CHAPTER 1
INTRODUCTION

Human activity recognition is considered to be a classification problem using time varying visual data. Visual data is separated from video groupings and spoke to in important highlights, which are utilized to coordinate with the highlights extricated from a gathering of marked reference arrangements speaking to commonplace exercises. During the extraction strategy, three sorts of highlights might be included: single item's highlights (i.e., position, speed, veins, shape, shading and so forth.), worldwide highlights of numerous articles (i.e., normal speed, district inhabitance, relative positional varieties and so on.), and the connections among them.

Recently, there has been increased demand in the field of action recognition system. These types of system lead to better surveillance and monitoring. Abnormal event is the one which stands out and requires more consideration and which relies upon the setting of scene considered. Abnormal human activities are the one which generally occur at low frequency and have irregularity in comparison to normal human activities. Anomalous human movement identification is in this way cultivated by finding the anomalies of the ordinary occasion because of low occurrence of such anomaly.

1.1 Problem statement

As the increasing population of older people and disabled ones in the society, the need of assistive systems is increasing day by day. Most of the people need independency and like to reside independently in homes, so a better technical solution to give assistance to them were the activity recognition systems. Through a camera assisted environment this can be carried out simply. The multiple camera environments give the better detection of each activity. The activities include the normal daily activity and the abnormal activities. Those two become detected by the proposed technique. The early techniques like accelerometers, help button, wearable sensors, etc. are also used by people. But now the technology was developed and the better ways by which the activity detection are emerged.

1.1.1 Motivation

There are large number of utilization of activity recognition systems around the world. Activity recognition has numerous applications in different fields like Health care, video surveillance, etc.

Video Surveillance: The present concern around the world relating to the security in urban cities and towns around the globe are of great concern to the individual and governments. The various activities like assault on people, burglaries, fight and provocation are few of the examples where activity recognition can play a vital role. High quality audio and video information processed are of great importance to police and security agents. Video surveillance are not used to effectively recognize these types of activities but also to prevent them

Figure 1.1: Surveillance

There has been rapid increase in demand for system for video surveillance and thus demand for activity recognition is also on rise.

Healthcare: Human activity recognition is turning out to be of great importance and use in healthcare and meditation field with the main goal of monitoring and observation of individuals. There is a great demand and interest for systems that process human activities, and easily identify and emerge about upcoming and existing physical and psychological well-being of patients. This is mainly important for older individuals, for whom such system help to live at home in much more safe and secure environment.

Figure 1.2: Fall Detection

Sports: Violence detection in non-crowded scenarios such as fights in stadiums where due to the size of stadium any delay in action against violent activities can cost someone's life either on or off field.

Figure 1.3: Fight between players

1.1.2 Objective

Our primary goal is to distinguish unusual human action from video.

So as to accomplish our primary goal, we should accomplish the following objectives:

- Our first objective is to identify and analyze existing local feature-based techniques so that we can choose one in which local features and which video portrayals work the best.
- Our second objective is to comprehend the confinements of the current methods and to propose new method for activity recognition in recordings to go past the cutting-edge limitations.
- Our third objective is to structure and build up a model for abnormal human activity recognition.
- Our last objective is to analyze the performance of the proposed model.

1.1.3 Scope

- Relevant features from the interest points are obtained from the video using feature extraction.
- Features are extracted from each video and a feature vector is obtained for every spatio-temporal region of an image.
- Video representation is the first and one of the most important sub-problems in video preprocessing. A good representation should include the key point and useful information for discrimination while discarding unnecessary information.
- Feature vector after video representation became the input of some classifier.
- It is not performed for global features it is only performed for local features.
- The dataset for our project will be collected from different online sources and will tested on model.

1.2 Related Work

Recently, there is an expanding enthusiasm for programmed investigation of a video transfer so as to produce alarms continuously when a "strange" occasion occurs. Such calculations might be utilized as a consideration system, which with legitimate discovery and bogus caution rates, will empower a solitary administrator to successfully "watch" an enormous number of cameras. Anomaly identification is a functioning region of exploration all alone. Different methodologies have been proposed, for both crowded and non-crowded scenes. They can be extensively ordered by the kind of scene portrayal embraced. One well known classification depends on trajectory modeling. It includes tracking each article in the scene and learning models for the subsequent item tracks. The two tasks are very troublesome on densely crowded scenes, for which these methodologies are not promising.

Interest points give smaller and theoretical portrayals of examples in a picture. Along these lines, to broaden the idea of spatial interest points into the spatio temporal area and show how the subsequent features regularly reflect fascinating occasions that can be utilized for a minimized representation of video information just as for its clarification. To identify spatio temporal events, we expand on the possibility of the Harris and Forstner interest point administrators and detect nearby structures in space-time where the picture esteems have critical local variations

in both space and time. We at that point gauge the spatio-temporal degrees of the recognized events and process their scale-invariant spatio-temporal descriptors. Utilizing such descriptors, we classify events and build video representation as far as named space-time points.

