Motion Recognition Using Silhouette Images

A Dissertation submitted in partial fulfilment of the requirement for the Award of degree of

MASTER OF TECHNOLOGY

IN

SOFTWARE ENGINEERING

Submitted By

Pranay Bajaj

(2K15/SWE/13)

Under the esteemed guidance of

Dr. Anil Singh Parihar

Assistant Professor

Department of Computer Science & Engineering

Delhi Technological University

Bawana Road, Delhi-110042

2015-2017

CERTIFICATE

This is to certify that the thesis entitled “Motion Recognition Using Silhouette Images” submitted by Pranay Bajaj(2K15/SWE/13) to the Delhi Technological University, Delhi for the award of the degree of Master of Technology is an implementation of (Songtao Ding & Shiru Qu, 2016) carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Place: DTU, Delhi

Date: _________

Dr. Anil Singh Parihar
Assistant Professor
Department of Information Technology
Delhi Technological University, Delhi

ACKNOWLEDGEMENT

Firstly, I would like to express my gratitude towards my supervisor Dr. Anil Singh Parihar, *Assistant Professor, Department of Information Technology* for his able guidance, support and motivation throughout the time. It would not have been possible without the kind support and help of many individuals and Delhi Technological University. I would like to extend my sincere thanks to all of them.

I would like to express my gratitude and thanks to Dr. Rajni Jindal (*Head of Dept.)* for giving me such an opportunity to work on the project.

I would like to express my gratitude towards my parents & staff of Delhi Technological University for their kind co-operation and encouragement which helped me in completion of this project.

My thanks and appreciations also go to my friends and colleagues in developing the project and people who have willingly helped me out with their abilities.

Pranay Bajaj
2K15/SWE/13
Dept. of Information & Technology
Delhi Technological University

ABSTRACT

Human activity recognition has been a significant subject in computer vision because of its various uses or application, for example, video reconnaissance; human machine association and video recovery. One of the main troubles behind these applications is mechanically recognizing low-level activities and abnormal state exercises of concern. This Thesis shows a strategy of naturally distinguishing the activity finished by the human. The speeded up robust features (SURF) algorithm is made utilization of, for extraction of the key-focuses in both spatial and worldly area which is the augmentation of the SIFT finder/detector. The Spatio-Temporal Difference-of-Gaussian (STDoG) pyramid is basically fabricated which is additionally used to discover the maxima and the minima focuses which give the intrigue focuses. The key focuses are start in the x-y, x-t and y-t planes where x-y matches to the spatial plane, x-t and y-t planes take after to the transient spaces. Experimental result is lead on a video covering numerous activity recognition on KTH dataset.

Keywords: KTH, Spatio-Temporal Difference-of-Gaussian, speeded up robust features, NBKNN etc.

TABLE OF CONTENTS

TABLES OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Human action acknowledgment is imperative regions of computer vision inquire about today. The goal of human activity acknowledgment is to normally analyze persistent exercises from an obscure video (i.e. a progression of picture diagrams).

The ability to see complex human exercises from recordings enables the advancement of a couple of basic applications. Automated perception systems without trying to hide places like air terminals and cable car stations require area of odd and suspicious exercises rather than run of the mill exercises. Acknowledgment of human exercises moreover engages the nonstop checking of patients, children, and elderly individuals. The improvement of movement based human PC interfaces and vision-based keen conditions winds up recognizably possible too with an activity acknowledgment system.

There are distinctive sorts of human exercises. Dependent upon their versatile quality, we skillfully sort human exercises into four remarkable levels: signals, exercises, joint efforts, and social event exercises. Movements are fundamental advancements of a man's body part, and are the atomic fragments delineating the critical development of a man. `Stretching an arm' and `raising a leg' are extraordinary instances of movements. Exercises are single individual exercises that may be made out of different movements dealt with momentarily, for instance, `walking', `waving', and `punching'. Affiliations are human exercises that incorporate no less than two individuals and also addresses. For example, `two individuals doing combating' is a relationship between two individuals and `a singular taking a sack from another' is a human-dissent participation including two individuals and one inquiry. Finally, collect exercises are the exercises performed by connected social occasions made out of various individuals or conceivably addresses. `A social affair of individuals is strolling ', `a total having a meeting', and `two groups doing combating' are ordinary instances of them. The objective of this paper is to give a whole survey of best in class human activity acknowledgment methods of insight. We discuss distinctive sorts of procedures planned for the acknowledgment of different levels of exercises. The past review created by (J &

M, 2011) has secured a couple of crucial low-level parts for the perception of human development, for instance, following and body posture examination. Regardless, the development examination procedures themselves were insufficient to depict and remark on constant human exercises with complex structures, and by far most of philosophies in 1990s focused on the acknowledgment of signs and clear exercises. In this new overview, we concentrate on unusual state activity acknowledgment systems proposed for the examination of human exercises, coordinated efforts, and social affair exercises, discussing late research floats in real life acknowledgment.

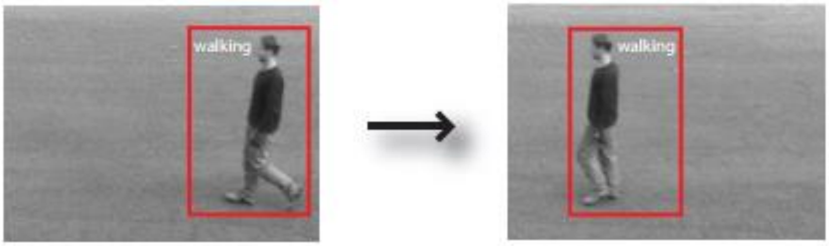


Figure 1.1: Single Action

For recognition of such an activity, we need to have clear background. So, the dataset must have a clear background. One thing that is to be noticed is that for the recognition of an action, there must be a clear background along with the clear identification of the human body and its pose. For identifying an activity clearly, there are three important things. The body posture must be identifiable. The poses need to be in a specific order. The speed of the body should be taken into account to clearly identify which pose it is. For instance, speed of the body is necessary to distinguish clearly between activities like walking, running and jogging. Also, strategies in view of spatio-transient formats generally focus on the posture of the human body, though techniques in light of dynamical models center their consideration regarding displaying the requesting of these stances in more prominent detail.

