisticated image processing and computer vision algorithms, the segmentation of the human performing the action was achieved and a set of descriptive features were extracted for each activity performed. Motion History Images (MHI) were used to describe these activities in a qualitative way and computed Hu moments, a widely used and well-known feature set to describe 2D or 3D shapes, for further processing. Several feature extraction and classification methods are compared using different combinations and the results are analyzed. These methods are Principle Component Analysis and Linear Discriminant Analysis.

The features with Support Vector Machines and K Nearest Neighbours classifiers are tested. The best results are taken in Hu moments for KNN classifier, K=1 and PCA for KNN. The both accuracy is 95 percent. The worst results are taken in PCA for SVM. The accuracy is 80.6 percent.

Key words: Human activity recognition, motion history images (MHI), Hu moments, principal components analysis (PCA), linear discriminant analysis (LDA)

# ÖZET

## İNSAN HAREKETİ ANALİZİ VE HAREKET TANINMASI

Nurullohođlu, Murat

Fenn Bilimleri Enstitüsü, Bilgisayar Mühendisliđi Yüksek Lisan Programı

Danışman: Doç.Dr. Adem Karahoca

Ađustos 2008, 62 sayfa

İnsan hareketi tanıma sistemlerinin ana amacı, önceden tanımlanmış koşma, el çırpma gibi hareketleri otomatik olarak tanınmasını sağlamaktır. Güvenlik için izlemeden, insan-makine etkileşimine kadar geniş bir yelpazede bu sistemler kullanılmaktadır. Bu sistemleri en önemli birimi örüntü tanıma işlemleridir.

Bu çalışmada, insan hareketlerinin tanınması için, iyi bilinen birkaç tanıma algoritması karşılaştırılmıştır. Öncelikle, önceden tanımlanmış olan ve farklı insanlar tarafından gerçekleştirilen koşma, jogging, el çırpma ve boks yapma gibi hareketlerden bir veritabanı oluşturulmuştur. Bu hareketlerin hepsi için geçmiş hareket imgelerinden Hu moment'leri hesaplanmıştır ve birkaç özellik çıkarma algoritması ve tanıma algoritmalarıyla farklı kombinasyonlar kullanılarak test edilip sonuçları analiz edilmiştir.

Test etmek için Support Vector Machines (SVMs) ve K Nearest Neighbours (KNN) algoritmaları kullanılmıştır. En iyi sonucu, Hu moment ve PCA ile elde edilen özelliklerin KNN ile testi sonucunu yüzde 95 başarı ile bulunmuştur. En kötü sonucu ise PCA ile oluşturulan özelliklerin SVMs ile test edilmesinde yüzde 80.6 ile elde edilmiştir.

Anahtarlar Kelimeler: Human activity recognition, motion history images (MHI), Hu moments, principal components analysis (PCA), linear discriminant analysis (LDA)

# TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>ii</b>
<b>ÖZET</b> .....	<b>iii</b>
<b>TABLE OF CONTENTS</b> .....	<b>iv</b>
<b>LIST OF TABLES</b> .....	<b>v</b>
<b>LIST OF FIGURES</b> .....	<b>vi</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
<b>2. LITERATURE REVIEW</b> .....	<b>4</b>
2.1 GENERIC HUMAN MODEL RECOVERY .....	4
2.1.1 Three-Dimensional Movement Recognition.....	6
2.2 APPEARANCE-BASED MODELS.....	7
2.3 MOTION-BASED RECOGNITION.....	10
<b>3. CLASSIFICATION OF THE HUMAN MOTION</b> .....	<b>12</b>
3.1 BACKGROUND SUBTRACTION .....	13
3.2 WHAT IS MHI?.....	16
3.3 WHAT ARE HU MOMENTS?.....	18
3.4 FEATURE EXTRACTION .....	22
3.4.1 Principle Component Analysis (PCA).....	22
3.4.2 Linear Discriminant Analysis (LDA) .....	27
3.5 CLASSIFICATION .....	32
3.5.1 K Nearest Neighbours.....	32
3.5.2 Support Vector Machines .....	33
<b>4. TESTING FEATURE EXTRACTION AND CLASSIFICATION</b> .....	<b>37</b>
4.1 MHI .....	37
4.2 CLASSIFICATION RESULTS USING PCA .....	43
4.3 CLASSIFICATION RESULTS USING LDA .....	50
<b>5. CONCLUSION</b> .....	<b>55</b>
<b>REFERENCES</b> .....	<b>57</b>
<b>ÖZGEÇMİŞ</b> .....	<b>62</b>

## LIST OF TABLES

Table 4. 1 : The result of KNN classifier for each motion classes .....	41
Table 4. 2 : The result of SVM classifier for each motion classes .....	42
Table 4. 3 : Variance explained for all principal components.....	44
Table 4. 4: The result of KNN classifier for each motion classes .....	47
Table 4. 5 : The result of SVM classifier for each motion classes .....	48
Table 4. 6 : The result of KNN classifier for each motion classes .....	51
Table 4. 7 : The result of SVM classifier for each motion classes .....	52
Table 4. 8 : The result of classifiers with their features .....	53



## LIST OF FIGURES

Figure 3. 1 : Running man.....	12
Figure 3. 2 : Mhi of running man .....	13
Figure 3. 3 : Orginal frame.....	14
Figure 3. 4 : Background subtraction .....	14
Figure 3. 5 : After removing noisy pixels .....	15
Figure 3. 6 : Running man.....	17
Figure 3. 7 : Mhi of running man .....	17
Figure 3. 8 : Handclapping man .....	17
Figure 3. 9 : Mhi of handclapping man .....	17
Figure 3. 10 : Class separation .....	27
Figure 3. 11 : Maximum-margin hyper planes for a SVM.....	34
Figure 4. 1 : Mhi of boxing man .....	38
Figure 4. 2 : Mhi of handclapping man .....	39
Figure 4. 3 : Mhi of jogging man .....	39
Figure 4. 4 : Mhi of running man .....	40
Figure 4. 5 : ROC Curve for each classes with KNN classifier .....	41
Figure 4. 6 : ROC Curve for each classes with SVM classifier .....	43
Figure 4. 7 : An example to explain and illustrate the theory of 2D PCA.....	45
Figure 4. 8 : Zoom of the small area, where the samples of jogging.....	46
Figure 4. 9 : ROC Curve for each classes with KNN classifier .....	47
Figure 4. 10 : Curve for each classes with SVM classifier.....	49
Figure 4. 11 : Will be used for LDA .....	50
Figure 4. 12 : ROC Curve for each classes with KNN classifier using.....	52

Figure 4. 13 : ROC Curve for each classes with SVM classifier .....	53
Figure 4. 14 : ROC Curve for each classifier methods and their features .....	54

## LIST OF SYMBOLS/ABBREVIATIONS

Hidden Markov Model	: HHM
K-Nearest Neighbours	: KNN
Linear Discriminant Analysis	: LDA
Motion History Images	: MHI
Principal Components Analysis	: PCA
Support Vector Machines	: SVM

# 1. INTRODUCTION

Human Activity Recognition is to characterize the behaviour of a person. The ability of recognizing human and activities is a key importance to design a machine capable of interacting intelligently with a human-inhabited environment. Human Activity Recognition is used in many applications like security surveillance, medical diagnosis, choreography of dance and ballet, sign language translation, sports analysis and smart rooms (Gavrila & Davis 1996).

Event detection and human action recognition have gained more interest, among video processing community because they find various applications in automatic surveillance, monitoring systems (Liu & Chellappa (2007), Tan et al (2005), Venkatesh & Ramakrishnan (2004)). Video indexing and retrieval, robot motion, human-computer interaction and segmentation, and human motion analysis related issues studied from (Hernández et al. (2007), Han et al. (2007), Starner and Pentland (1995), Shavit and Jepson (1993)). Human motion analysis based on computer vision techniques has gained much attention to the modern researchers for its various practical applications. There are two fundamental questions in these applications. The first one is that how are things (pixels or cameras) moving? , the second one is “What is happening?”(Czyz and Liao (2006), Cao et al. (2004), Campbell et al. (2004), Bobick (1995), Lai et al. (2006)).

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are both well-known techniques for feature extraction or sometimes referred to as dimensionality reduction. (Urtasun et al. (2006), Ormoneit et al. (2005)) used PCA in human motion analysis. PCA provides a compact representation by projecting the data onto a feature space that captures the most representative features. In the new feature space, the first principal component is associated with the largest eigen values of the covariance matrix that corresponds to the direction where data have the largest variance (Smith 2002). Dick and Brooks (2003) used LDA in their human motion analysis work. LDA requires class labels for each data sample. It uses Fisher’s discriminant ratio to compute the

lower dimensional subspace that best discriminates data. Hu moments are statistical description of images using moment-based features, which have reasonable shape discrimination in a translation- and scale-invariant manner (Hu 1962).

Various approaches have been proposed for human motion analysis. We divided the prior work into three main classes. These are *generic model recovery (model based)*, *appearance-based models*, and *direct motion-based* recognition. The approach of the generic model recovery is to recover the pose of the person or object at each time instant using a three-dimensional model. The main advantage of model based approach is evidence gathering techniques used the whole image sequence before making a choice on the model fitting. On the other hand, the computational costs, due to the complex matching and searching are the disadvantage of implementing a model based approach. The disadvantage of implementing a model based approach is that the computational costs, due to the complex matching and searching that has to be performed are high (Dawson 2002). Because of the fact that increasing computing power, this can be seen as less of a disadvantages especially in non real-time applications and most algorithmic implementations help to reduce this costs.

Another model of human analysis is the appearance-based models. The approach of this model is to use only the two dimensional appearance of the action of two-dimensional static used in a multitude of frameworks where a human action is seen by consecutive images of two-dimensional instances/poses of the object. The major advantage of using appearance based methods is the simplicity of parameter computation. However, the matching may not be one-to-one, and the losing of precise spatial information is the main disadvantage of the appearance-based models (Du & Charbon 2007)

The last model is the direct motion-based recognition model. The approach of this model is to characterize the motion itself without reference to the underlying static poses of the body. The main advantage of this approach is the lower computation complexity and simpler implementation and also it doesn't require explicit feature and it determines the spatial transformation between sequences more accurately than the trajectory-to-trajectory approach does. On the contrary, since trajectories contain

explicit geometric information, the trajectory-to-trajectory approach better determines large spatial temporal misalignments, can align video sequences acquired by different sensors and is less affected by changes in background (Rao et al. 2003).

This thesis presents an appearance-based approach to the recognition of human movement. In this study, main objective is to benchmarking the different combinations of feature extraction and classification algorithms and compare and discuss the outcomes. We used Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA) algorithms as feature extractors to better represent the feature space by reducing the dimensionality. We made tests on the resulting feature spaces using a number of classifiers such as K-Nearest Neighbours (KNN) and Support Vector Machines (SVM) with a linear kernel.

The thesis is organized as follows. Section 2 reviews related work on human motion recognition. Section 3 presents our methods for pre-processing and classification. Section 4 reports the detailed results using the classification algorithms with different classifier. Finally, conclusions are drawn in Section 5.

## **2. LITERATURE REVIEW**

The number of approaches to recognizing motion, particularly human movement, has recently grown at a tremendous rate. As mentioned earlier, prior works can be categorized into three categories.

- A. Generic model recovery,
- B. Appearance-based models,
- C. Direct motion-based recognition.

### **2.1 GENERIC HUMAN MODEL RECOVERY**

One of the most common techniques to attain the three dimensional movement information is to recover the pose of the person or object at each time instant using a three-dimensional model. The model proposed is driven by attempting to minimize a residual measure between the projected models and object contours (e.g., edges of body in the image). The strong segmentation of foreground/background and also of the individual body parts are required to help the model organized process. Extending to the blurred sequence for these techniques is so difficult (Bobick & James 2001).

Wagg (2006) used model-based approach to automated extraction of walking people from video data, under indoor and outdoor capture conditions. He thought about Generalized Expectation-Maximization algorithm with the same solution approach. He used similar algorithm for local and global modelling strategies employed in a recursive process. Human shape and gait motion was used for extracting process. The extracted shape and motion information is applied to construct a gait signature and it is enough for recognition purposes.

Dawson's (2002) work is related with semi-automatic gait recognition using model based approach. The spatial and temporal metrics extracted from the model. Such that

people walking pattern's amplitude and variation in angles of the limp and transformed in eigenspace and he used Principle Component Analysis.