These methods depend on the 2D and 3D data extracted from the human body party., which is extremely hard to accomplish in practical and unconstrained recordings. Despite the fact that cost-effective depth cameras help in the extraction of human body joint focuses, they likewise bring enormous impediments, for example to the scope of the depth sensor, and this is actually why we don't utilize them in this proposition.

1.3 Challenges

Because of unpredictable and inconsistent nature of human activities, human activity recognition remains a difficult task. Some of the challenges faced in human activity recognition are-

- **Intra and inter class variation of human activities -** There are various ways in which a activity can be performed, for example, person may have different way of walking, wearing different apparels (different pattern, size and fitting) or walk at different speed. Again, classes of exercises can be different for individual, some may perform task in altogether different from other. Model for human activity recognition should be such that it accommodates every case of action performed by the individual and discriminate different kinds of activities.

- **Viewpoint -** The viewpoint also plays a vital role in human activity recognition. For an instance in some case individual might be facing camera directly while in some cases it may be the side-on view of the camera. There is another issue of viewpoint relating to ground and aerial viewpoint of the camera which is a quite common in video surveillance system involving multiple camera monitoring the same area.

- **Brightness variation -**Brightness is another factor which may vary with change in condition or in a similar condition because of lighting conditions. Brightness in open area is totally different compared to a home environment.

- **Camera movement and jitter -** While thinking about a static, fixed camera this is once in a while an issue; however, in some genuine situations it likely could be an issue, for instance, an observation camera joined to an enormous post may experience the ill effects of some development in blustery conditions. Another such model is video recorded from hand-held gadgets, for example, camcorders or advanced mobile phones. For the most part, adjustment calculations can lessen the impacts of camera development, in spite of the fact that obviously no adjustment calculation is perfect in such manner and a few relics may even now engender through and influence the human movement model.

- **Complex and dynamic backgrounds -** In reality the background may not be simple like involving single individual or against a plain background. But in reality, the background may involve many individuals surrounded by objects and different individual moving in different directions. In all these cases it is extremely difficult for a system to focus on a single individual and predict the activity.

- **Partial and full occlusions –** In real world scenario it may frequently happen where a object may block part or entire body of an individual. It may also happen like a individual may be impeded by another person. This type of scenario might lead to some kind of indecision with the system like which individual is involved with the action.

1.1 ORGANIZATION OF THE REPORT

Chapter 2: Research and study of work related to AbHAR and action recognition. It categorizes past work with respect to action representation model and classification algorithm used.

Chapter 3: This chapter describes the system design and methodology of proposed system. It contains detailed description of all the major steps involved in abnormal human activity recognition.

Chapter 4: This chapter contains the implementation and results of proposed system. It draws comparisons of our results to that of previously used methods.

Chapter 5: This chapter draws conclusion to the results and applications of our project and ends with the list of references related to proposed system.

CHAPTER 2
LITERATURE SURVEY

In this chapter, a review of literature is presented and discussed in order to provide a theoretical background about the project. Various types of features can be used in the process for identifying the abnormalities. These features can be: Selective Spatial-Temporal Interest Points (STIP), Spatial-Temporal Pyramid Features, Optical Flow, Spatial Temporal Texture Map (STTM), Shape and motion features of human silhouette. Features can be represented by using many ways such as: Discrete Cosine Transform (DCT)-based image signatures, Bag of Features(BOF), Spatial-Temporal Pyramid Networks(STP Net), Histogram of Optical Flow Orientations (HOFO), MODEC Model And Harris 3D, Multi-layer Gaussian model, 3D- Scale Invariant Feature Transform (3D-SIFT) detector. Various algorithms such as SVM, k-NN etc. can be used for classification.

2.1 AbHAR Based on STIP and Saliency

The proposed system is used to identify between normal and abnormal human activities from the visual input video sequence. Human activities which are considered to be normal include activities like sitting, walking, cleaning, jumping, relaxing on bed, etc. Human activities which are considered to be abnormal include activities like falling on ground, fall from a height, fall from standing position. In this system saliency map is calculated for each frame of video and interest points are extracted using selective STIP. 3-D image gradients are quantized to obtain feature vector. Activity description is done by building vocabulary using Bag of Feature (BoF) technique. Support Vector Machine (SVM) classifies activity as normal or abnormal.