Perceiving single activities is a moderately more straightforward issue contrasted with complex activities; it is generally less demanding to secure preparing information for distinguishing single activities. Furthermore, current datasets accessible just manage static foundations where forefront human figures are effectively extractable for additionally handling.

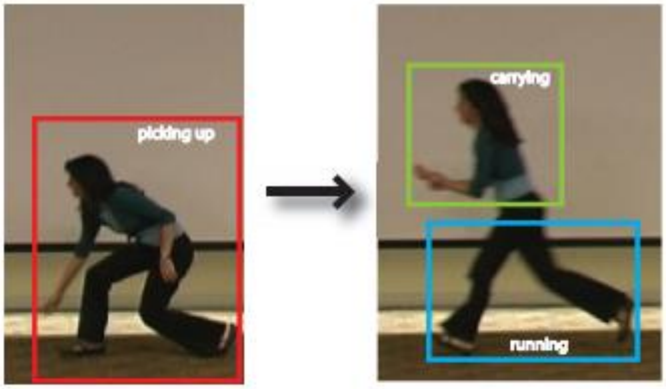


Figure 1.2: Complex action

Under these conditions, we trust that an extremely minimized portrayal ought to be sufficient to acclimate the requirements of single activity recognition, and displaying such conservative portrayals is our main thing in the initial segment of this proposal.

1.2 MOTIVATION OF THE WORK

The improvement of computer vision has empowered the event of various novel recognition strategies in both 2D images and 3D video classifications. In spite of the fact that it is as yet difficult to perceive a particular protest from a dataset of images because of perspective change, enlightenment, incomplete impediments, and intra-class distinction et cetera, numerous effective strategies have been proposed. In any case, for the video recognition issue, the present strategies still need change, particularly for reasonable motion images which have wide varieties in individuals' stance and garments, dynamic foundation, and fractional impediments. Instinctively, a clear way is contrasting an unknown video and the preparation tests by processing connection between the entire recordings.

This approach makes great utilization of geometrical consistency, however it is not practical when managing camera movements, zooming, intra-class contrasts and non-stationary foundations. Indeed, activity recognition has turned out to be one of the most smoking examination ranges in PC vision and amazing advancement has been made toward this path. Be that as it may, advance is principally restricted to a controlled test condition, which may prompt troubles when we move to perceiving and breaking down activities in more practical situations. To comprehend the conceivable troubles, let us initially look at a few presumptions which have been made in customary activity recognition (i.e. controlled condition):

1. Pre-processing Assumption: For a computer vision problem, choosing appropriate visual features and representation is the first step to solving the problem. In most cases, the feature extraction requires some pre-processing steps. In action recognition, this pre-processing step can be the detection and tracking of body parts or a moving person, or the segmentation of the region of interest. However, if these pre-processing steps fail, the methods based on them will breakdown.

2. Data Assumption: Most action recognition systems are based on statistical machine learning methods, which learns a classifier from a set of training data. In the usual case, sufficient labelled training data is assumed to be available. However, when the labelled training data is insufficient or unavailable or the data can only be obtained from more complex settings, say, from an ambiguously annotated dataset, the system structure of the training process will need to be changed accordingly.

3. Model Assumption: To mathematically model an action, we often make the assumption that an action can be viewed as an equivalent simplified vision/machine learning problem. For example, if an action is represented by a set of silhouettes, an underlying assumption is that an action can be characterized by the temporal evolution of 2D shapes. If we model an action by a bag of local features, we assume that an action can be characterized by the order-less local spatial temporal patterns. Of course, the assumption does not always hold in many applications.

1.3 RESEARCH OBJECTIVES

Our main objectives in this dissertation are three-fold:

- To develop an algorithm to handle intra-class variance and inter-class similarity in human actions.

- To develop an approach for considering semantic relations in bag-of-words framework by computing a concept space in which visual words are more semantically distributed.

- To develop an algorithm for multi-view action recognition this helps in overcoming occlusion.

1.4 THESIS ORGANIZATION

This proposition is isolated into five sections. These are depicting beneath:

Part 1 – Introduction: This section gives the proposal diagram and foundation to the subject of human activity recognition. The key point, inspiration and work extent of the proposal are additionally characterized alongside the postulation structure.

Part 2 – Literature Review: This section distinguishes slants and gives association of affiliation that as of now existing work. It will likewise research movement recognition techniques, for example, GMM, HMM, space time intrigue point and limited state machine.

Part 3 – Approaches: This section will give details about our Algorithm and proposed idea.

Part 4 – Outcome: This part includes experiments and results.

Part 5 – Conclusion: This part will sum up the work proposed in the thesis.

CHAPTER 2

LITERATURE SURVEY

The study of a motion recognition system is the most important aspect for designing a new one. A motion recognition system composes of three major steps i.e. Human body extraction, analysing the image sequences using a technique and classifying the motion using a classifier.

The HAR techniques are based on key points, global and local features representation, spatial-temporal features, bag-of-words etc. (J & M, 2011).

In (Dinesh K & Kuldeep, 2016), shape and motion features of human silhouette have been used for human action recognition. SDGs and SDPs of AESI were used to compute shape information and R-Transform was used for motion information. The key poses were identified to and the test input was compared with these key poses. The temporal content was measured using R-Transform. PCA was used for dimensionality reduction because the R transform produces results in high dimensionality. SDPs is the sum of directional pixels and is an efficient way to calculate the density of the region that is the key area to identify a pose.