Rehg, Kanade (1995) described to the local tracking of articulated objects using a layered template representation of self-occlusions. Their approach is based on a kinematical model. This model was used to order the templates by their visibility to the camera. The effects of occlusion were captured to each template by attaching window functions. The experimental results are presented for 3D hand tracking under a significant amount of self-occlusion.

Goncalves (1995) presented a promoted 3D tracking of the human arm against a uniform background using a two cone arm model and a single camera. Their approach was about addressing the problem which estimated the human position and a human arm's motion using its behaviour. The modelled arm was two truncated right-circular cones connected using spherical joints. The using recursive estimator for arm position was proposed. So the estimator with error signals obtained by comparing the projected estimated arm position with that of the actual arm in the image is provided. The system is tested on a real image sequence.

Rohr (1994) extended their approach to the whole body as claimed. He used a full-body cylindrical model for tracking walking humans in natural scenes. The model-based approach for the recognition of pedestrians was introduced. They represent the human body by 3D-Model consisting cylinder. They use data from medical motion studies for modelling the movement of walking. Kalman filter was applied the estimation of model parameters in consecutive images. Single number was given all the poses of a walking action.

Gavrila and Davis (1996) extended this work to a full-body model for tracking human motion against a complex background. Their approach is about dividing humans and recognizing their activities to two major components;



- 1 - Body poses recovery and tracking
- 2 - Recognition of movement patterns

They present a vision system for the 3D model-based tracking of unconstrained human movement. They used image sequences acquired simultaneously from multiple views. The decomposition approach and best search technique for the high dimensional pose parameter space was used. They recover the 3D body pose at each time instant without the use of markers.

### **2.1.1 Three-Dimensional Movement Recognition**

Bottino et al. (2007) approach is based on 3D motion data. The motion captured in 3D by means of a model-based technique was divided into sequences. Each sequence was classified into one of the recognizable classes using PCA based method

Campbell and Bobick (1995) used three-dimensional data of human body for commercially system. Their approach is about identifying sets of constraints of movement expressed using body-centered coordinates such as joint angles and in force only during a particular movement. Their techniques were developed using defined by space curves for a representation of movements. There were axes of joint angles and torso location and attitude in the phase space. Their system used this representation for recognizing movements in unsegmented data.

Siskind's (1995) approach is like known object configurations. The input of his system developed consisted of line-drawings of a person, table, and ball. The moving object's positions, orientations, shapes, and sizes are known at all times. The approach uses support, contact, and attachment primitives and event logic to determine the actions of dropping, throwing, picking up, and putting down. The problem of recognizing actions were addressed by this approach when the precise configuration of the person and environment is known while the methods from the previous section concentrate on the recovery of the object pose.

## 2.2 APPEARANCE-BASED MODELS

"View-based Approach" is an addition to three - dimensional recognition approach. When an human setting out is seen by consecutive images of two-dimensional instances of the object, view-based approach attempts to use only the two dimensional appearance of the action of two-dimensional static used in a multitude of frameworks. In many methods, the normalized image of the object is necessary.

Silhouette techniques using multiple cameras and shape are used by Weinland at al. (2006) and new motion descriptors are presented. Motion history volumes which fuse action cues, as seen from different viewpoints and over short time periods, into a single three-dimensional representation sets up the main idea of these new motion descriptors.

Spatial-temporal templates are used by Dimitrijevic at al. (2006). The templates that they used represent the specific part on the walking cycle where the angle between the legs is greatest and the feet are on the ground. Six different camera views, excluding frontal and back view and different scales are covered by the templates.

Hand shapes commonly found in the American Sign Language that are image-based were worked by Rupe (2005). Describing the shape of the hand for region-based shape descriptor is completed by the generic Fourier descriptor as the single camera was for recognizing the hand shape. For recognition of the hand shapes Multi-class Support Vector Machine was used and to extract the arm from hand region before feature extracting he developed the wrist detection algorithm.

For object tracking and pose estimation, two different algorithms are used by Pradit Mittrapiyanuruk (2004). Two instances in a pair of images and performs 3D reconstruction followed by 3D pose estimation was the first method's principal part. An extension of the original Active Appearance Models takes part of the second method's matching algorithm. The principal of this extension is without any restriction on its geometry its allowing for the estimation of the 3D pose of any object.

Shape analysis and tracking are used by Hariataoglu et al. (2006). As for ordering the body parts on the silhouette boundary, shape analysis and tracking were used to construct models of people appearance.

For representation the action, action mostly hand motion in the actual greyscale images without background is used by Cui et al. (1995) and, Wilson and Bobick (1995). Due to obvious natural variations and different clothing, actions that include the appearance of the total body are not as visually consistent across different people as the hand looks are still rather similar over a wide range of people except the skin colour.

Cui's approach is a spatiotemporal event included two kinds of information, the object of interest and the movement of the object. The movement of the object was divided into two components: global and local motions. The global motion captured gross motion in terms of position and also the local motion represented deformation, orientation and gesture changes.

The approach of Wilson and Bobick (1995) is to allow for multiple models. The model is about a systematic way for describing a set of existing sensor data and a method for measuring how well it described new sensor data. The approach is based on multiple models simultaneously, where the type of models might be quite distinct. To characterize images in an image sequence was used eigenvector decomposition of sets of images, orientation histograms, peak temporal frequencies, tracked position of objects in the frame, and optic flow field summaries. In addition to this, Hidden Markov Model (HMM) algorithm was used for recognition motion.

Yamato *et al.* (1992) searched body silhouettes, and Akita (1984) employed body contours/edges, As opposed to using the actual raw greyscale image. The feature of image sequence of Yamato is based on bottom up approach with HMM. Silhouettes of human actions in a Hidden Markov Model (HMM) where silhouettes of background-subtracted images are vector quantized and used as input to the HMMs were used.

In Akita's work (1984), the use of edges and some simple two-dimensional body configuration knowledge (e.g., the arm is a protrusion out from the torso) are used to determine the body parts in a hierarchical manner (first, find legs, then head, arms, trunk) based on stability. The chaining local contour information found Individual parts. Although these two approaches decreased some of the variability between people, some problems seemed, such as the disappearance of movement that happens to be within the silhouetted region and also the varying amount of contour/edge information that arises when the background or clothing is high versus low frequency (as in most natural scenes). Also, the problem of examining the entire body, as opposed to only the desired regions, still exists, as it does in much of the three dimensional work.

### 2.3 MOTION-BASED RECOGNITION

Direct motion recognition (Dait & Qiant 2004; Dawson 2002) is another approach for recognition motion. This approach is to characterize the motion itself without using body's static poses. Both single "blob-like" entity and the tracking of predefined regions (head, leg, etc.) are two main approaches for the analysis of the body region using motion instead of structural features.

Efros (2003)' work was about computing optical flow measurements in a spatial-temporal recognize human activities in a nearest-neighbours framework volume using tennis and football sequences recognize human activities in a nearest-neighbours framework.

The work of Polana and Nelson (1994), Shavit and Jepson (1993) and, also, Little and Boyd (1995)'s works are similar about "blob-analysis". They recognize periodic walking motions using the repetitive motion. They gathered a feature vector of people in outdoor using the entire body and periodicity measurements. The nearest centroid algorithm was applied to recognize. By assuming a fixed height and velocity of each person, their approach was about being extended to tracking multiple people in simple cases. The approach of Shavit and Jepson's work was about using the gross overall motion of the person. The body is coarsely modelled as an ellipsoid. Optical flow measurements were used to help create a phase portrait for the system, which was then analyzed for the force, rotation, and strain dynamics. Little and Boyd (1995) was used similar approach. Their system was about recognizing people walking by analyzing the motion associated with two ellipsoids fit to the body. One ellipsoid was for fitting motion region and the other ellipsoid was about fitting motion magnitudes as weighting factors. The relative phase of various measures like centroid movement, weighted centroid movement, torque for each of the ellipses over time characterizes the gait of several people.

The region-based motion properties were used by some researches. They focused on motion for facial expression such as motion of the mouth, eyebrows, and eyes (Dait & Qiant (2004), Black & Yacoob (1995), Dawson (2002)). The aim of these researches is to is to recognize human facial expressions. These approaches characterize the expressions using the underlying motion properties rather than represent the action as a sequence of poses or configurations.

Yacoob and Davis (1996) and, also Black and Yacoob (1995) used optical flow measurements to help track predefined polygonal patches placed on interest regions. The parameterization and location relative to the face of each patch was given a priori. According to positive or negative intervals, they described qualitatively the temporal trajectories of the motion parameters. They used these qualitative labels in rule based, temporal model for recognition to determine expressions.

Ju *et al.* (1996) have extended this work with faces to include tracking the legs of a person walking. Despite of the simple, independent patches used for faces, an articulated three-patch model was needed for tracking the legs. Many problems appeared like large motions, occlusions, and shadows and it made motion estimation in that situation more challenging than for the facial case.

Our approach is such a (Bobick & Davis 1996) for recognizing not part of body and also whole body movements. It is an attempt to generalize the face patch tracking technique. The basic strategy to hypothesize movements is to use the overall shape of the motion. Essa and Pentland (1997) was used optical flow for physically-based model of the face to estimate muscle activation on a detailed. Similarity measure expressions of the typical patterns of muscle activation were classified by recognition approach. Another recognition method matched motion energy templates derived from the muscle activations. The activity sequence was compressed by these templates. In our work we developed similar templates, but these templates were used to incorporate the temporal motion characteristics.

### 3. CLASSIFICATION of the HUMAN MOTION

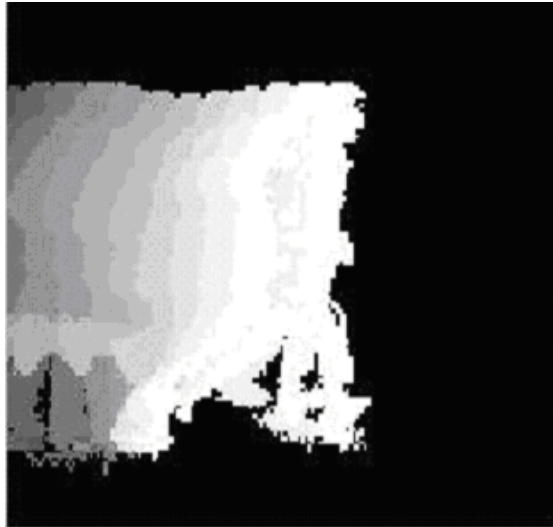
In this chapter, the methods used in this thesis were presented. Our goal is to construct an abstract, view-specific representation of motion, where the background is assumed as static.

The main idea is creating a vector-image to be able to match it against stored representations of known movements. Therefore, we first segmented the action from the background image using background subtraction method. Then we extracted Motion history images (MHI) from segmented video sequences that allow representing the motion in multiple frames in a more powerful way. A sample video sequence consisting of ten frames is shown in Figure.3.1 (a) with MHI in Figure3.1 and 3.2, respectively.



Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 3. 1 : Running Man**



Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 3. 2 : Mhi of running man**

In order to evaluate the accuracy of our system, we divided our data into two as training and test sets, where we have 80 samples for training and 80 samples for testing, for each action. We used PCA and LDA as feature extraction methods, and SVM and KNN as classifiers and tested the accuracy of the system for different combinations using a feature extractor and a classifier.

### **3.1 BACKGROUND SUBTRACTION**

One of the basic methods for motion detection in video sequences is background subtraction. In this method, corresponding pixels of two frames is compared, and their differences is taken and the pixels which their value more than a certain threshold are changes pixels (McIvor 2000). Moving objects are those whose pixel values have changed significantly.

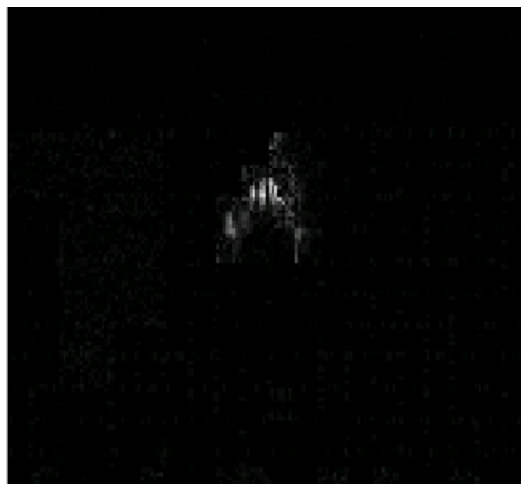
Therefore, after comparing the corresponding pixels, all connected components (objects) that have fewer than  $P$  pixels, where  $P$  is the threshold value determined experimentally as 15 are removed from the images. Because it was assumed that these are the pixels that don't belong to the motion.