2.1.1 Descriptor/Technique

The saliency is calculated for each frame of input video sequence using DCT based algorithm. After this feature extraction and description are performed using Selective Spatio-Temporal Interest Points (STIP) and Hidden Markov Model (HMM) descriptor

2.1.2 Dataset

As the real dataset containing abnormal activities of elderly people is difficult to acquire, two benchmark datasets such as Dataset UR-fall detection (URFD) and Dataset S provided by Le2i CNRS are used to estimate the performance of the pro- posed system to detect abnormal activity. UR fall dataset: The dataset consists of 70 (30 fall activities + 40 activities of daily living) RGB and depth videos. Dataset S by Le2i CNRS: The entire dataset contains total 221 videos out of which 126 are of abnormal activities and 95 are of normal activities. The proposed system uses only RGB videos for processing. Holdout method is used for dataset partitioning with 50% samples under both categories are used for training and remaining 50% are used for testing.

2.1.3 Accuracy

For UR Fall dataset the system is giving 100% accuracy whereas in case of Dataset S by Le2i CNRS, though a system is giving false alarms, the rate of correct classification for abnormal activities is 96.83%.

2.1.4 Limitations/Drawbacks

1. Saliency computation may omit the useful interest points from the frames.
2. Sometimes it may produce false alarms. However, chances are very low.

2.2 AbHAR based on Movement Analysis from Video Sequence

The novel anomaly indicator is derived from HMM in which learning takes place from the histogram of optical flow orientation from the frames obtained through input videos. The similarity if found out between normal and observed frames. The proposed system has been evaluated and tested on numerous surveillance dataset.

2.2.1 Descriptor/Technique

Feature descriptor focusing on movement information from region of interest is proposed. The descriptor used is Histogram of Optical Flow Orientations (HOFO). Firstly, low level movement information is obtained by finding out the optical flow from the resulting frames. Classifier based on Hidden Markov Model (HMM) is proposed to differentiate normal event from the abnormal event based on probabilistic properties of the HOFO.

2.2.2 Datasets

The system is validated and tested on UMN and PETS dataset. The UMN dataset has activities relating to crowded escaping events. The abnormal event is like an individual running in different direction while normal activity is like normal walking. PETS dataset is multisensor sequences containing different crowd activities. The main goal is to validate and test the system for crowd surveillance in real-world environment.

2.2.3 Accuracy

For UMN dataset: The detection accuracy is 97.24%.

For PETS dataset: The detection accuracy is 97.27%.

2.3 Spatio-Temporal Descriptor for AbHAR

Author proposed a descriptor based on spatio-temporal features known as Spatio-Temporal Descriptor (STD). The novel descriptor relies on STTM and it is based on 3D Harris. Even small variation in Spatio-Temporal domain are captured by this methodology

Descriptor described here is motivated from Spatio-Temporal Interest Point (STIP). Author prosed use of Spatio-Temporal Texture Map and Spatio-Temporal Descriptor for anomaly detection in human activity.

Dataset used - UCSD dataset

Accuracy – EER(Equal Error Rate)

Table 2.1 Equal Error Rate

Descriptor	Ped1	Ped2	Average
MIDT-temp	22.9%	27.9%	25.4%
Proposed STD	32.4%	28.5%	30.5%
MPPCA	35.6%	35.8%	35.7%
Force flow	36.5%	35.0%	35.8%
MIDT-spat	43.8%	28.7%	36.3%
LMH	38.9%	45.8%	42.4%

Table 2.2 Computational cost

Descriptor	Time taken per frame
MIDT-temp	114.343s
Proposed STD	5.145s

2.4 AbHAR in Crowded Scenes

Feature extraction is performed using Harris or Fast detector from the frames of input video sequence. The system is evaluated and tested using UCSD pedestrian dataset and VIRAT dataset. The result obtained are shown in table

Table 2.3 Result for HARRIS and FAST Detector

Dataset	**Using Harris Detector**		**Using Fast Detector**	
	Detection Rate	False Alarm	Detection Rate	False Alarm
PED1	7.88	19.39	6.67	33.33
PED2	51.30	32.61	44.78	41.74
VIRAT	56.67	60	53.33	70

2.5 A 3-D SIFT Descriptor and its Application to Action Recognition

3D SIFT descriptor is proposed by the author or 3-D imaging like Magnetic Resonance Imaging (MRI) data. This descriptor is capable of representing the 3D characteristics of visual info in an efficient manner. Authors shows how 3D SIFT outperforms other descriptors.

Action Dataset is used for evaluation and testing.

Accuracy –

Table 2.4 Precision for different descriptors

Descriptor	**Average Precision**
2D SIFT	30.4%
Multiple 2D SIFT	47.8%
Gradient Magnitude	67.4%
3D SIFT	82.6%

The above table 2.3 shows the average precision on action dataset for different descriptors

2.6 AbHAR based on MODEC Feature

Author proposed method for cheating detection during examination. For this they proposed MODEC models and Harris 3D Interest point detector to extract the interest points and then do further processing and classification.