In (Songtao Ding & Shiru Qu, 2016), Harris Laplace method was used to produce the interest points that were used in the spatio-temporal technique to calculate the feature vector and analyse the human activity. High response regions were detected using Gabor filter. This technique was based on identifying the key poses and storing them as dictionary. The spatio-temporal technique is based on local features. The dictionary built was hence called the bag-of-visual (BoV) model. This model is integrated with a SIFT descriptor to build the visual vocabulary. The dataset used was KTH which is a benchmark dataset and the results were compared with other techniques that have worked on the same dataset and it proved to be better.

(Lianliang Cao, YingLi Tian, Benjamin Yao, Zhengyou Zhang, & Thomas S. Huag, 2010) used multiple STIP features for human activity recognition. GMM (Gaussian Mixture Models) and Branch and bound model were combined to locate the action efficiently. The basic idea was that the relevant features must be used to detect the action and the irrelevant features must not be used. This would adversely affect the efficiency. The dataset used was KTH and so the results were directly compared with other techniques that used the same dataset.

(Di & Ling , 2012) proposed an approach human activity recognition that combined both local and global features. The method modified the existing bag-of-words approach to a bag-of-correlated poses. For using as a feature, the histograms of image sequences were modified to correlogram of image sequences to represent the visual words. For dimensionality reduction, PCA (Principal Component Analysis) was used. LDA (Linear Discriminant Analysis) was also used for locaing the interest points. K means clustering is used to differ the different clusters based on Mahalanobis distance is used. The correlogram matrix is constructed then using the MHI (Motion History Image) and GEI (Gait Energy Information). Bag-of-words is then called the Bag-of-Correlated Poses (BoCPs). Again PCA is used for dimensionality reduction. IXMAS dataset is used for implementing the technique and getting the accuracy of the proposed technique.

(De, 2015) proposed a simple approach based on extracting features, training the classifier and then performing the recognition. Silhouette is extracted using Background Subtraction. The foreground is extracted using Karman filters and Gaussian background models. The duration of action is then estimated using Local Linear Embedding (LLE) which is a non-linear dimensionality reduction method K nearest neighbours are then assigned to each point. The silhouette image sequences are analysed and then the Silhouette Principal Component Image(SPCI) is constructed. PCA is used for dimensionality reduction. The multi-class problem of action recognition is then handled using multi-class SVM. Two class samples are selected of k-classes every time and they are fused together. So it take k(k-1)/2 cycles for classification and action recognition. The technique was implemented using the KTH dataset.

CHAPTER 3

METHODOLOGY

3.1 GENERAL OVERVIEW

In general, the process of identifying an action involves the extraction of the frames and then from those frames, the human body is extracted. The object is identified in the image. The sequence of images is used to check the ordering of poses of the human body. For silhouette creation, there are various methods that can be used. Using Silhouette images for activity recognition is an efficient and accurate way to judge the pose of the human body. We have used silhouette images in our proposed work as well.

Figure 3.1 depicts the silhouette images for a human body that exhibits the action of walking. Similarly, we can create silhouette images of other actions as well. The key point in identifying any action is dependent on the ordering of the images and also the speed of the body which affects the identification process.

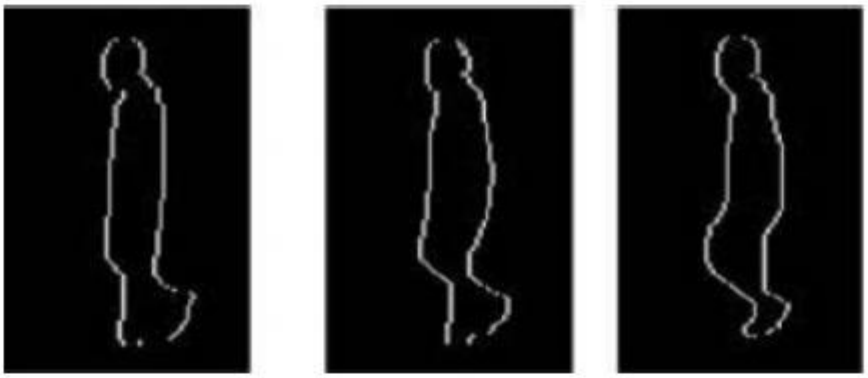


Figure 3.1 Silhouette Images for a human body that exhibits walking

3.2 ACTIVITY RECOGNITION

For activity recognition, we must identify what part of the video contains the activity so that we can analyse it and recognize the activity. The different approaches are discussed as follows:

Table 3.1: A comparison table of various methodologies

REFERENCES	METHODOLOGY	ADVANTAGES	FUTURE WORK
[3]	multiple instance-SVM	Enhances the neighborhood highlight based action acknowledgment by utilizing propelled machine learning methods which are not the same as sack of-highlights based portrayal	incorporates spatio-temporal information and different descriptors
[5]	subspace clustering approach	can deal with multi-dimensional information that are impractical with the run of the mill grouping strategy	points in intertwining a huge logical data, for example, feelings, wellbeing conditions.
[6]	Bayes net combined with trajectory	capacity to distinguish action even in medium determination recordings	It can be stretched out to security applications
[9]	Kinematics Model-Based	Takes just the conspicuous human represents that dispatches the	handling of similar action recognition

		data and takes out the rest	
[9]	Kinematics Model-Free	Learns key poses depending on Contour	Has high resistance to inter actor variance
[6]	Physics Model-Based	Dynamic elements are registered and by utilizing these elements activity classes are ordered in wording torques	Dynamically applying features to gait.
[9]	Kinematics Model-Based Approach	Versatile vision-based human activity acknowledgment strategy is proposed.	Adaptive learning ought to be contrasted with other benchmarking incremental learning what's more, constant adjustment techniques.
[9]	Kinematics Model-Based	Another skeletal portrayal that particularly models the 3D geometric connections among different body fragments utilizing pivots and	increase the framework to show complex method.

		interpretations in 3D space.	