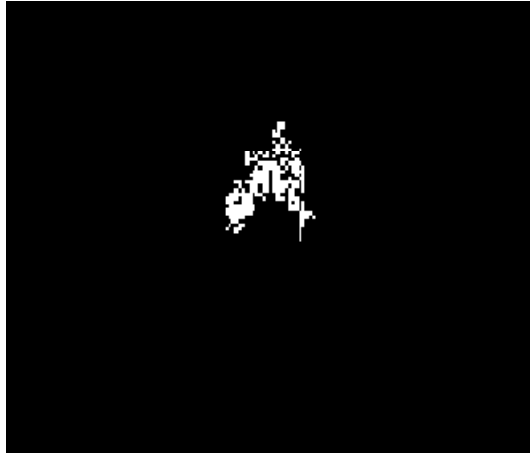
Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 3.3 : Original frame**



Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 3.4 : Background subtraction**



Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 3. 5 : After removing noisy pixels**

To make this technique suitable for real-time applications, it requires small memory space (one frame) and a small number of operations per pixel (integer subtraction), although this technique is not reliable for recognizing the moving object

$$B[x, y] = \begin{cases} 1 & f|I[x, y, t + 1] - R[x, y, t]| \\ 0 & otherwise \end{cases} \quad (3.1)$$

$I[x, y, t+1]$  is values of  $t+1$  image frame.  $R[x, y, t]$  is values of  $t$  frames

Although the background is static, the motion region can be isolated from the background by first capturing the background image and then subtracting it from each subsequent frame. While subtracting each subsequent frames, we removed small sized objects after performing connected components labelling followed by a threshold based on area. Pixel groups that are not connected to a component more than 15 pixels are removed to segment the motion from video sequences in the best way. In the end of this step we had subsequent frames to constitute MHI.

### 3.2 WHAT IS MHI?

The image that obtained by squashing the image-time volume onto a single image is named as MHI. The intensity values in the MHI shows the time of the pixels which last motion happened or object presence. It represents motion the image is moving  $H_t$ , pixel intensity is a function of the temporal history of motion at that point. MHI's were calculated for 0-t time period for each motion, after motion was extracted from video frames. An advantage of MHI is that although it is a representation of the history of pixel-level changes, only one previous frame needs to be stored. (Bobick & James 2001)

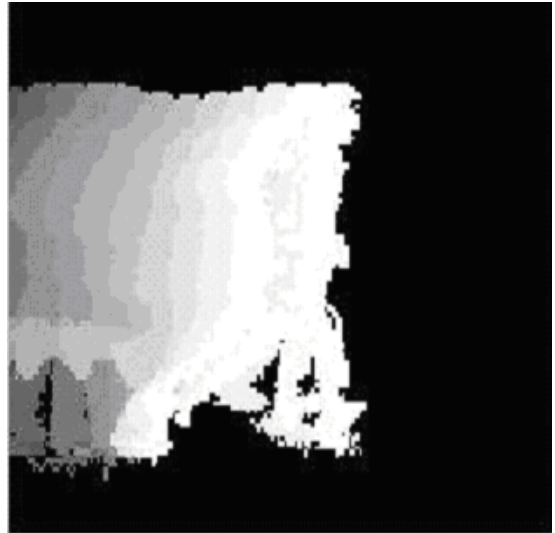
For the results presented here, we use a simple replacement and decay operator:

$$H_t = \begin{cases} T & \text{if } D(x, y) = 1 \\ \max(0, H_t(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad (3.2)$$

The result is a scalar-values image. More recently changed pixels are shown brighter. Examples for MHIs are presented in Figure 3.6, 3.7, 3.7 and, 3.9



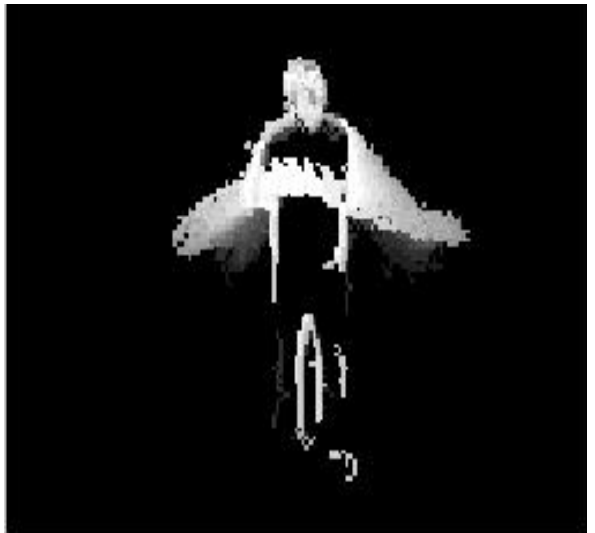
Source: Schuldt C., Laptev I., Caputo B., 2004  
**Figure 3. 6 : Running man**



Source: Schuldt C., Laptev I., Caputo B., 2004  
**Figure 3. 7 : Mhi of running man**



Source: Schuldt C., Laptev I., Caputo B., 2004  
**Figure 3. 8 : Handclapping man**



Source: Schuldt C., Laptev I., Caputo B., 2004  
**Figure 3. 9 : Mhi of handclapping man**

### 3.3 WHAT ARE HU MOMENTS?

The use of moments as invariant binary shape representations was first proposed by Hu (1962). He classified handwritten characters successfully by using this technique. Hu moments show that recognition schemes based on these invariants could be truly position, size and orientation independent, and also flexible enough to learn almost any set of patterns. The concept of moments is used extensively in classical mechanics and statistical theory. In addition to this that, central moments, size normalization, and principal axes are also used.

In the pattern recognition field, centroid and size normalization have been exploited for “pre-processing.” Orientation normalization has also been attempted. The Hu moments presented here achieves orientation independence without ambiguity by using either absolute or relative orthogonal moment invariants. To characterize each pattern for recognition, the method uses “moment invariants” or invariant moments (moments referred to a pair of uniquely determined principal axes)

Training examples of each movement running, handclapping, boxing, jogging is collected. Statistical descriptions of these images are computed by using moment-based features in a given set of MHIs for each movement combination. Our current choice is 7 Hu moments because of the fact that they are known to yield reasonable shape discrimination in a translation- and scale-invariant manner.

The two-dimensional moment,  $m_{pq}$  of order  $p + q$  of a density distribution function,  $f(x, y)$ , (image intensity) is defined as

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) d_x d_y \quad (3.3)$$

For  $p, q=0, 1, 2, \dots$

The central moments  $\mu_{pq}$  are defined as

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q p(x, y) d(x - \bar{x}) d(y - \bar{y}) \quad (3.4)$$

Where

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (3.5)$$

It is well-known that under the translation of coordinates, the central moments do not change, and are therefore invariants under translation. It is quite easy to express the central moments  $\mu_{pq}$  in terms of the ordinary moments  $m_{pq}$  for the first four orders, we have

$$\mu_{00} = m_{00} \equiv \mu \quad (3.6)$$

$$\mu_{10} = 0 \quad (3.7)$$

$$\mu_{01} = 0 \quad (3.8)$$

$$\mu_{20} = m_{20} - \mu \bar{x}^2 \quad (3.9)$$

$$\mu_{11} = m_{11} - \mu \bar{x} \bar{y} \quad (3.10)$$

$$\mu_{02} = m_{02} - \mu \bar{y}^2 \quad (3.11)$$

$$\mu_{30} = m_{30} - 3m_{20} \bar{\chi} + 2\mu \bar{x}^3 \quad (3.12)$$

$$\mu_{21} = m_{21} - m_{20} \bar{y} - 2m_{11} \bar{x} + 2\mu \bar{x} \bar{y}^2 \quad (3.13)$$

$$\mu_{12} = m_{12} - m_{02} \bar{\chi} - 2m_{11} \bar{y} + 2\mu \bar{x} \bar{y}^2 \quad (3.14)$$

$$\mu_{03} = m_{03} - 3m_{02} \bar{y} + 2\mu \bar{y}^3 \quad (3.15)$$

To achieve invariance with respect to orientation and, scale, we first normalize for scale defining  $\eta_{pq}$  :

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (3.16)$$

Where  $\gamma = (p+q)/2 + 1$  and  $\langle p+q \rangle = 2$  the first seven-orientation invariant Hu moments are defined as:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (3.17)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (3.18)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (3.19)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (3.20)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})^2 (\eta_{30} + \eta_{12}) \left( (\eta_{03} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right) + (3\eta_{21} - \eta_{03}) (\eta_{21} + \eta_{03}) \left( 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) \quad (3.21)$$

$$\phi_6 = (\eta_{20} + \eta_{02}) \left( (\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) + 4\eta_{11} (\eta_{30} + \eta_{12}) (\eta_{21} + \eta_{03}) \quad (3.22)$$

$$\phi_7 = (3\eta_{21} + \eta_{03}) (\eta_{30} + \eta_{12}) \left( (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right) - (\eta_{30} - 3\eta_{12}) (\eta_{21} + \eta_{03}) \left( 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) \quad (3.23)$$

These moments can be used for pattern identification independent of position, size, and orientation.



### **3.4 FEATURE EXTRACTION**

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

#### **3.4.1 Principle Component Analysis (PCA)**

One of the image classification techniques used for the face recognition is Principle Component Analysis (PCA). The most of descriptive features of an image are principal components. In this method, a new set of variables, that is called principal components, are generated. Each if this principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data here are an infinite number of ways to construct an orthogonal basis for several columns of data. The main advantage of PCA is compressing data by reducing the number of dimensions without much losing of data information. The second one is identification of groups of inter-related variables, to see how they are related to each other. On the other hand, the main disadvantage of PCA is that it is only able to searching for linear relationships in the data. When the data is clustered, it might be more useful to apply PCA. Furthermore, the other main disadvantage of PCA is to have high computational complexity for large data sets (Lie-Wei et al. 2005).

The first principal component is a single axis in space. Each observation on that axis results values form a new variables. And the variance of this variable is the maximum among all possible choices of the first axis.

The second principal component is another axis which is perpendicular to the first one. Observations on this axes generates another new variable. The variance of this variable gives the maximum among all possible choices of this axis.

The full set of principal components is as large as the original set of variables. But it is common place for the sum of the variances of the first few principal components to exceed 90% of the total variance of the original data. By examining plots of these few new variables, researchers often develop a deeper understanding of the driving forces that generated the original data.

Training procedure in any classification system is significant and can be beneficial. Using various algorithms training can be performed. Algorithm used in current image classifier system for training is principal component analysis, whose major emphasis is to locate and depict the principal features of the given sample image The Eigenvectors is a key capability used in PCA analysis algorithm. Eigenvectors are defined to be a related set of spatial characteristics that computer uses to recognize a specific motion type. PCA technique uses training and testing sets of images. Eigenvectors of the covariance matrix is computed from the training set of images. These eigenvectors represent the principal components of the training images. These eigenvectors are often ortho-normal to each other. In the context of clouds classification, these eigenvectors would form the motion space. They can be thought of as a set of features that together characterize the variation between motion images.

The description of PCA is below. The aim is to reduce data set  $X$  of dimension  $M$  to data set  $Y$  of dimension  $L$ . Put it differently, our goal is to achieve the values of matrix  $Y$ , which in other words Karhunen-Loeve transform (KLT) of matrix  $X$ .

$$Y = KLT(X) \tag{3.24}$$

We have comprising data set observations of M variables and we want to reduce each observations to describe with L variables  $L < M$ . The data form is  $x_1, \dots, x_N$ .  $x_N$  Represent a single grouped observation of the M variables.  $x_1, \dots, x_N$  Column vectors is placed to the matrix X of dimension M x N. After these operations, we calculate the empirical mean along dimension  $m=1, \dots, M$  and store mean values into mean vector u of dimensions  $M \times 1$ .

$$\mu[m] = \frac{1}{N} \sum_{n=1}^N X[m, n] \tag{3.25}$$

We subtract the mean vector  $\mu$  each column of the data matrix X and store data in the M x N matrix B

$$B = X - \mu \tag{3.26}$$

Next step is to find covariance matrix from the outer product of matrix B with itself.

$$C = E[B \otimes B] = E[B \cdot B^*] = \frac{1}{N} B \cdot B^* \tag{3.27}$$

E is the expected value operator,  $\otimes$  is the outer product operator, and  $*$  is the conjugate transpose operator. After calculated covariance matrix, we find the eigenvectors and eigenvalues of the covariance matrix.

$$V^{-1}CV = D \quad (3.28)$$

V is the matrix of eigenvectors which diagonalizes the covariance matrix C. D is the diagonal matrix of eigenvalues of C. Matrix D's form is M x M diagonal matrix, where

$$D[p, q] = \lambda_m \text{ For } p=q=m \quad (3.29)$$

$\lambda_m$  is the mth eigenvalues of the covariance matrix C, and

$$D[p, q] = 0 \text{ For } p \neq q \quad (3.30)$$

Matrix V contains M column vectors, that have M length, which represent the M eigenvectors of the covariance matrix C. Eigenvectors and eigenvalues might be considered paired, the mth eigenvalue corresponds to the mth eigenvector then we sort the columns of the eigenvalue matrix D in order of decreasing eigenvalue, because eigenvalue represent distribution of the source data's energy among each of the eigenvectors. The cumulative energy g is calculated.

$$g[m] = \sum_{q=1}^m D[p, q] \text{ For } p=q \text{ and } m=1 \dots M \quad (3.31)$$

We use g vector to find value for L. We want to choose threshold such as %90 etc.

$$g[m = L] \geq \%90 \quad (3.32)$$

After these operations we normalize the data using empirical standard deviation vector  $s$  the from square root of each element along the main diagonal of the covariance matrix  $C$

$$s = \{s[m]\} = \sqrt{C[p, q]} \quad \text{For } p=q=m=1 \dots M \quad (3.33)$$

$$Z = \frac{B}{s.h} \quad (\text{divide element by element}) \quad (3.34)$$

The last step is to project the data onto new feature space.