2.6.1 Technique

Feature are extracted by MODEC Model and Harris 3D and Classification is done with the help of MCMC LDA.

2.6.2 Data Set Used

Authors used their own dataset consisting of 49 Videos

Table 2.5 Count of videos by activity in the dataset

No	Activity	Abbrev.	No. of Videos
1	Cheat sheet	CHSH	13
2	Hand signal or code	CODE	6
3	Exchanging answer sheet	EXCH	7
4	Verbal communication	TALK	6
5	Looking at others' answer	LOOK	7
6	No cheating	NOCH	10
	Total		**49**

2.6.3 Accuracy

When Harris3D used the accuracy can be shown as:

Table 2.6 Accuracy with Harris3D

Experiment	Accuracy	Precision	Recall	F1-score
1	.429	.514	.417	.433
2	.571	.572	.556	.525
3	.357	.389	.333	.329
4	.571	.514	.528	.510
5	.357	.444	.389	.394
6	.429	.417	.389	.394
7	.500	.431	.500	.460
8	.500	.444	.472	.456
9	.429	.306	.389	.333
10	.571	.406	.528	.436
Average	**.471**	**.444**	**.447**	**.425**

When MODEC used the accuracy can be shown as:

Table 2.7 Accuracy with MODEC

Experiment	Accuracy	Precision	Recall	F1-score
1	.643	.625	.667	.621
2	.571	.489	.556	.514
3	.643	.611	.639	.611
4	.500	.350	.500	.403
5	**.714**	**.750**	**.722**	**.711**
6	.571	.583	.611	.573

Experiment	Accuracy	Precision	Recall	F1-score
7	.571	.594	.583	.561
8	.500	.514	.500	.489
9	.643	.639	.694	.662
10	.571	.653	.611	.600
Average	**.593**	**.581**	**.608**	**.574**

2.7 Activity Representation based on Multiple Instance Dictionary Learning

In order to identify positive features for every activity category Multiple Instance SVM (Mi-SVM) is used and codebook is generated using K-means algorithm. After this the features are encoded into the generated codebook using locality-constrained linear coding. This is followed by Spatio-Temporal pyramid pooling to convey the spatio-temporal stats. Now, SVM is used to classify the videos. Two Datasets were used i.e. a) KTH b) Hollywood2

Accuracy can be shown as:

Table 2.8 On KTH dataset

Approach	Average Accuracy
Proposed Method	**92.83%**
Wang et al. [18]	86.10%
Laptev et al. [?]	91.8%
Xiaojing et al. [21]	92.59%
Niebles et al. [30]	81.50%
Wang et al. [7]	94.2%
Le et al. [10]	93.9%

Table 2.9 On Hollywood2 dataset

Approach	Mean AP
Proposed Method	**51.8%**
Wang et al. [18]	47.4%
Laptev et al. [15]	45.2%
Le et al. [10]	53.3%
Wang et al. [7]	58.2%

2.8 Recognition of AbHAR Using the Changes in Orientation of Silhouette in Key Frames

Simple and effective method of abnormal activity recognition is proposed by the author which can be used for old age monitoring. Surveillance system in old age homes and for elderly persons has become an important point of research and development area in recent time for the safety of elders. Background subtraction is performed followed by the silhouette extraction. K-NN is used which is a learning classifier which means that it learns as more and more videos are processed using it. Author created his dataset of 10 videos for training and testing purpose. The dataset used has some limitation like activities are performed in good lightning and background variation is limited.

Table 2.10 Confusion matrix

Actual Activity/Recognized Activity	Fall towards left	Fall toward Right	Chest Pain	Headaches
Fall towards left	90%	0%	0%	10%
Fall toward Right	0%	90%	10%	0%
Chest Pain	0%	30%	70%	0%
Headaches	30%	0%	0%	70%

Table 2.11 Taxonomy of AbHAR Literature

Year	Author	Description	Features	Feature Representation method	Classifier	Dataset	Remark
2019	Smriti H. Bhandari and Navneet S. Babar	2D visual saliency map is created for each frame which is used for further processing for features extraction and description [2]	Selective Spatial-Temporal Interest Points (STIP)	Discrete Cosine Transform (DCT)-based image signatures, Bag of Features (BoF)	Support Vector Machine (SVM)	UR fall Dataset and Dataset S by Le2i CNRS	Accuracy of 100% for URFD dataset and 96.83 % for Dataset S.
2019	Zhenxing Zheng et. al.	Proposed a new spatial pyramid module to aggregate inherent multi-scale features of a CNN for action recognition.[1]	Spatial-Temporal Pyramid Features	Spatial-Temporal Pyramid Networks (S-TPNet)	Temporal Pyramid Module	UCF101	Accuracy vary between 84.5-85.1 %
2017	Tian Wanga et. al.	Proposed an algorithm based on an image descriptor which encodes the movement information and the classification method.[3]	Optical Flow	Histogram of Optical Flow Orientations (HOFO)	Hidden Markov Model(HMM)	UMN and PETS	Accuracy of 97.24% for UMN and 97.27% for PETS.