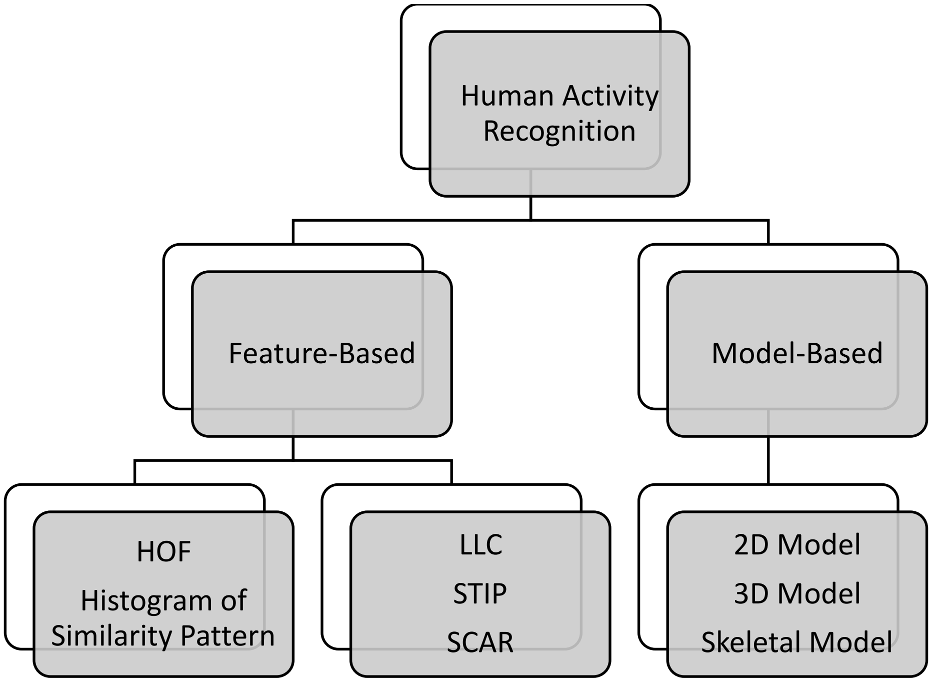


Figure 3.2 Human Activity Recognition Techniques

A. Feature Based Approaches

a. HOF (Histogram Of Optical Flow)

b. Histogram of Similarity Pattern

c. LLC (Local Linear Coding)

d. STIP (Spatio-Temporal Interests Points)

e. SCAR

B. Model Based Approaches

a. 2D Model

b. 3D Model

c. Skeletal Model

3.3 SILHOUETTES

Silhouettes are important and informative features in recognizing actions and have been used widely. One of the common methods for extracting silhouettes is background subtraction [2]. It is a popular method for segmenting the foreground moving regions from the background by constructing a model for scene background and comparing each frame to it. In simple and homogeneous backgrounds, silhouettes are extracted by background subtraction. But for more complicated backgrounds or for moving camera, extracting correct foreground is not easy. In order to obtain shape information from silhouettes, they should be described using appropriate descriptors. Several methods based on regions, boundary, skeleton and bounding box have been used to encode the shape from extracted blobs. Medial axis transform skeleton is another method for representing the shape information using skeleton curves. Furthermore, the pixels within the bounding box have been used to model the shape.

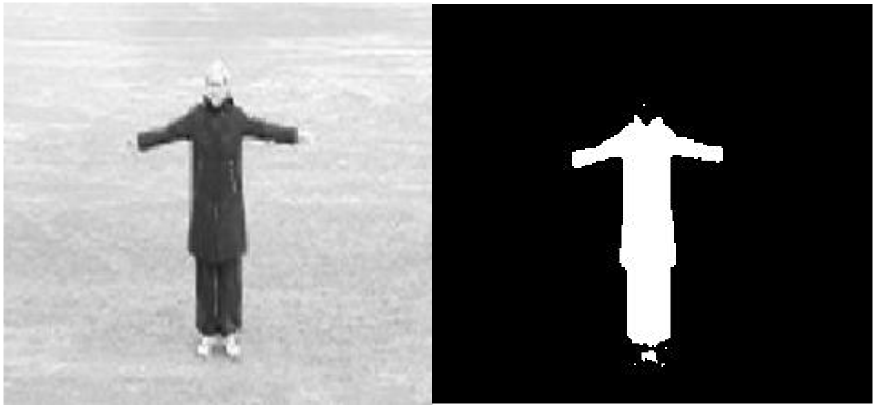


Figure 3.3 Silhouette Image for Arm Waving

These pixels are vectorized and embedded into proper subspaces using dimension reduction techniques [3-5]. This method of description has the benefit of robustness with respect to boundary noise. All these methods describe individual silhouettes from each frame. There are also methods which use all the silhouettes of a sequence to construct a volume and extract features from the entire space-time shape [8].

3.4 KNN CLASSIFIER

KNN is based on the instances. It relates unknown instances to the known instances based on a similarity or dissimilarity criteria. The instances are related to each other depending on a distance function that calculates the distance of an instance from a class. Here No information is abstracted from the training data. In learning phase, the data is encapsulated. It is also referred to as nearest neighbour classifier because it involves the instance to be located in the instance space.

In this process, k neighbours are located and the majority is allowed to vote depending on which the unknown instance is classified. The smoothness of the classifier is based on the value of k. The classifier is smoother and less locally sensitive with higher values of k.

But the disadvantage of increasing k to a very high number is that as k approaches n, all the unknown samples will be classified to a single class that is most frequently occurring class. So, an optimal value of k must be selected so that the classifier works correctly and does not classify all the unknown samples or instances to the same class that is most frequently occurring.

Our method uses Eucledian Distance as the distance function to classify the unknown instances to known classes S.

It is a very simple but effective classifier. It classifies the unknown instances to the known classes depending on the minimum value of the distance from that class.

The k-NN classifier has been explained as follows:

The data is represented as a matrix A = N x S, containing S scenarios b^1, $b^2, \dots, b^S$.