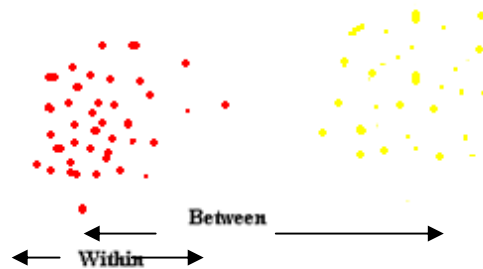
$$Y = W^* . Z = KLT\{X\} \quad (3.35)$$

We used another dimension reduction algorithms called LDA using Hu moments features.

### 3.4.2 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is the way for extracting features and dimension reduction of data for classification. LDA which maximizes the ratio of between-class variance to within-class variance in any particular data set finds the linear combination of features for separating two or more classes. LDA tries to divide region between classes. LDA wants to know the distribution of the feature data. The main advantage of LDA is that reduces the dimension of a data set to reveal its essential characteristics and also computing cost is very low compared with PCA and in addition to this, LDA shows powerful performance in the low dimension. However, LDA lose some discriminative information in the high dimensional space and furthermore, each class is thought to has a normal distributions with a common covariance matrix. (Balakrishnama & Ganapathiraju 1998)

The approach of LDA is about involving maximizing the ratio of overall variance to within class variance. Below Figure 3.5 shows good class separation



**Figure 3.10 : Class separation**

This approach makes the data sets transform to use one optimizing criterion; hence all data points irrespective of their class identity are transformed using this transform. In this type of LDA, each class is considered as a separate class against all other classes.

The sample sets below are analyzed using LDA with the mathematical operations. This approach was applied to a two-class problem for easily understanding.

Set1 is data set and set 2 is the test sets, they are be classified in the original space. The data sets are represented as a matrix form that consist of features

$$\text{set1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \dots & \dots \\ \dots & \dots \\ \dots & \dots \\ a_{m1} & a_{m2} \end{bmatrix} \quad \text{set2} = \begin{bmatrix} b_{11} & b_{21} \\ b_{21} & b_{22} \\ \dots & \dots \\ \dots & \dots \\ \dots & \dots \\ b_{m1} & b_{m2} \end{bmatrix} \quad (3.36)$$

Both the mean of each data and mean of entire data set must be calculated. V1 and V2 are mean of data set 1 and data set 2. V3 is also mean of the whole data. P1 and P2 are probabilities of the classes.

$$\tilde{\mathbf{o}}_3 = \mathbf{p}_1 \times \tilde{\mathbf{o}}_1 + \mathbf{p}_2 \times \tilde{\mathbf{o}}_2 \quad (3.37)$$

In LDA, within-class and between-class scatter are calculated for the class separability. Within-class scatter is measured by covariance of each of the classes

$$S_w = \sum_j p \times (\text{COV}_j) \quad (3.38)$$

Therefore, for the two-class problem,

$$S_w = 0.5 \times \text{COV}_1 + 0.5 \times \text{COV}_2 \quad (3.39)$$

One of the properties of the covariance matrices are symmetric.  $COV1$  and  $COV2$  represent the covariance of data set 1 and data set 2. Covariance matrix is calculated using the following equation.

$$COV_j = (x_j - \mu_j)(x_j - \mu_j)^T \quad (3.40)$$

The following equation calculates the between-class scatter.

$$S_b = \sum_j (\mu_j - \mu_3) \times (\mu_j - \mu_3)^T \quad (3.41)$$

The ratio of between-class scatter to the within-class scatter is the optimizing criterion of LDA. This criterion must be maximized to reach solution. This solution defines the axes of the transformed space. In addition to this, equations 3.42 and 3.41 are used to compute the optimizing criterion for the class-dependent transform. The optimizing factor is measured for class dependent type using below equations

$$criterion_j = inv(COV_j) \times S_b \quad (3.42)$$

For the class independent transform, the optimizing criterion is computed as

$$criterion = inv(S_w) \times S_b \quad (3.43)$$

An eigen vector that is a subspace of the vector space represents a 1-D invariant subspace and also this eigen vector is linearly independent. Linear dependency is represented by zero eigen value. Non-zero eigen values are considered and one of the



zero eigen values is neglected. The eigen vector matrix of the different criteria defined in Equations 3.42 and 3.43

We have L-1 non-zero eigen values for L-class problem. We transform the data sets using the single LDA transform or the class specific transforms whichever the case may be. The transforming the entire data set to one axis provides definite boundaries to classify the data. The decision region in the transformed space is a solid line separating the transformed data sets thus for the class dependent LDA,

$$transformed\_set\_j = transform\_j^t x\_set\_j \quad (3.44)$$

For the class independent LDA,

$$transformed\_set = transform\_spec^T x\_data\_set^T \quad (3.45)$$

The Euclidean distance of the test vectors from each class mean is used to classify and to transform to the test vectors. Both the original data sets and transformed data set are showed. Transformation provides a boundary for classification. In this example the classes were properly defined but in other cases, there might be overlap between classes. In some case, defined decision region is very difficult and in such cases importance of the transformation is considered. Transformation along largest eigen vector axis is the best transformation.

Euclidean distance or RMS distance is used to classify data points after LDA transforms. Equation 3.34 shows how Euclidean distance is computed.  $\mu_{ntrans}$  is the mean of the transformed data set, is n the class index and x is the test vector. N Euclidean distances are obtained for each test point for n classes.

$$dist\_n = (transform\_n\_spec^T) \times \chi - \mu_{ntrans} \quad (3.46)$$

The smallest Euclidean distance among the distances classifies the test vector as belonging to class.

All the things we made to now are for classification to create database to compare new motion. We used two algorithms to recognize human motion. One of these is Support Vector Machines (SVM), another is knn classifier.

### **3.5 CLASSIFICATION**

In the field of Machine Learning, the goal of classification is to group items that have similar feature values, into groups. A linear classifier achieves this by making a classification decision based on the value of the linear combination of the features.

In our thesis, we used K Nearest Neighbours and Support Vector Machines classifier for recognition and compared these techniques

#### **3.5.1 K Nearest Neighbours**

The  $k$ -nearest neighbours classifier is both the simplest of all machine learning algorithms and the simplest to implement. An object is classified by a closeness of its neighbours, with the object being assigned the class most common amongst its  $k$  nearest neighbours.  $K$  is a positive integer. If  $k$  is chosen one, then the object is belong to assigned the class of its nearest neighbours. It is helpful to choose  $k$  to be an odd number as this avoids difficulties with tied votes (Venkatesh & Ramakrishnan 2004)

The main advantage of KNN classifier is the simplest to implement and also K-nearest can predict both qualitative attributes that is the most frequent value among the  $k$  nearest neighbours and quantitative attributes that is the mean among the  $k$  nearest neighbours. In addition to this, it doesn't require creating a predictive model for each attribute with missing data and takes in consideration the correlation structure of the data. Furthermore it has robust performance.

On the other hand, KNN can have poor run-time performance if the training set is large, KNN is very sensitive to irrelevant or redundant features and also slow testing and scale (metric) dependent.

### 3.5.2 Support Vector Machines

Support vector machines (SVMs) being a sentence in connection standing supervised learning methods, those for classification and regression. They belong to a family of the generalized linear classifier and can be supposed as a special case of the Tikhonov regularization. A special characteristic from SVMs is that they lower the empirical classification error at the same time and maximize the geometrical side edge; therefore they are alias maximum margin classifiers. The main advantage of SVMs is the possibility to provide greater generalization performance using nonlinear classifiers in high dimensional spaces over a small training set. The input dimensionality of the problem is not related to the error of the SVMs; however the margin separates the data. This indicates SVM's good performance even with a large number of inputs. However, SVM has limitations that are speed, size both in training and test data. SVMs is the high algorithmic complexity and extensive memory requirements (Christopher & Burges 1998)

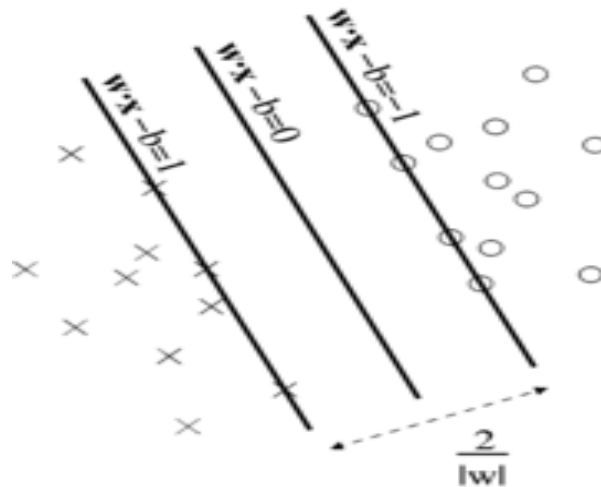
Support vector machines diagram entrance vectors to a higher measure area, in which a maximum separating hyper plane is designed. Two parallel hyper levels are designed on each side of the hyper level, which separates the data. The separating hyper level is the hyper level, which maximizes the distance between the two parallel hyper levels. An acceptance is formed that, which more largely the side edge or the distance between these parallel hyper levels, which is better the generalisation error of the classifier. A comparison of the SVM to other classifiers through van der Walt and Barnard pool of broadcasting corporations are accomplished frequently a machine-learning process is used by us to classify data. Each data point belonging to only one of two categories is showed by a P-dimensional vector (a list of the p numbers). We would like to see if we can separate them with a "p minus 1" measure hyper plane. This is a typical form of the linear classifiers. There are many linear classifiers, which could fulfil this characteristic. However we additionally interested, on to find out, if we can obtain maximum separation (margin) between the two categories. By this we mean that we select the hyper plane, so that the distance from the hyper plane to the next data point is

maximized. That is called that the next distance between one point in a separate hyper plane and one point in the other separate hyper plane is maximized.

Data points of the form is regarded

$$\{(x_1, c_1), (x_2, c_2), \dots, (x_n, c_n)\} \quad (3.47)$$

Where  $c_i$  is either 1 or -1, a constant showing the class where the point  $X_i$  belongs.  $X_i$  is a  $p$ -dimensional real vector which normalize  $[0, 1]$  or  $[-1, 1]$  values. To save against variables (attributes) with larger variance may control classification. We can called this as training data, that indicate the correct classification by means of the dividing (or separating) hyper plane, which takes the form



source: [http://en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine) [cited 28 June 2007]

**Figure 3. 11 : Maximum-margin hyper planes for a SVM**

Maximum-margin hyper planes for a SVM trained with samples from two classes. Samples along the hyper planes are called the support vectors.

$$w \cdot x - b = 0 \tag{3.48}$$

The vector  $w$  represents the perpendicular to the separating hyper plane. Adding the parameter  $b$  is to increase the margin. In its absence, it makes the hyper plane pass through the origin, restricting the solution.