Year	Author	Description	Features	Feature Representation method	Classifier	Dataset	Remark
2017	Abdulamir A.Karim and Narjis M.Shati	Proposed HARRIS or FAST detector to extract list of pairs of interest points from the frames of video clips.[4]	Interest Points	Harris Detector	Seed Filling Technique	UCSD pedestrian and VIRAT.	Accuracy of 56.67% for UCSD and 70% for VIRAT.
2016	Janson Hendryli, Mohamad Ivan Fanany	To detect Cheating activity in examination MODEC models and Harris 3D is used to extract the interest points.[5]	Spatial-Temporal Interest Points (STIP)	MODEC Model And Harris 3D	MCMCLDA (Multi-Class Markov Chain Latent Dirichlet Allocation)	Own CCTV dataset of 49 videos.	Accuracy is 57.1% and 71.4% for Harris 3D detector and MODEC Model respectively y.
2015	Fam Boon Lung, and Mohamed Hisham Jaward	Proposed a Spatio-Temporal Descriptor(STD) based on spatio-temporal features of an image sequence. It is able to capture subtle variations in the spatio-temporal Domain.[6]	Spatial Temporal Texture Map (STTM)	Spatio-Temporal Descriptor (STD)	Gaussian Hidden Markov Model (HMM)	UCSD	Accuracy: Equal error rate of 30.5% with low computation cost of 5.1 seconds.

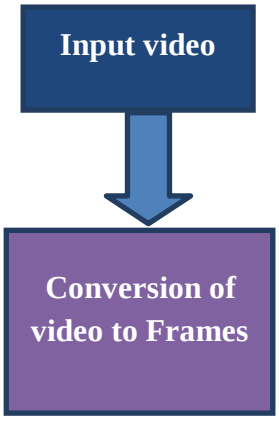
Year	Author	Description	Features	Feature Representation method	Classifier	Dataset	Remark
2015	Dr. Dinesh Kumar Vishwakarma	Proposed silhouette extraction by thresholding the difference image, obtained after background subtraction.[7]	Shape and motion features of human silhouette.	Multi-layer Gaussian model	K-NN	Own dataset of 10 videos.	Accuracy of 80% is obtained.
2014	Lakshmi Priya and Smitha Suresh	Human activity detection based on edge point movements and spatio-temporal features.	Spatio-Temporal Interest Points (STIP)	Bag of Features (BoF)	Support Vector Machine (SVM)	Own dataset	The working of the proposed system was done using an experimental video clip.
2007	Paul Scovanner et. al.	Proposed a 3-dimensional (3D) SIFT descriptor for video or 3D imagery such as MRI data.	Spatio-Temporal Interest Points.	3D- Scale Invariant Feature Transform (3D-SIFT) detector	Support Vector Machine (SVM)	Action	Accuracy is 82.6%.

CHAPTER 3
SYSTEM DESIGN AND METHODOLOGY

In this chapter, detailed explanation of the methods used in this project along with its architecture is provided. There were many alternatives present to the methods used in this project but the methods selected works with greater efficiency than any other alternatives. This chapter will throw light on proposed methodology, advantage of using the proposed methods, system architecture.

3.1 Proposed Methodology

The proposed method works for anomaly detection in the home environment. Human activities which are considered to be normal include activities like sitting, walking, cleaning, jumping, relaxing on bed, etc. Human activities which are considered to be abnormal include activities like falling on ground, fall from a height, fall from standing position. The overall methodology is explained via the following diagram:



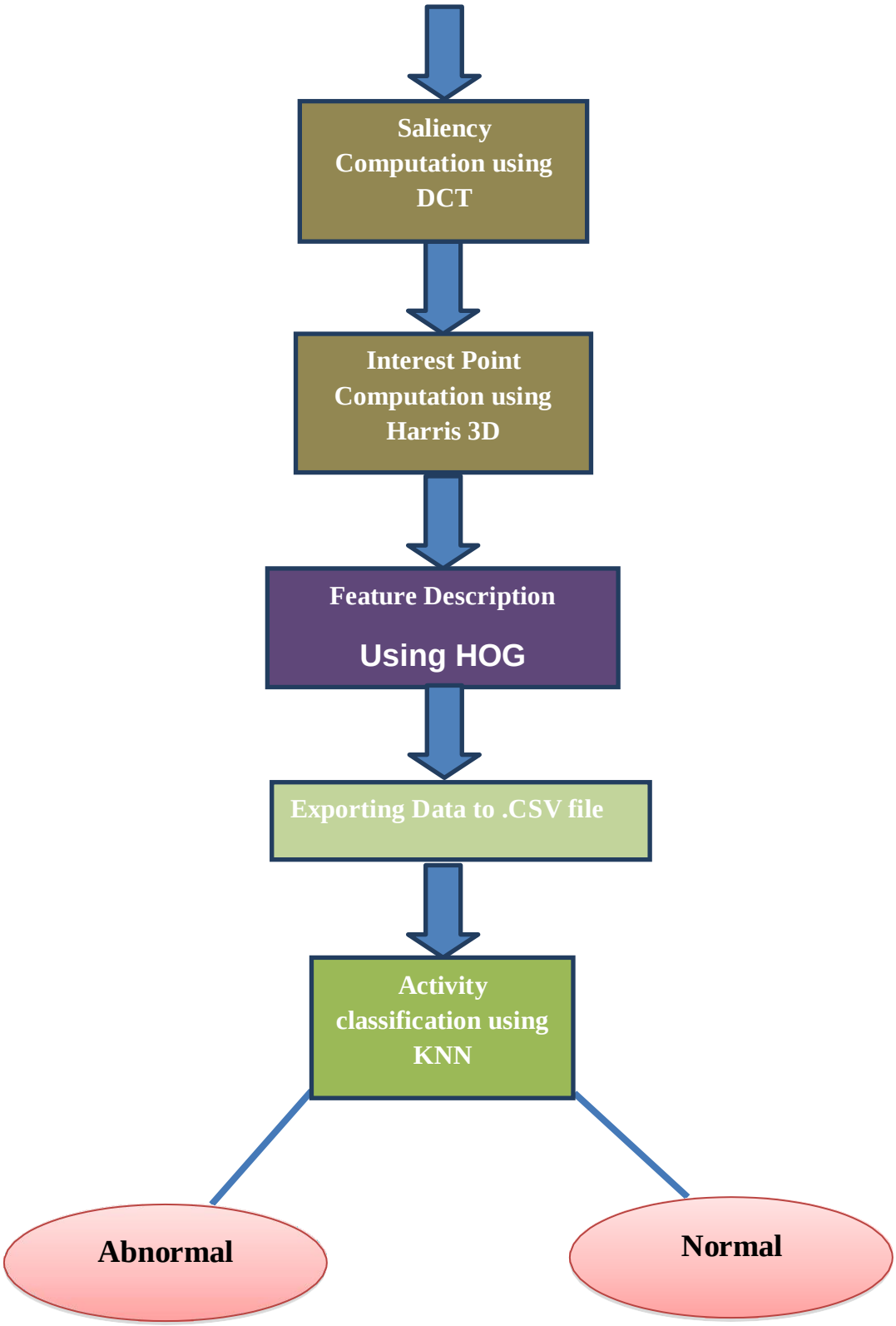


Fig 3.1 Methodology

Firstly, input videos sequences are converted to frames. Further, we find salient regions from each frame of input video. The saliency for each frame is computed using DCT algorithm for each frame of input video. Once salient region is computed for each frame the next task is to find out the interest points for each salient. We use Harris Stephens Algorithm to compute the interest point in each frame. Intermediate step of saliency computation helps us to remove unwanted interest point from each frame. Further, based on interest points, feature description is done using appropriate descriptor. Histogram of Oriented Gradients (HOG) is used as a feature descriptor in our proposed system. The features are returned in a 1-by-N vector, where N is the HOG feature length. The returned features encode local shape information from regions within an image. Finally, Decision tree is used as a classifier for classifying the activities into two classes whether normal or abnormal.

The detailed methodology is described in the following text.

3.1.1 Saliency Computation

Salient region is the one that gives some meaning full information or semantic information in an image. Part of an image is considered to be salient based upon the tendency by which it can be differentiated from its neighborhood. Saliency computation helps us to remove unwanted objects from the frame. This helps us in calculating only those interest points which are crucial for our work and avoiding false or misleading points. For our proposed system we need to detect salient object from the frame of input video. Hou et al. [23] proposed a method for saliency computation based upon Discrete Cosine Transform (DCT) algorithm. The detailed methodology for saliency computation is defined below-

Let say grey scale images can be decomposed as

$$i = x + a \tag{1}$$

The image signature of that image is as shown below:

$$\hat{x} = DCT(i)$$

$$ImageSignature(x) = sign(\hat{x}) \tag{2}$$

Given an image which can be decomposed as in (1), the support of x can be taken as the sign of the mixture signal i in the transformed domain and then computing the reconstructed image in

spatial domain using inverse DCT.

$$\hat{x} = \text{IDCT}(\text{sign}(\hat{x})) \quad (3)$$

Foreground of an image is assumed to be visually apparent and discernible with respect to its background, then we can form a saliency map m by smoothing the squared reconstructed image as in (4):

$$m = g * (x \quad \bar{x}) \quad (4)$$

where g is the Gaussian kernel. '*' is convolution operator and 'o' is Hadamard (entry wise) product operator.