Each scenario b^i contains N features b^i={ b_1^i, b_2^i, … , b_N^i }. This matrix is accompanied by a vector e with length S of output values e = $\{e^1, e^2, .., e^S\}$, that lists the output of each scenario.

The input is a set of k training objects, B, and the object to be tested is t = (a’,b’). The distance between the test object and all the points in the training sets is calculated. The distance to be calculated is based on the distance function that we are using. The set of k training objects closest to the test object are selected. We name them as $B_{t.}$

In this system, KNN algorithm is used the suitable result by mixing the Euclidean distance among the various kinds of distance metric. The Euclidean distance is as shown in below :

$$b_{ij} = \sqrt{\left(a_{i1} - b_{j1}\right)^2 + \left(a_{i2} - b_{j2}\right)^2 + \cdots + \left(a_{ip} - b_{jp}\right)^2} \tag{3.1}$$

Where

d_{ij} = the distance between the training objects and test object

ai = input data for test object

bj = data for training objects stored in the database

In KNN algorithm, there are several advantages and disadvantages:

- Advantages
  - It can handle noise in the training data.
  - KNN is suitable where there are multiple classes or the data has multiple modes.
  - KNN is easy to understand, simple but effective.

3.5 SURF (Speeded up Robust Features)

SURF is a feature detector as well as a descriptor. It is based on finding the interest points. A template image is used a reference here. A threshold is used for matching the current image with the template image. An optimal threshold is chosen to perform accurate matching. The interest points are detected using Hessian matrix.

For any point t = (x,y) in image I, a scale is used. The scale σ is used to find the Hessian matrix as:

$$K = \begin{bmatrix} H_{xx}(t,\sigma) & H_{xy}(t,\sigma) \\ H_{xy}(t,\sigma) & H_{yy}(t,\sigma) \end{bmatrix} \tag{3.2}$$

Where $H_{xx}(t,\sigma)$ is the convolution result of second derivative of Gaussian filter $\frac{\partial y}{\partial x}\frac{\partial y}{\partial x}g(\delta)$ and I(x,y),

$$g(\delta) = \frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{3.3}$$

,

H_{xy} , H_{yy} are calculated using the same way.

At different scales, different interest points can be found out. After locating interest points, assignment of a direction and feature descriptor is done. A vector is formed V_{sub} = $(\sum dx\, , |\sum dx|, \sum dy\, , |\sum dy\,|\,)$.

The normalisation of the feature vectors is done to 1.

The comparison of descriptors from different images is done to find the matching pairs.

3.6 HARRIS CORNER DETECTOR

A corner detecor is generally used for detecting sharp corneers or edges or finding shapes that are different from each others. Corners are basically junctions or contours. A corner is detected if there is a large change in the direction. According to the corner theory, there are 3 types of regions:

- Flat

  A flat region is identified where there is no change of intensity in all directions as we move the window.

- Edge

  If there is no change along the edge direction, then it is recognized as an edge. That means if we are moving in the direction of a line then the direction along the edge is not changed.

- Corner

  If there is a significant change in all directions as we move along the edge then it is identified as a corner.

Harris Corner Detector is a mathematical way of detecting a corner.

Harris corner detection algorithm has many applications in the area of machine vision. It is realized by computing pixel's gradient value. A particular pixel is judged as a corner pixel if the absolute gradient value in both the directions is above the set threshold. Let us define a square patch N ∈ I centered on x_0 and y_0 for the given 2D image I. To compute the image gradient in both horizontal and vertical directions, the sum of the squared difference (SSD) between N and a shifted window $N_{(\Delta x,\Delta y)}$ is computed for all the 16 x 16 windows:

$$\boldsymbol{SSD} = \sum_{(\boldsymbol{xi},\boldsymbol{yi})} \boldsymbol{I}(\boldsymbol{a_i}, \boldsymbol{b_i}) - \boldsymbol{I}(\boldsymbol{a_i} - \Delta \boldsymbol{a}, \boldsymbol{b_i} - \Delta \boldsymbol{b}) \tag{3.4}$$

Matrix M is given by

$$\boldsymbol{M} = \begin{bmatrix} \sum_{(\boldsymbol{xi},\boldsymbol{yi})\in \boldsymbol{N}} \boldsymbol{K}\rho(\mathrm{i})(\nabla_i^h) * (\nabla_i^h) & \sum_{(\boldsymbol{xi},\boldsymbol{yi})\in \boldsymbol{N}} \boldsymbol{K}\rho(\mathrm{i})(\nabla_i^h) * (\nabla_i^v) \\ \sum_{(\boldsymbol{xi},\boldsymbol{yi})\in \boldsymbol{N}} \boldsymbol{K}\rho(\mathrm{i})(\nabla_i^h) * (\nabla_i^v) & \sum_{(\boldsymbol{xi},\boldsymbol{yi})\in \boldsymbol{N}} \boldsymbol{K}\rho(\mathrm{i})(\nabla_i^v) * (\nabla_i^v) \end{bmatrix} \tag{3.5}$$

Matrix M is a symmetric matrix. ∇_i^h, ∇_i^v represent first order partial derivatives of image I along horizontal and vertical directions at pixel (x_i, y_i), respectively. The tensor product ∇I x ∇^T_I over the window N is calculated and the matrix M is computed by averaging, and then multiplied with Gaussian function Kρ(i) where, ρ is the standard deviation which makes it positive semi-definite. The eigenvalues ($ǰ_1$ and $ǰ_2$) of the matrix M provide the variation corresponding to the partial derivatives in orthogonal directions.

3.7 PROPOSED APPROACH

a. Harris Corner Detection

- Detecting response at any shift (x,y) is done using gaussian instead of 0-1 weights.
- The eigen values, $ǰ_1$ and $ǰ_2$, of matrix M are then computed.
- Gradient vectors are a set of (dx,dy) with set of centre of mass at (0,0).
- Ellipse is used for displaying the points.