When we look at in the maximum margin, we are interested in the support vectors. Parallel hyper planes are closest to these support vectors in either class. The below equations show these parallel hyper planes

$$\begin{aligned} w \cdot x - b &= 1, \\ w \cdot x - b &= -1. \end{aligned} \tag{3.49}$$

If we separate linearly the training data, we can select these hyper planes. There are no points between them and then try to maximize their distance. By using geometry, we minimize  $|w|$ , and find the distance between the hyper planes is  $2/|w|$ . We need to ensure that for all  $i$  either

$$\begin{aligned} w \cdot x_i - b &\geq 1 \quad \text{or} \\ w \cdot x_i - b &\leq -1, \end{aligned} \tag{3.50}$$

This can be rewritten as:

$$C_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n \tag{3.51}$$

Primal form, the problem now is to minimize  $|w|$  subject to the constraint (1).

Minimize  $(1/2)\|w\|^2$ , subject to

$$c_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \quad 1 \leq i \leq n \quad (3.52)$$

The factor of  $\frac{1}{2}$  is used for mathematical convenience

Dual Form, writing the classification rule in its dual form reveals that classification is only a function of the support vectors, i.e., the training data that lie on the margin. The dual of the SVM can be shown to be

$$\max \sum_{i=1}^n a_i - \sum_{i,j} a_i a_j c_i c_j \mathbf{x}_i^T \mathbf{x}_j \quad \text{Subject to } a_i \geq 0 \quad (3.53)$$

Where the  $a$  terms constitute a dual representation for the weight vector in terms of the training set:

$$\mathbf{w} = \sum_i a_i c_i \mathbf{x}_i \quad (3.54)$$

## 4. TESTING FEATURE EXTRACTION AND CLASSIFICATION

Methodologies have been implemented on a database consisting of 80 video sequences of four different motions, i.e. running, jogging, boxing and handclapping (Schuldt, Laptev and Caputo, 2004). Common characteristic of all these videos is that the background is static and there is no object moving, except for human motion in all videos. All videos were taken with a static camera at 25fps frame rate and stored using AVI file format in RGB without compression each of which are approximately four seconds on average duration with a spatial resolution of 160x120 pixels leading to an average size of 1.5 megabytes.

In these videos, each different motion has been performed by four different people. We separated the dataset into two groups as training and testing sets to perform the recognition tests. Training set consists of randomly selected 80 video sequences, 20 for each distinct motion. It has been used to construct the feature space and train the particular classifiers as mentioned in Sections 3.4 and 3.5, respectively. The remaining 80 test samples have been used for testing using the methodology explained in detail in Section 3

### 4.1 MHI

The intensity values in the MHI are indicative of the time at which that pixel last witnessed motion or object presence. There are a few problems to create MHI of videos. The big problem is the repeating motion in video. During the training phase, we measure the minimum and maximum duration that a movement may take,  $t_{\min}$  and  $t_{\max}$ . If the test motions are performed at varying speeds, we need to choose the right value for  $t$  for the computation of the MHI.

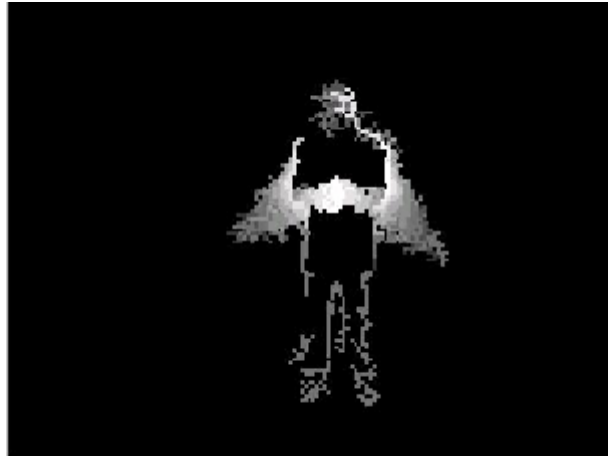


To choose right  $t$  is very important in handclapping videos since the human moving can repeat one more handclapping in one video. Therefore, we took first motion from the beginning frame to the end frame of first handclapping motion accordingly. Similar problems in jogging and running videos were handled in the same way. In such videos some people can run or jog in angle of camera between  $0-t$  periods, where the camera can capture the returning motion of some people in the same period, too. Similarly, we only took first complete cycle of the action to avoid such problems. Although capturing first motion of boxing is very hard, we took just 4 frames of each boxing videos and created MHI's using these frame periods. Figures 4.1, 4.2, 4.3 and 4.4 show some example MHIs constructed for boxing, hand clapping, jogging and running sequences, respectively.



Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 4. 1 : Mhi of boxing man**



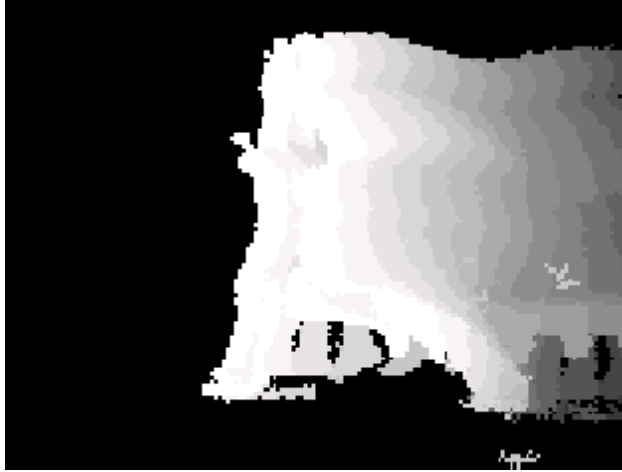
Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 4. 2 : Mhi of handclapping man**



Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 4. 3 : Mhi of jogging man**



Source: Schuldt C., Laptev I., Caputo B., 2004

**Figure 4. 4 : Mhi of running man**

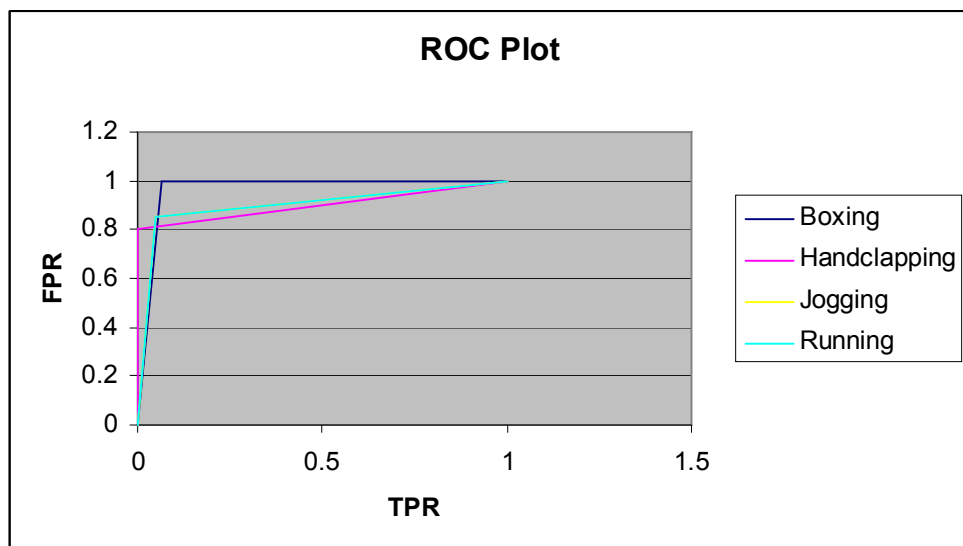
Once the MHI's are constructed, they are used to compute seven Hu moments. These moments used to characterize these motions quantitatively. Besides, since Hu moments are invariant to translation, rotation and scale, we are not required to apply pre-processing such as resizing or moving the motion to centre of image frame. As mentioned in Section 3, Hu moments of each motion have been calculated and these values have been stored to create the training database, where they have been used as input for the feature extraction step using PCA and LDA. In our training database, first 20 samples correspond to boxing, samples between 21 and 40 correspond to handclapping, 41 and 60 represent to jogging and the rest 20 correspond to running actions.

The final stage for human motion recognition is the classification step, where every test sample is classified using the parameters learned from the training data and the accuracy of the system is evaluated. Before applying any feature extraction step, first we performed an independent test to compare the classification accuracies of two classifiers: KNN and SVM, which are discussed in Sections 3.5.1 and 3.5.2. The independent test results for each motion have been presented in Tables 4.1 and 4.2.

**Table 4. 1 : The result of KNN classifier for each motion classes**

Class	Specificity	Sensitivity	Accuracy	Empiric ROC Area
Boxing	93.3 %	100 %	95 %	0.97
Handclapping	100 %	80 %	95 %	0.90
Jogging	95 %	85 %	92.5 %	0.90
Running	95 %	85 %	92.5 %	0.90

As can be seen from the results reported in Table 4.1 and Figure 4.5, the proposed approach was able to classify all four types of human activities with a remarkable overall accuracy of 93.8 percent using the KNN classifier. Classifying all boxing sequences correctly, the system was most sensitive on detecting samples of this action with 100 percent perfect accuracy. However, the specificity of this type of action is the lowest as expected. On average, both the overall specificity and sensitivity; therefore the overall classification accuracy of the system as indicated by the ROC area under the curve was significantly high, meaning the system was quite accurate identifying human actions of different types from video sequences.

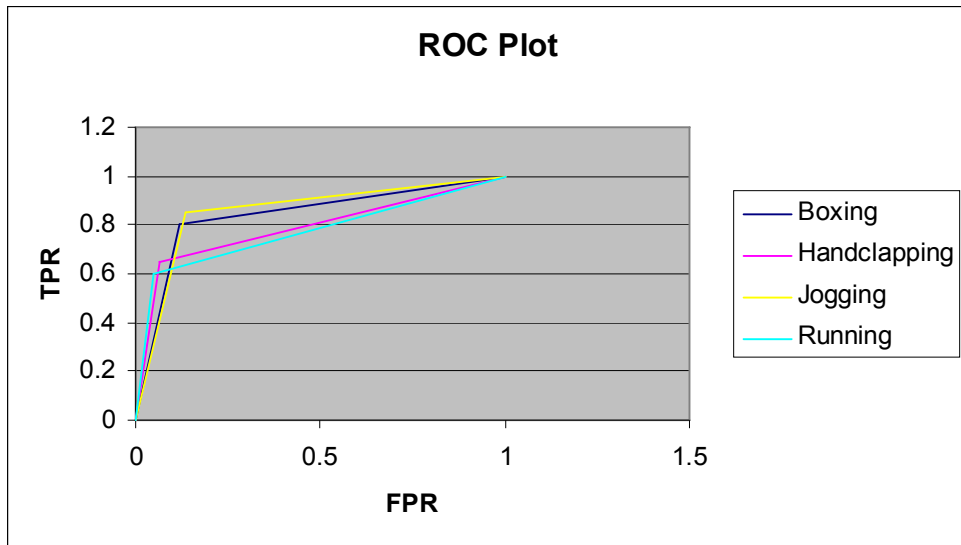


**Figure 4. 5 : ROC Curve for each classes with KNN classifier**

**Table 4. 2 : The result of SVM classifier for each motion classes**

<b>Class</b>	<b>Specificity</b>	<b>Sensitivity</b>	<b>Accuracy</b>	<b>Empiric ROC Area</b>
Boxing	88.3 %	80.0 %	86.3 %	0.84
Handclapping	93.4 %	65.0 %	86.4 %	0.79
Jogging	85.5 %	85.0 %	85.3 %	0.85
Running	95.0 %	60.0 %	86.3 %	0.78

On the other hand, the classification accuracies obtained using the SVM classifier, were not as accurate as the results obtained using a KNN classifier with a decrease around 86.3 percent, which is quite significant. Accordingly, the area under the curve for each different action was smaller when compared to using the KNN classifier. This is mostly due to the complex nature of the SVM classifier and probably the high dimensional feature space, which is even increased to a certain level where the class boundaries are linearly separable using a kernel function so that the SVM classifier can maximize the margin between the support vectors of each class. On the contrary, the KNN classifier only estimates the local densities in the feature space, leading to more accurate results. One other possible failure reason for the SVM classifier can be the insufficient number of training samples to construct an accurate mapping to the higher dimensional space, where classes are linearly separable.



**Figure 4. 6 : ROC Curve for each classes with SVM classifier**

To investigate the effects of the dimensionality of the feature space, we conducted similar experiments with a feature extraction step, where we compare two different well-known methods: PCA and LDA. In the following sections, we will discuss the results obtained by using different feature extraction steps.

#### **4.2 CLASSIFICATION RESULTS USING PCA**

In this section we analyzed the classification accuracy of human motions using PCA, which is one of the most popular feature extraction and dimension reduction algorithm and tested the output with KNN and SVM classifiers.