3.1.2 Interest Points Detection

We used Harris-Stephens algorithm to compute the interest point in each saliency computed frame of input video sequence in our proposed system.

3.1.3 Feature Description

Histogram of Oriented Gradients (HOG) is used as a feature descriptor in our proposed system. MATLAB inbuilt function **[features, validPoints] = extractHOGFeatures(I,points**) is used for HOG feature descriptor where I refers to image and points refers to interest points which returns **features** which is 1-by-N vector where *N* is the HOG feature length. The returned features encode local shape information from regions within an image.

3.1.4 Training and Testing

The proposed model is trained and tested using Decision Tree Algorithm.

The main aim of Decision Tree is to create model that can predict the class or value by learning decision rules based on the data used for training the system. For predicting the class label, we start from root of the tree and then the record's attribute value is compared with value of the root attribute. On the basis of that comparison we follow that branch corresponding to that value and then jump to the next node.

CHAPTER 4
IMPLEMENTATION AND RESULT

4.1 Implementation

To Implement the discussed Methodology MATLAB Environment used for saliency computation, Interest point detection, Feature description and data preprocessing then preprocessed data is used in python environment for learning (Training) and testing.

4.1.1 In MATLAB environment:

Step 1: Select the dataset folder in the environment and select one video for further processing.

Step 2: For selected video process video by iterating over every frame of it , read frame in every iteration then calculate Saliency and interest point.

Step 3: Saliency calculation for frame is done by using image signature as image descriptor

Saliency computation Algorithm:

Step 1:- i=imread('example.jpg');

Step 2:- m=rgb2gray(i);

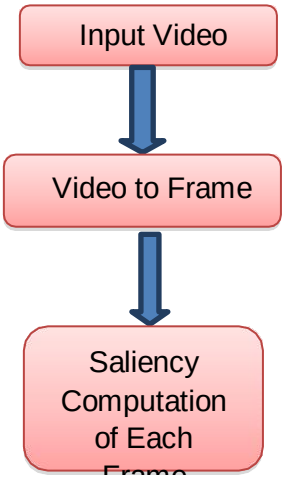
Step 3:- y=dct2(m);

Step 4:- z=idct2(sign(y));

Step 5:- a=imgaussfilt(z);

In the above algorithm Input image **i** taken then **i** is converted into grayscale by MATLAB in built function **rgb2gray(image)** which returns **m** grayscale of input image. This grayscale image **m** is currently in Spatial domain and Discrete cosine transform is used to transform domain of image to frequency so to achieve this MATLAB inbuilt Function **dct2(image)** is used which

returns **y** frequency domain image, Now to calculate **image signature** signum function of **y** is taken **sign(Y)** then this image signature is inversely transform into spatial domain by using inbuilt function **idct2(image signature)** which returns **z** inverse discrete cosine transform of image signature then to form a saliency map **a** by smoothing the squared reconstructed image by using **imgaussfilt(z)** which returns saliency map of image **.**



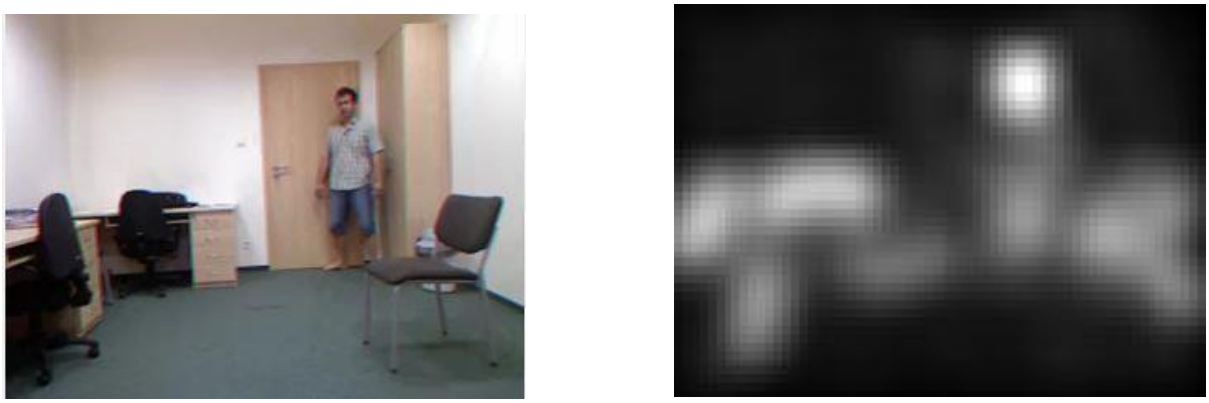


Figure 4.1: Frame before and after saliency computation

Step 4: Now saliency mapped image is used for interest point detection to calculate interest points, MATLAB inbuilt function **detectHarrisFeatures(Image)** is used which returns **c** corner points object which contains the information of spatial location of points.