The corners are identified depending on the value of eigen values of M.

- If $ǰ_1$>> $ǰ_2$, vertical edge is identified.
- If $ǰ_2$ >> $ǰ_{1,}$ horizontal edge is identified.
- If $ǰ_1$~ $ǰ_2$ both are very large, then SSD increases im all directions that implies that it is a corner.
- If $ǰ_1$~ $ǰ_2$ and both are very small, it implies it is a flat reegion.

b. Steps for Silhouette Images and Interest Points

- Silhouette Images are generated for all the frames. The silhouette extraction involves the following steps:
    1. Obtain the negative texture mask image
    2. The image obtained in previous step is converted into a grayscale image.
    3. Convert the above into binary image.
    4. Find the ROI from the Harris Corner Points obtained earlier and resize all images to same size.
- Create the image width map of the image sequences from the silhouette images obtained. A width map is the number of picels stored in both x and y directions to find the centre of mass as well identify the key pose for the action.

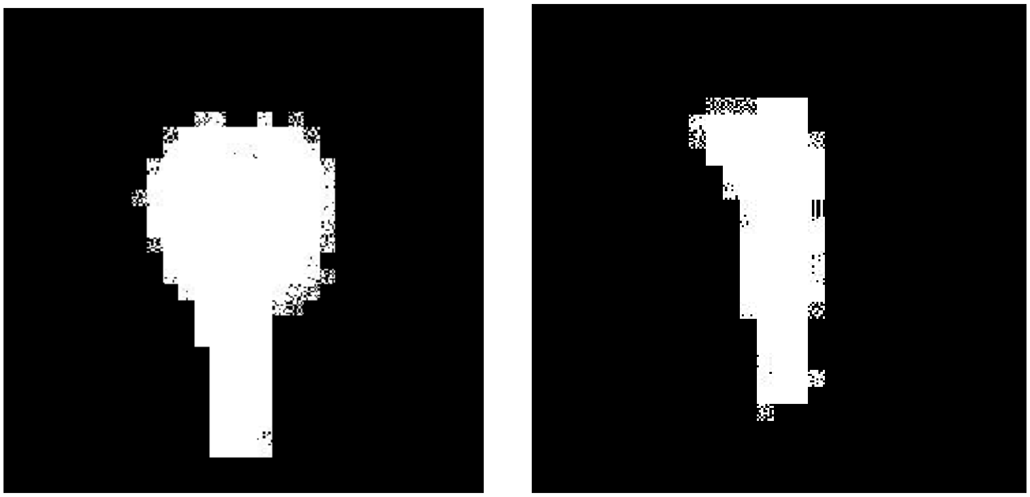


Figure 3.4 Image width map of handwaving and boxing

- The Harris Points obtained are the interest points obtained and we depict them using ellipses.
- Removal of Background Points is a key point here. The key points obtained that are not near the human silhouette depending on a neighbourhood value are removed from consideration which increases the accuracy of the system.
- The Sum of Pixels is calculated from the image width map obtained.
- Based on number of the windows that overlap, a feature is extracted and also the gradient's angle is computed using the following equation:

$$D_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * B \quad \text{and } D_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * B \tag{3.6}$$

- D_x and D_y are horizontal and vertical derivatives and * is convolution with the input image B.
- The Magnitude and angle of gradient are calcualted using following equations:

$$\text{Gradient magnitude, } D = \sqrt{D_x^2 + D_y^2} \tag{3.7}$$

$$\text{Gradient Direction, } \delta = \text{atan}\left(\frac{D_x}{D_y}\right) \tag{3.8}$$

- The histogram of gradients is built using horizontal nearest point and vertical nearest point.

c. Classifying the action

- The image width map is divided into N x N overlapping windows.
- KNN is used to assign label y to x.
- The action is classified depending on the K nearest neighbours to that image width map feature.

CHAPTER 4

RESULT & IMPLEMENTATION

4.1 SIMULATION RESULT

In this section, we present the results obtained testing the activity recognition proposal within KTH dataset.