We applied PCA to perform a linear transformation of the original seven dimensional feature space constructed by computing the Hu moments of MHIs for each motion into a set of uncorrelated basis variables. PCA make the variance of linear combinations of original variables maximize by using eigenvalues and eigenvectors. Both the eigenvector that has the largest eigenvalue and the first principal component have the same direction and also the second largest eigenvalue is the second principal component, and so on.

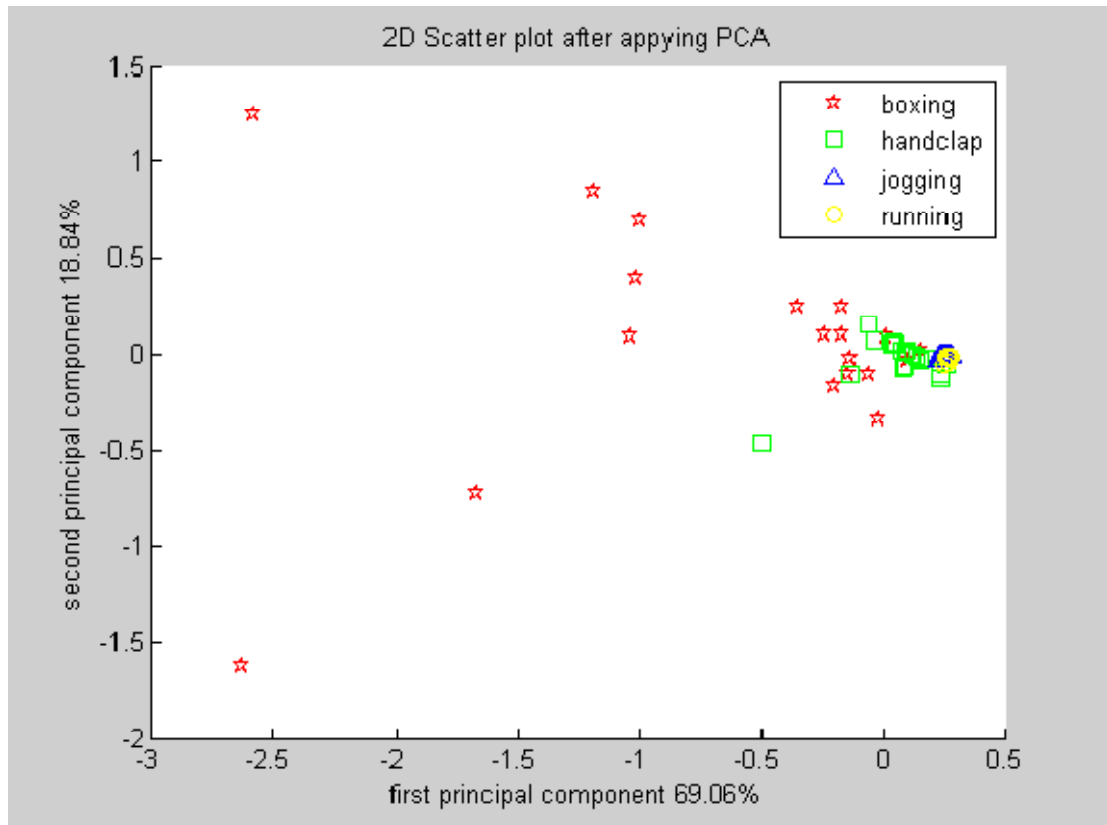
The diagram below helps to determine the optimal number of principal components to represent the whole data. We used the first n principle components that cover 90 percent of the total variance. In our training set, three principal components were selected as predictors summing up a variance of 93.6 percent in the seven original variables, as shown in Table 4.3.

**Table 4. 3 : Variance explained for all principal components**

Components	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	0.313	69.058	69.06
2	0.085	18.85	87.90
3	0.026	5.72	93.62
4	0.018	3.86	97.48
5	0.007	1.55	99.03
6	0.004	0.77	99.80
7	0.001	0.19	100.00

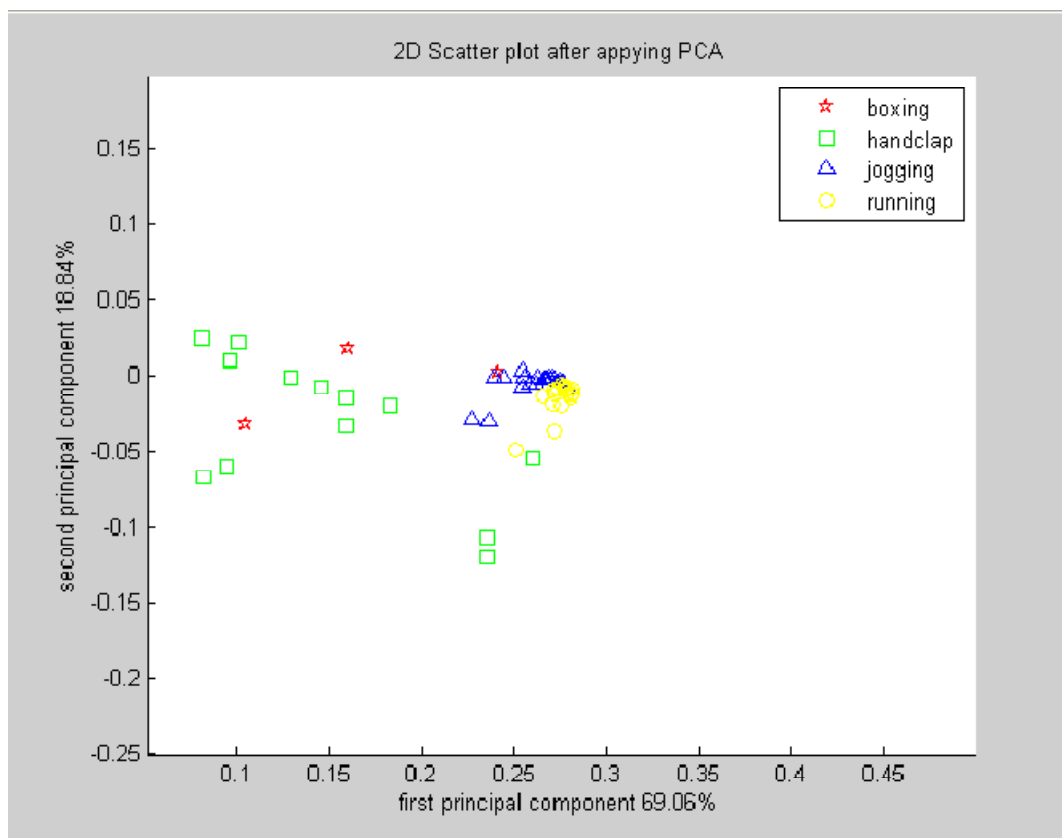
Once the transformation was computed, the training data is mapped onto this reduced dimensional feature space, where the inter-class variance is maximized. Figure 4.7 shows the 2D scatter plot of the transformed feature space, where X and Y axes is the first and second principal components, respectively. Samples corresponding to different classes are plotted using different shapes and colours as indicated in the legend of the plot. As can be seen from this figure, boxing and handclap motions can be discriminated easily where the samples of running and jogging motions are very close to each other making the discrimination harder. This result is quite intuitive since it is almost the same case in real life, where the motions of jogging and running are very similar performed almost identically with an only difference at speed. These results show that the proposed method is able to represent the human motions reasonably in this semantic context. However, we are still required to be able to discriminate these motions. To

better visualize the distribution of these two motions, Figure 4.8 shows zoom of the small area, where the samples of jogging and running are.



**Figure 4. 7 : An example to explain and illustrate the theory of 2D PCA**





**Figure 4. 8 : Zoom of the small area, where the samples of jogging and running are**

Principal Component Analysis (PCA) provides a compact representation by projecting the data onto a feature space that captures the most representative features. In the new feature space, the first principal component is associated with the largest eigen values of the covariance matrix that corresponds to the direction where data have the largest variance. Therefore, we can select the projection matrix composed of the first few eigen vectors associated with the largest eigen values. By doing so, we can reduce the dimensionality of our feature space without losing much representative information.

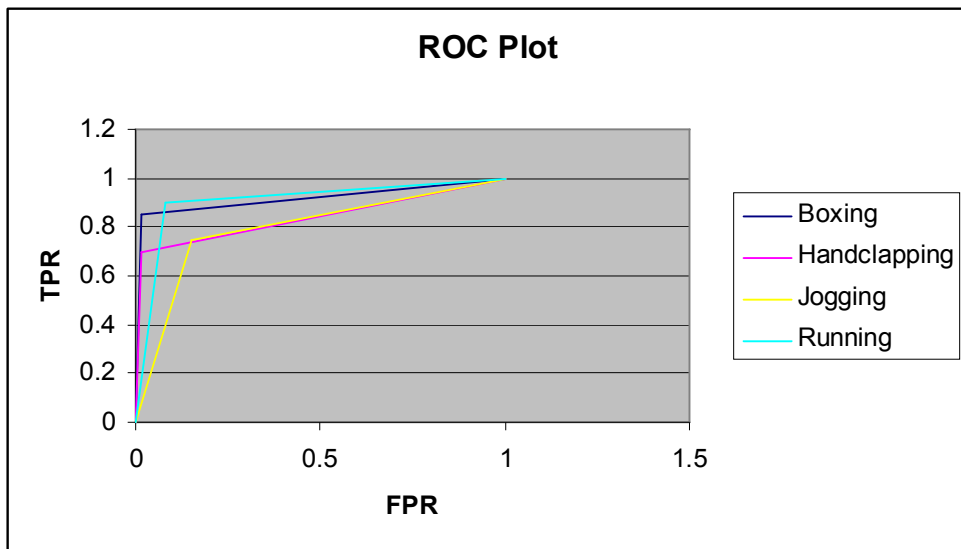
Although running and jogging classes were very similar to each other, the most representative features of these classes were seen very close to each others.

We tested PCA results with KNN classifier (k=1) and the results are given in Table 4.4.

**Table 4. 4: The result of KNN classifier for each motion classes**

Class	Specificity	Sensitivity	Accuracy	Empiric ROC Area
Boxing	98.3 %	85.0 %	95.0 %	0.917
Handclapping	98.8 %	70.0 %	91.3 %	0.842
Jogging	85.0 %	75.0 %	82.5 %	0.800
Running	91.7 %	90.0 %	91.3 %	0.908

When we tested PCA features with KNN classifier, the results are shown Table 4.4. We reached the best accuracy and specificity in boxing class with 85 percent and 98.3 percent. Although the boxing class has the best accuracy and specificity, we found better sensitivity in running class with 90 percent. Figure 4.9 shows us ROC curve for each motion with KNN classifier.



**Figure 4. 9 : ROC Curve for each classes with KNN classifier using PCA features**

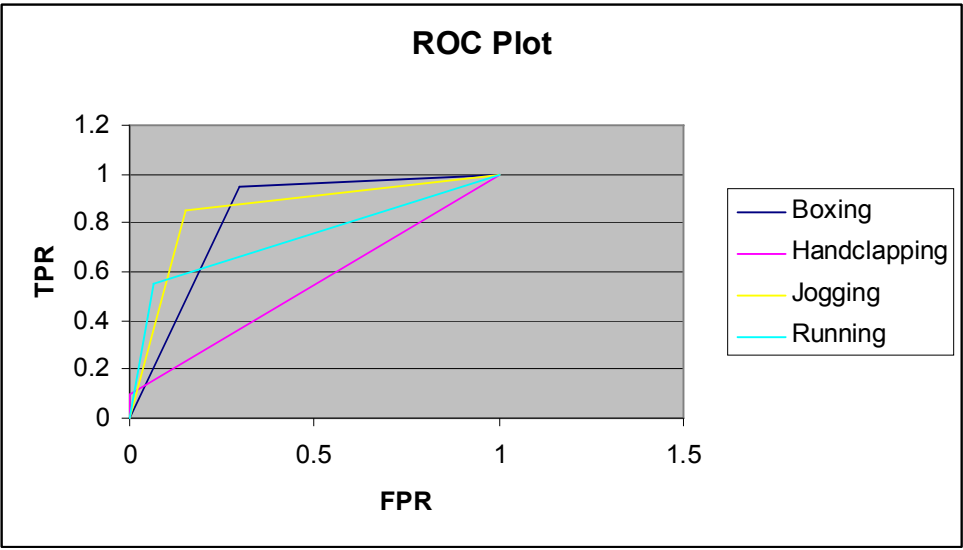
Similarly, Table 4.5 shows the classification accuracies after using SVM classifier instead of KNN. The results are quite interesting since the classification accuracy has drastically fallen to 76.3 percent at least for boxing. This can be explained with the concept of generalization. It is unlikely that the complex decision boundary in produced by SVM would provide good generalization, since it had been “*tuned*” to the particular training samples, rather than some underlying characteristics or true model of all these

motions that will have to be separated. KNN yielded better results because of its highly adaptive behaviour since it uses local information and does not rely on prior probabilities providing accurate estimates of the underlying true distribution of data.