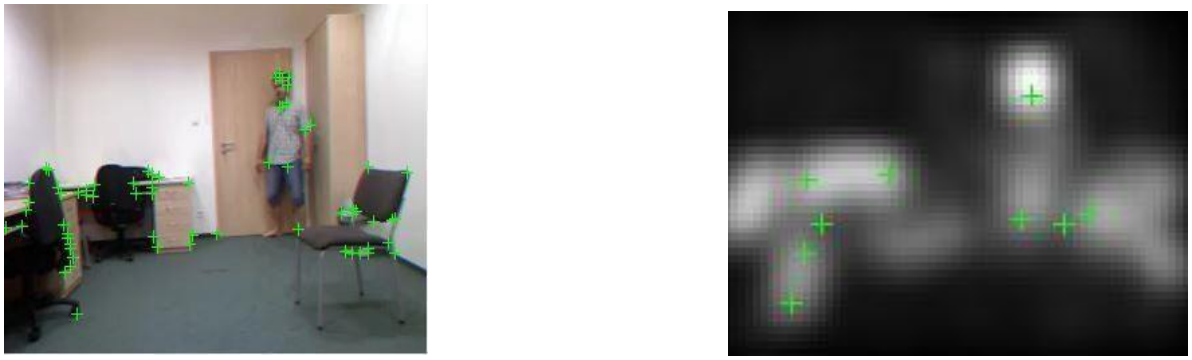


Figure 4.2: Demonstration of Intrest point with and without saliency in a video frame

Step 5: Now calculated corner points or interest points are passed to feature descriptor HOG which shapes the features into 1 x N vector where N is the HOG feature length, MATLAB inbuilt function **[features, validPoints] = extractHOGFeatures(I,points)** where I refers to image and points refers to interest points which returns **features** which is 1 by N feature vector.

Step 6: Now the mean of the feature vector for each frame is taken and stored in a csv file.

Step 7: So, for one video 1 x N vector is saved in a **.csv** file and now we take mean of the entire data again. Therefore, single video is now represented by data in one row.

This whole process is repeated for all videos in the training dataset and then this .csv file is imported for training and testing in python environment.

4.1.2 In python environment:

Step 1: Training Data obtained from the above steps is fitted in Decision Tree model for learning.

Step 2: Train test split module is used to partition the data for training and testing.

Step 3: After the model training testing phase starts, for training 85% of the dataset is used and the rest of 15 % is used for testing purposes. In testing, the feature vector of video is given as input and model output 1 for abnormal and 0 for normal activity.

Step 4: For accuracy of model predicted outcomes is compared with the actual output.

4.2 Results

4.2.1 Dataset

The dataset involving abnormal human activities relating to older individual was difficult to find so a genuine dataset such as UR- Fall dataset was used to evaluate and test the proposed system. The detailed description of dataset used is given below

The dataset used in our system consist of 63 videos. Out of the 63 videos, 30 videos are of fall activity while 33 videos are of normal day to day activities. RGB recording are used for our proposed system. In our system, fall activity is considered to be abnormal event while events of day to day life are considered to be normal event. Normal activities are like walking, sitting on chair, jumping, cleaning, etc. while abnormal activities are the one like falling on ground, falling from chair, etc. The videos of dataset are recorded in home environment.

4.2.2 Result

UR Fall dataset is used for experimentation purpose. For classification, Decision Tree classifier is used. To partition the data for training and testing, 'Train test split' method is used. 85% of the data is used to train the machine and the remaining 15% data is used for testing.

Accuracy of 90% is obtained on testing data.

CHAPTER 4
CONCLUSION

This work proposes a method for the anomaly detection from video in the home environment. Human activities which are considered to be normal include activities like sitting, walking, cleaning, jumping, relaxing on bed, etc. Human activities which are considered to be abnormal include activities like falling on ground, fall from a height, tumbling down from standing position.

The proposed framework targets anomaly activity detection in the home condition. This is a genuine target towards building of an emotionally supportive system helping old individuals which can provide them sense of safety and security living alone. The system firstly involves saliency computation followed by feature exaction and description from the input video sequence. After feature extraction and description Decision Tree (DT) is used to classify the detected human activity as normal or abnormal event. The main goal is to identify irregular action effectively and results obtained are highly encouraging. The data is partitioned as 85% of it used for training the system while remaining 15% used for testing the same. The system gives an accuracy of 90%. Further improvements will be made in the future to increase the accuracy and reduce computational cost for the betterment of the system.