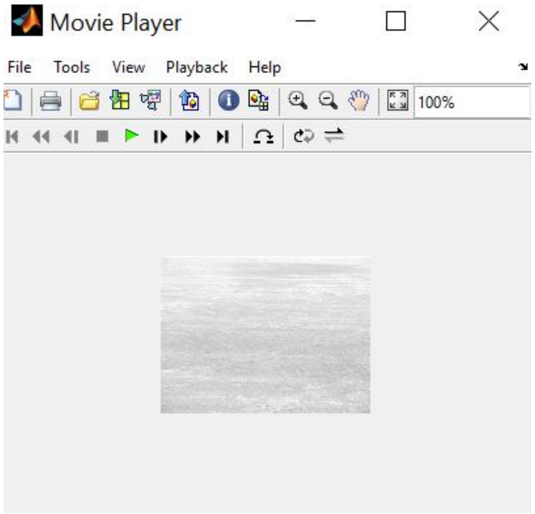


Figure 4.1: Input video from KTH dataset

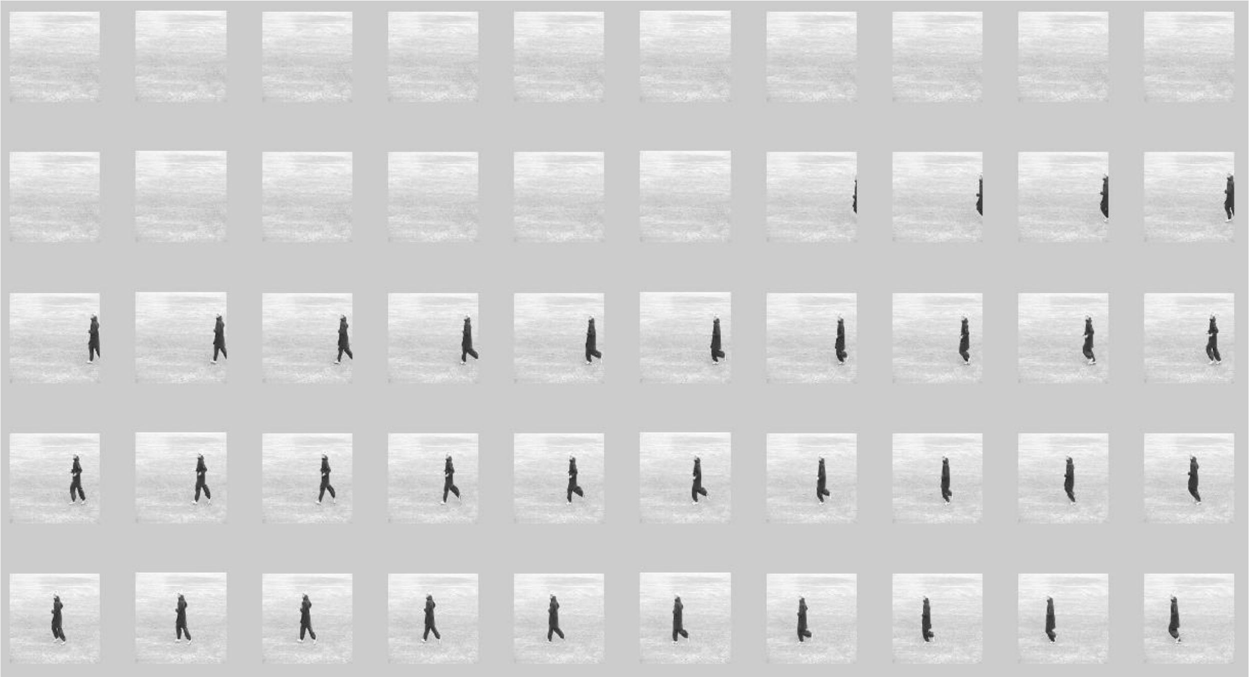


Figure 4.2: Framing of input video

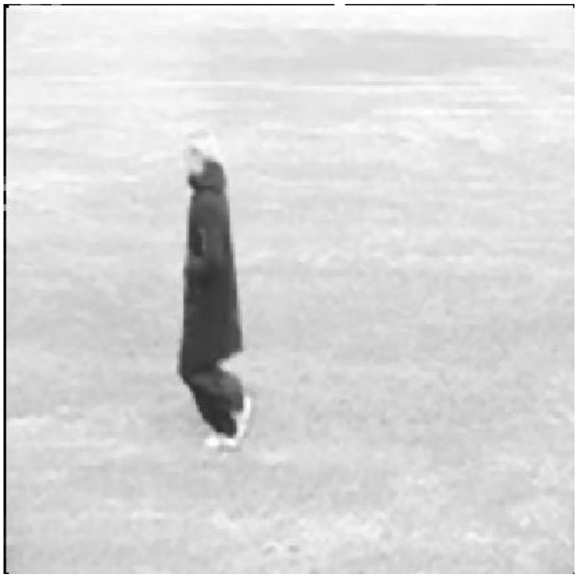


Figure 4.3: Pre-processing of input video

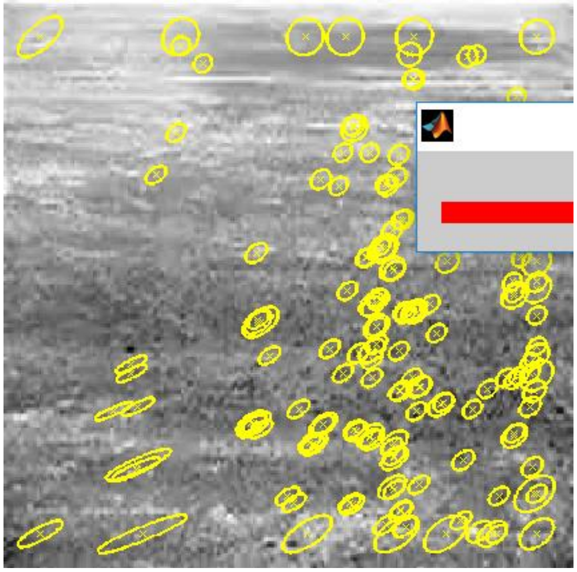


Figure 4.4: Corner detection using Harris points

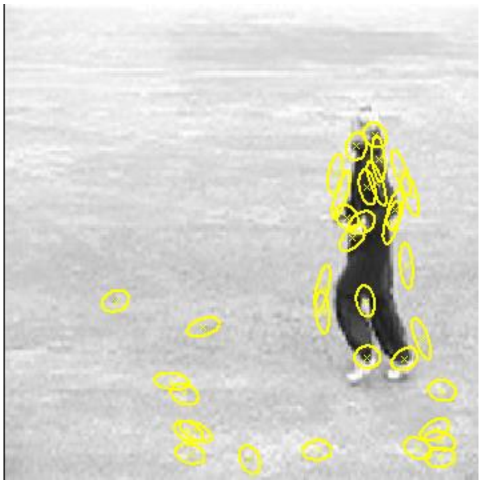


Figure 4.5: Pruning steps of video

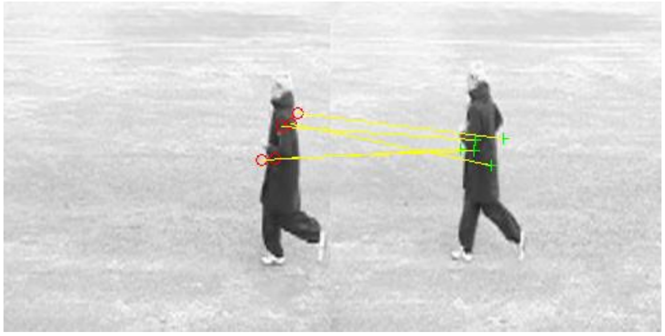
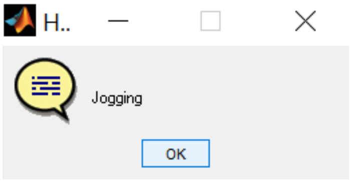
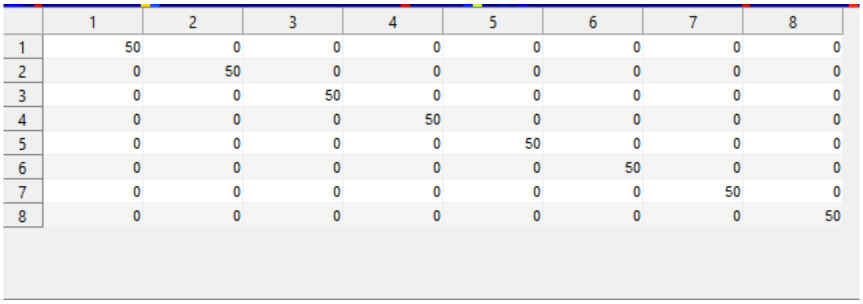


Figure 4.6: Surf description of input video



Jogging

Figure 4.7: Motion detected



	1	2	3	4	5	6	7	8
1	50	0	0	0	0	0	0	0
2	0	50	0	0	0	0	0	0
3	0	0	50	0	0	0	0	0
4	0	0	0	50	0	0	0	0
5	0	0	0	0	50	0	0	0
6	0	0	0	0	0	50	0	0
7	0	0	0	0	0	0	50	0
8	0	0	0	0	0	0	0	50

Figure 4.8: Matrix of input video

CHAPTER 5

CONCLUSION

In this Thesis, our proposed strategy for moving human body will find the moving thing perfectly in the incorporated way. It can be proficient with high exactness and relentless quality. To restrict or keep up a vital separation from the issues moving closer in moving article disclosure, we used SURF system to see moving thing, establishment presentation and invigorate the present picture continuously. In which the moving human body acknowledgment is the most basic bit of the human body development examination, the question is to perceive the moving human body from the establishment picture in video groupings, and for the consequent treatment, for instance, the goal arrange, the human body following and lead understanding, its effective disclosure expect a fundamental part. Human development examination concerns the area, following and acknowledgment of person's practices, from picture groupings including individuals.

According to the result of moving thing acknowledgment analyse on video progressions, this wander presents another calculation for recognizing moving articles from a static establishment scene, to recognize moving thing in light of establishment subtraction. We set up a NBKNN technique to confine the effect of lighting up. Starting now and into the foreseeable future, Harris corner area is begun to corner establishment interruption inconvenience then the moving human bodies are exactly and reliably perceived. The trial comes to fruition show that the proposed method runs rapidly, decisively and fits for the on-going disclosure. This system has also a nice effect on the finish of racket and shadow, and has the ability to expel the aggregate and exact picture of moving human body. The propagation comes to fruition by MATLAB exhibit that the establishment subtraction is important in both recognizing and following moving articles, and the establishment subtraction calculation runs more quickly.