**Table 4. 5 : The result of SVM classifier for each motion classes**

<b>Class</b>	<b>Specificity</b>	<b>Sensitivity</b>	<b>Accuracy</b>	<b>Empiric ROC Area</b>
Boxing	70.0 %	95.0 %	76.3 %	0.82
Handclapping	100.0 %	10.0 %	77.5 %	0.55
Jogging	85.0 %	85.0 %	85.0 %	0.85
Running	93.3 %	55.0 %	83.8 %	0.74

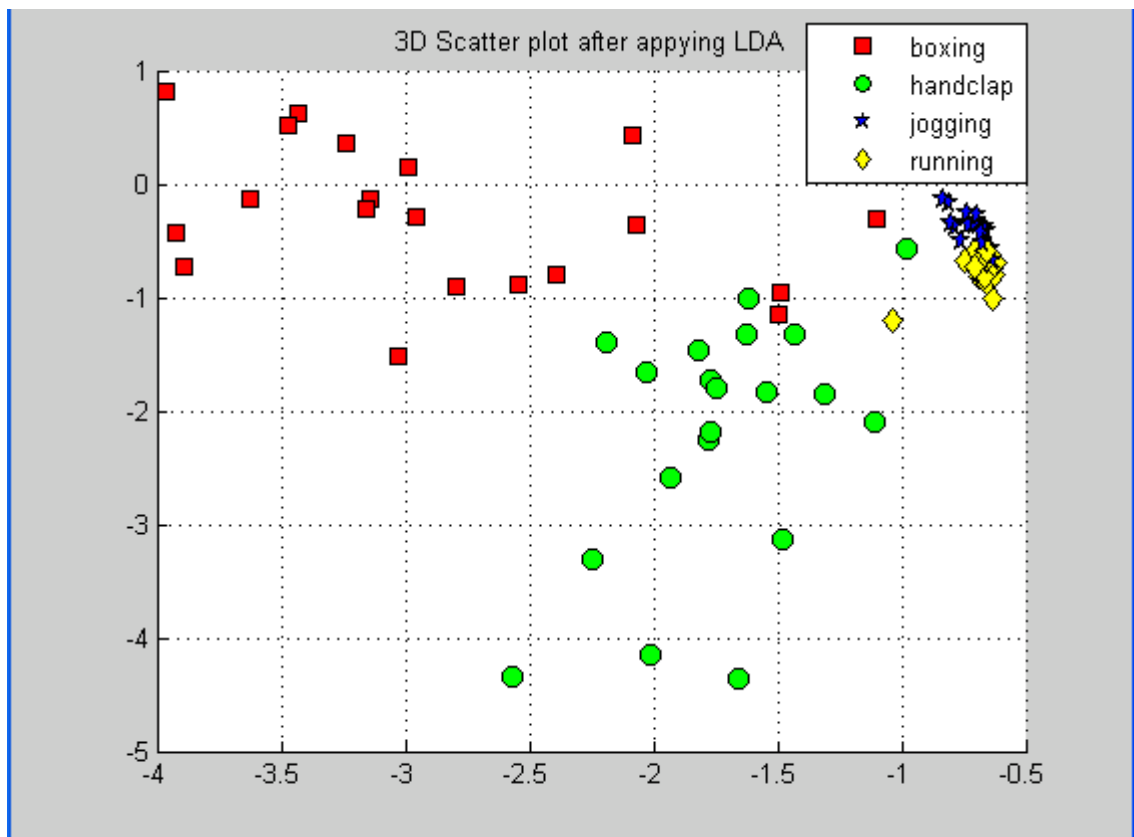
In these test works, the accuracy results decreased. We get the best accuracy rate in Jogging class in addition to this, the sensitivity and a specificity result of jogging class is not negative. We had very low sensitivity rate of handclapping class and running class, because of the fact that as seen in representation of PCA, the boxing class and handclapping class is very close and also running class and jogging class is close. Test videos of jogging class were recognized in running class and also handclapping videos were recognized in boxing class. Figure 4.10 shows us ROC curve for each classifier.



**Figure 4. 10 : Curve for each classes with SVM classifier using PCA features**

### 4.3 CLASSIFICATION RESULTS USING LDA

In this section, human motions were analyzed using LDA that reduces the dimensionality by constructing the most discriminant features based on inter- and intra-class distances. Again, outputs of LDA were tested the using KNN and SVM classifiers.



**Figure 4. 11 : Will be used for LDA**

Linear Discriminant Analysis (LDA) requires class labels for each data sample. It uses Fisher's discriminant ratio to compute the lower dimensional subspace that best discriminates data. LDA assumes that the samples in each class are distributed as a Gaussian distribution and all these distributions are defined by the same covariance matrix  $S_w$ , with different means (i.e., the data are homoscedastic). When this assumption holds, there is no loss in classification when reducing from  $p$  dimensions to  $C-1$ , where

$C$  is the number of classes. The goal is to find a projection where the Euclidean distance between classes are maximized, while the distance between samples in the same class is minimized. These measures are captured, respectively, using between-class scatter matrix  $S_b$  and within class scatter matrix  $S_w$  as defined in [1]. The resulting projection is given by the leading eigenvectors of matrix  $S_w^{-1}S_b$ . For a problem defined by  $C$  classes, this matrix has at most  $C-1$  nonzero eigen values, defining a feature subspace of  $C-1$  dimensions

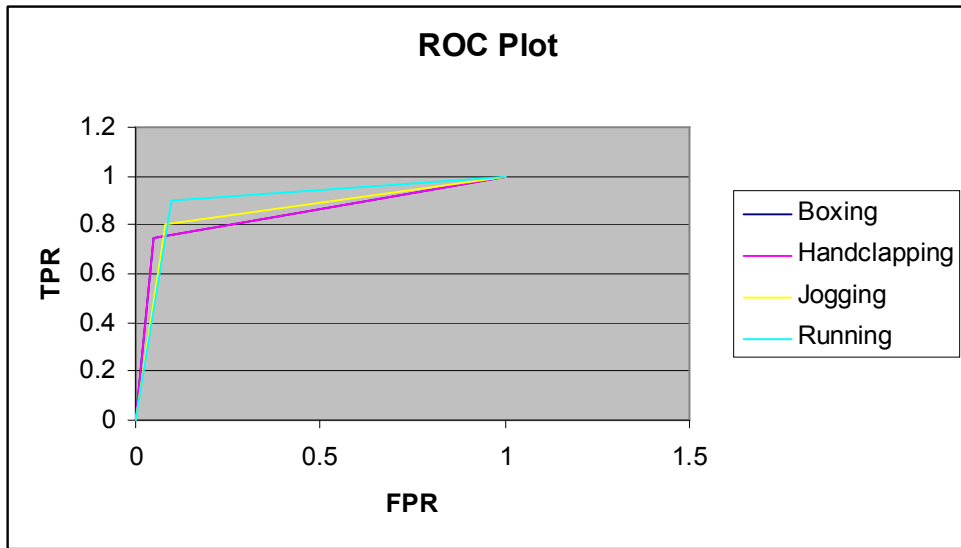
Firstly, we tested LDA features with KNN classifier,  $K=1$ . Mean of each classes were calculated and the distance of test videos to mean of classes were measured. The closest distance was taken.

The below Table 4.6 shows results applied by LDA and KNN classifier.

**Table 4. 6 : The result of KNN classifier for each motion classes**

<b>Class</b>	<b>Specificity</b>	<b>Sensitivity</b>	<b>Accuracy</b>	<b>Empiric ROC Area</b>
Boxing	95.0 %	75.0 %	90.0 %	0.85
Handclapping	95.0 %	75.0 %	90.0 %	0.85
Jogging	91.7 %	80.0 %	88.8 %	0.86
Running	90.0 %	90.0 %	90.0 %	0.90

We reached same and the highest accuracy in boxing, handclapping and running class. In this test, running class has the highest value for sensitivity with 90 percent. This rate fell into 75 percent in boxing and handclapping class. When we looked at specificity rate, the best value we had is in boxing and handclapping class with 95 percent.



**Figure 4. 12 : ROC Curve for each classes with KNN classifier using LDA features**

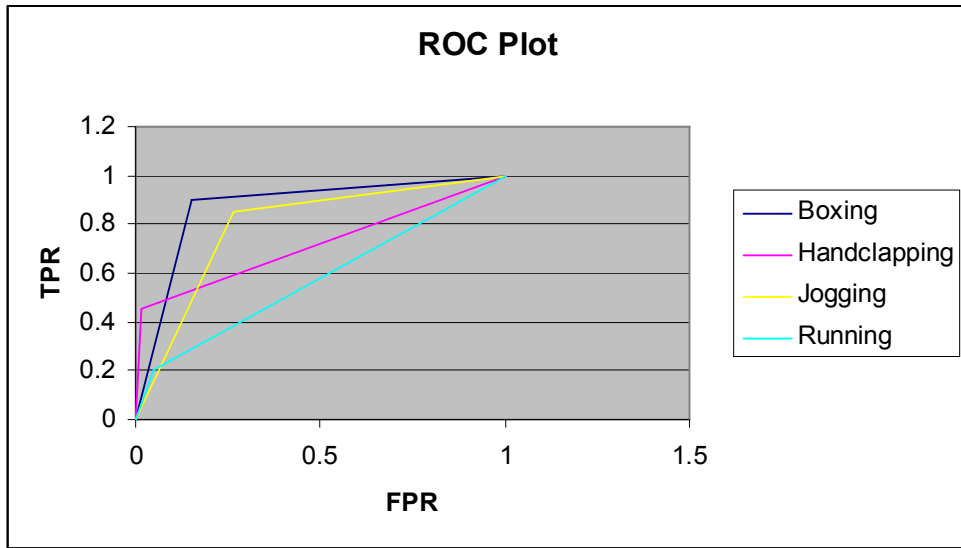
Secondly, we tested LDA features with SVM classifier.

The below Table 4.7 shows results applied by LDA and SVM classifier.

**Table 4. 7 : The result of SVM classifier for each motion classes**

Class	Specificity	Sensitivity	Accuracy	Empiric ROC Area
Boxing	85.0 %	90.0 %	86.3 %	0.88
Handclapping	98.3 %	45.0 %	85.0 %	0.72
Jogging	73.3 %	85.0 %	76.3%	0.79
Running	95.0 %	20.0 %	76.3%	0.58

As seen in representation of LDA, the boxing class and handclapping class is very close and also running class and jogging class is close. Because of this reason, the sensitivity of handclapping and running class fell into 45 percent and 20 percent. In these tests, we reached 86.3 percent accuracy in boxing class, and also boxing class has the highest sensitivity with 90 percent. Although handclapping has 45 percent sensitivity, it has the best rate for specificity with 98.3 percent. Figure 4.13 shows these results in ROC curve.



**Figure 4. 13 : ROC Curve for each classes with SVM classifier using LDA features**

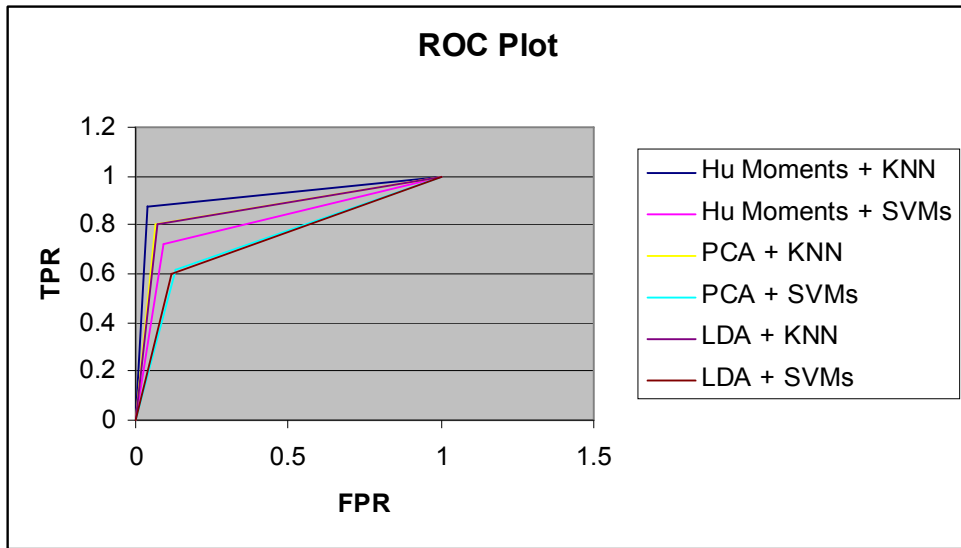
All in all, when the feature extraction methods were looked at with classifiers, Table 4.14 shows us the results for used classifier with its features

**Table 4. 8 : The result of classifiers with their features**

Methods	Specificity	Sensitivity	Accuracy	Empiric ROC Area
Hu Moments + KNN	95.8 %	87.5 %	93.8 %	0.92
Hu Moments + SVMs	90.8 %	72.5 %	86.3 %	0.82
PCA + KNN	93.3 %	80.0 %	90.0 %	0.87
PCA + SVMs	87.1%	61.3%	80.6%	0.74
LDA + KNN	92.9%	80.0%	89.7%	0.87
LDA + SVMs	87.9%	60.0%	80.9%	0.74

Table 4.8 and Figure 4.14 show us the best accuracy in Hu moments using KNN classifier, with 93.8 percent. Furthermore, the best sensitivity and specificity rate were taken using Hu moments with KNN classifier.