References

1. Bangpeng Yao, & Li Fei-Fei. (2012). "*Recognizing Human Object Interactions in Still Images by Modelling the Mutual Context of Objects and Human Poses*". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1691 - 1703.

2. D. K. Vishwakarma, & R. Kapoor. (2015). "*Integrated Approach for Human Action Recognition Using Edge Spatial Distribution, Direction Pixel and R-Transform*". *Advanced Robotics*, vol. 29,no. 23, pp. 1551-1561.

3. De, Z. (2015). "*Recognizing Human Actions via Silhouette Image Analysis*". *Chinese Control and Decision Conferece (CCDC),* IEEE, Beijing, China. pp. 5870 - 5874, DOI: 10.1109/CCDC.2015.7161859.

4. Di , W., & Ling , S. (2012). "*Silhouette Analysis-Based Action Recognition Via Exploiting Human Poses*". *IEEE,* Trans. Circuits Syst. Video Technol., vol. 23, no. 2, pp. 236–243, Feb. 2013

5. Dinesh K, V., & Kuldeep, S. (2016). "*Human Activity Recognition Based On Spatial Distribution Of Gradients at Sub-Levels of Average Energy Silhouette Images*". *IEEE*. vol. PP, no. 99, pp. 1-1

6. Han Su, Jiayun Zou, & Wenjie Wang. (2013). "*Human Activity Recognition Based on Silhouette Analysis Using Local Binary Patterns*". *International Conference on Fuzzy Systems and Knowledge Discovery.* IEEE, pp. 924-929, DOI: 10.1109/FSKD.2013.6816327.

7. J, A., & M, R. (2011). "*Human Activity Analysis: A Review*". *ACM Computing Survey* (pp. 16-43). IEEE pp. 165-169, DOI: 10.1109/ACIRS.2017.7986086.

8. Jian Cheng, HaijunLiu, Feng Wang, Hongsheng Li, & Ce Zhu. (2015). "*Silhouette Analysis for Human Action Recognition Based on Supervised Temporal t-SNE and Incremental Learning*". *IEEE Transacitons On Image Processing*, vol. 24, no. 10, pp. 3201-3217.

9. K. Charalampous, & A. Gasteratos. (2014). "*Online Deep Learning method for Action Recognition*". *Pattern Anal. App.*, pp. 165-170, DOI: 10.1109/HUMANOIDS.2016.7803273.

10. Lianliang Cao, YingLi Tian, Benjamin Yao, Zhengyou Zhang, & Thomas S. Huag. (2010). "*Action Detection USing Multiple Spatial-Temporal Interest Points*". *ICME 2010.* IEEE, pp. 340-345, DOI: 10.1109/ICME.2010.5583562.

11. Songtao Ding, & Shiru Qu. (2016). "*An improved Interest Point Detector for Human Action Recognition*. *Chinese Control and Decision Conference (CCDC).* IEEE. pp. 4335-4360, DOI: 10.1109/CCDC.2016.7531750.