**Figure 4. 14 : ROC Curve for each classifier methods and their features**

## 5. CONCLUSION

The comprehensive comparison of different combinations of feature extraction and classification algorithms for a human motion recognition system was presented. The recognition framework is based on MHIs, an intuitive representation of motion using 2D, time-stamped templates followed by a feature construction step using HU moments, which are strong shape descriptors invariant to translation, rotation and scale. The MHI based representation allows the description of movements by their appearances. MHI templates extracted from image sequences provides a temporal template that captures the the motion properties at the corresponding spatial locations in the image sequence. This information -the presence of the motion in every image frame and their recency- leads to key clues to recognize particular actions against stored instances of similar known actions. The MHI based approach automatically performs the temporal segmentation and is invariant to linear changes in speed, and computationally. Its appearance-based approach compared to many efforts to recover the 3D reconstruction of the human form to interpret motion as mentioned in Section 2, is designed directly to support recognition of the motion itself. On the other hand, this view-based approach fails when there is missing motion information for example in case of any occlusion. However, the occlusion problem can be avoided at the segmentation step and independent from the MHI approach. PCA and LDA were used as feature extraction algorithms for dimensionality reduction. Finally, algorithms were tested these with KNN and SVM classifiers and discussed the varying accuracies obtained different combination of methods.

We determined four different motions including running, jogging, boxing and handclapping. The common speciality of these videos is to have static background. The dataset has been separated into two groups as training and testing sets. First step is to characterize each motion using Hu moments extracted from MHIs. Initially, we used these features to recognize the motions using KNN,K=1 and SVMs, where we reached the highest accuracy in boxing and handclapping class with 95 percent using KNN classifier and in boxing and running class with 86.3 percent using SVM classifier.

We applied PCA algorithms using Hu moments for input to classify motion features then we tested these features again with KNN, K=1 and SVMs. The accuracy of Knn is better than the accuracy of SVMs classifier. We reached 95 percent accuracy in boxing class with Knn. This rate fell into 85 percent in jogging class with SVMs classifier.

We also applied LDA algorithms for classification using Hu moments. The result of KNN, K=1 classifier for LDA features is 90 percent accuracy in boxing, handclapping and, running class and when we tested with SVMs, the highest accuracy rate we had is 86.3 percent in boxing class.

When we combined the results for all features extraction methods with their classifier, we reached the best accuracy rate in Hu moments using KNN classifier, with 93.8 percent. Furthermore, the best sensitivity and specificity rate were taken using Hu moments with KNN classifier.

## REFERENCES

### *Articles*

- Akita, K., 1984. Image sequence analysis of real world human motion. *Pattern Recognition*, vol. 17, no. 1, pp. 73-83.
- Black, M.J. & Yacoob, Y., 1995. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image Motion, *Proc. Int'l Conf. Computer Vision*, pp. 374-381.
- Bobick, A. F. & James, D. W., 2001. The recognition of human movement using temporal templates. IEEE Computer Society.
- Bobick, A., & Davis, J., 1996. An appearance-based representation of action, *Proc. Int'l Conf. Pattern Recognition*, pp. 307-312.
- Bottino, A., Simone, M., & Laurentini A., 2007. Recognizing human motion using eigen sequences, Politecnico di Torino Dipartimento di Automatica e Informatica.
- Campbell, L. & Bobick, A., 1995. Recognition of human body motion using phase space constraints. *Proc. Int'l Conf. Computer Vision*, pp. 624-630.
- Cao, D., Masoud, T.O., Boley, D., & Papanikolopoulos N., 2004. Online motion classification using support vector machines. Department of Computer Science and Engineering, University of Minnesota, Minneapolis.
- Cui, Y., Swets, D., & Weng, J., 1995. Learning-based hand sign recognition using Shoslif-m. *Proc. Int'l Conf. Computer Vision*, pp. 631-636.
- Dawson, M.R., 2002. Gait recognition. Department of Computing Imperial College of Science, Technology & Medicine, London.
- Dick, A. & Brooks, M., 2003. Issues in automated visual surveillance, School of Computer Science, University of Adelaide, and CRC for Sensor, Signal and Information Processing, Technology Park.
- Dimitrijevic, M., Lepetit, V., & Fua, P., 2006. Human body pose detection using Bayesian spatio-temporal templates. Computer Vision Laboratory.

- Du, H. & Charbon, E., 2007. 3D Hand model fitting for virtual keyboard system. Ecole Polytechnique Fédérale de Lausanne, Switzerland.
- Efros, A., Berg, A., Mori, G., & Malik J., 2003. Recognizing actions at a distance. *IEEE International Conference on Computer Vision*.
- Essa, I. & Pentland A., 1997. Coding, analysis, interpretation, and recognition of facial Expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757-763.
- Gavrila, D.M. & Davis, L.S., 1996. 3D Model-based tracking of humans in Action: A multiview approach. *Proc. Computer Vision and Pattern Recognition*, pp. 73-80.
- Goncalves, L., DiBernardo, E., Ursella, E., & Perona P., 1995. Monocular tracking of the human arm in 3D. *ICCV Proceedings of the Fifth International Conference on Computer Vision*.
- Liu, H. & Chellappa R., 2007. Markerless monocular tracking of articulated human motion, object video. University of Maryland.
- Han, M., Xu, W., & Gong, Y., 2007. Video foreground segmentation based on sequential feature clustering. NEC Laboratories America.
- Hariataoglu, I., Harwood, D., & Davis, L., 2000. W4: A Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8.
- Hernández, P.C., Czyz, J., Marqués, F., Marichal, X., & Macq, B, 2007. Bayesian approach for morphology-based 2-d human motion capture.
- Hu, M., 1962. Visual pattern recognition by moment invariants, *IRE Trans on Information Theory*, vol. IT-8.
- Ju, S., Black M., & Yacoob Y., 1996. Cardboard People: A parameterized model of articulated image motion. *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 38-44.
- Lai, Y., Czyz J., & Liao H.Y.M, 2006. Human motion recognition using clay representation of trajectories. Department of Computer Science, National Chiao-Tung University, Taiwan, Institute of Information Science, Academia Sinica, Taiwan
- Little, J. & Boyd, J., 1995. Describing motion for recognition. *Int'l Symp. Computer Vision*, pp. 235-240.
- McIvor, A. M., 2000. Background subtraction techniques. Reveal Ltd PO Box 128-221, Remuera, Auckland, New Zealand.

- Mittrapiyanuruk, P., DeSouza, N.G., & Kak C.A., 2004. Calculating the 3d-pose of rigid-objects using active appearance models, Robot Vision Laboratory, School of Electrical and Computer Engineering, Purdue University.
- Ogata, T., Rahman, M., Tan, J.K., & Ishikawa S, 2004. Real time human motion recognition based on a motion history image and an eigenspace, Kyushu Institute of Technology, Sensuicho, Tobata, Kitakyushu, Japan.
- Ormoneit, D., Black M.J., Hastie, T., & Kjellstro, H , 2005. Representing cyclic human motion using functional analysis ,Swedish Defence Research Agency, Sweden.
- Polana, R. & Nelson R., 1994. Low level recognition of human motion. *Proc. IEEE Workshop Non-Rigid and Articulated Motion*, pp. 77-82.
- Rao, C., Gritai, A, Shah, M, & Syeda-Mahmood T, 2003. View-invariant alignment and matching of video sequences, University of Central Florida, IBM Almaden Research Center, San Jose, CA.
- Rehg, J., & Kanade T., 1995. Model-based tracking of self-occluding articulated objects. *Proc. Int'l Conf. Computer Vision*, pp. 612-617.
- Rohr, K., 1994. Towards model-based recognition of human movements in image sequences. *CVGIP, Image Understanding*, vol. 59, no. 1, pp. 94-115.
- Rupe, C. J., 2005. Vision-based hand shape identification for sign language recognition, Department of Computer Engineering at Gleason College of Engineering Rochester Institute of Technology Rochester.
- Schuldt, C., & Laptev I., & Caputo B., 2004. Recognizing human actions: A local svm approach, *Proc. ICPR'04*.
- Shah, M. & Jain, R., 1997. Motion based recognition, Kluwer.
- Shavit, E. & Jepson A., 1993. Motion understanding using phase portraits. *Proc. IJCAI Workshop: Looking at People*.
- Siskind, J.M., 1995. Grounding language in perception, *ArtificialIntelligence Rev.*, vol. 8, pp. 371-391.
- Starner, T. & Pentland A., 1995. Real-time american sign language recognition from video using hidden Markov models. MIT Media Lab, TR-375.
- Tan, J.K., Houman, S., & Ishikawa, S., 2005. Human motion representation using eigenspaces, Department of Mechanical and Control Engineering , Kyushu Institute of Technology.

- Urtasun, R., Fleet, D. J., & Fua, P., 2006. Temporal motion models for monocular and multiview 3-d human body tracking. Computer Vision Laboratory. Swiss Federal Institute of Technology, Department of Computer Science University of Toronto
- Wilson, A. & Bobick A., 1995. Learning visual behavior for gesture analysis, Proc. IEEE Int'l. Symp. Computer Vision.
- Weinland, D., Ronfard, R., & Boyer E., 2006. Free viewpoint action recognition using motion history volumes. Perception-GRAVIR, INRIA Rhone-Alpes, 38334 Montbonnot Saint Martin, France.
- Wagg, D.K., 2006. Local and global models for articulated Motion analysis. University of Southampton,
- Venkatesh, B.R. & Ramakrishnan K.R., 2004. Recognition of human actions using motion history information extracted from the compressed video. *Image and Vision Computing*, Vol 22, No. 8, pp. 597-607
- Yamato, J., Ohya, J., & Ishii, K., 1992. Recognizing human action in time sequential images using hidden markov models. *Proc. Computer Vision and Pattern Recognition*, pp. 379-385.
- Yacoob, Y. & Davis, L., 1996. Recognizing human facial expressions form long image sequences using optical flow. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 636-642.
- Lie-Wei, W., Xiao W., Ming C., & Ju-Fu F., 2005. Is Two-Dimensional PCA a New Technique?. Center for information Sciences, Peking University

***Other Publications***

Balakrishnama, S. & Ganapathiraju A., 1998. Linear discriminant analysis – a brief tutorial, Institute for Signal and Information Processing. Department of Electrical ,Computer Engineering Mississippi State University.

Christopher, J.C., 1998. "A Tutorial on Support Vector Machines for Pattern Recognition". Data Mining and Knowledge Discovery 2:121 – 167

Smith, L., 2002. A tutorial on principal components analysis



## **ÖZGEÇMİŞ**

**Adı Soyadı** : Murat Nurullohođlu

**Sürekli Adresi** : Merkez Efendi Mah. Tercüman Sit. A4 Blok D75 Cevizlibađ-IST

**Dođum Yeri ve Yılı** : Istanbul-1980

**Yabancı Dili** : İngilizce

**İlk Öđretim** : Bađlar İlköđretim Okulu 1994

**Orta Öđretim** : Pertevniyal Lises 1998

**Lisans** : Yıldız Teknik Üniversitesi- Bilgisayar Mühendisliđi 2005

**Çalıřma Hayatı** : Tagsoft 2006 Ekim-.....

InspireIt 2005 Haziran-2006 Eylül

MngKargo 2005 nisan-2005 Haziran.