

HUMAN ACTION RECOGNITION USING MULTI-CHANNEL SPATIO-TEMPORAL INTEREST POINTS

A

Dissertation report

*submitted in partial fulfilment of the
requirements for the Award of Degree of*

Master of Technology

in

Artificial Intelligence & Artificial Neural Networks

by

Ayush Purohit

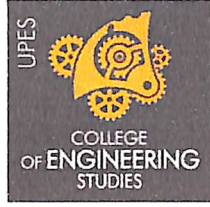
Under the Esteemed Guidance of

Dr. Venkatadri Marriboyina



**Centre for Information Technology
University of Petroleum & Energy Studies
Bidholi, Via Prem Nagar, Dehradun, UK**

April – 2016

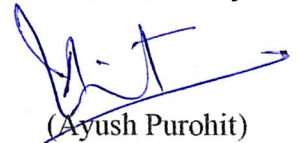


The innovation driven
E-School

CANDIDATE'S DECLARATION

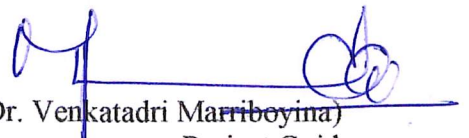
I hereby certify that the project work entitled " Human Action Recognition using Multi-Channel Spatio-Temporal Interest Points" in partial fulfilment of the requirements for the award of the Degree of MASTER OF TECHNOLOGY in ARTIFICIAL INTELLIGENCE AND ARTIFICIAL NEURAL NETWORK and submitted to the Department of Computer Science & Engineering at Center for Information Technology, University of Petroleum & Energy Studies, Dehradun, is an authentic record of my work carried out during a period from January, 2016 to April, 2016 under the supervision of Dr. Venkatadri Marriboyina, Assistant Professor, University of Petroleum and Energy Studies.

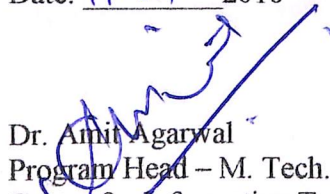
The matter presented in this project has not been submitted by me for the award of any other degree of this or any other University.


(Ayush Purohit)
R102214002

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date: 11.04. 2016


(Dr. Venkatadri Marriboyina)
Project Guide


Dr. Amit Agarwal
Program Head – M. Tech. AI ANN
Center for Information Technology
University of Petroleum & Energy Studies
Dehradun – 248 001 (Uttarakhand)

ACKNOWLEDGEMENT

I wish to express our deep gratitude to our guide Dr. Venkatadri Marriboyina, for all advice, encouragement and constant support he has given us throughout our project work. This work would not have been possible without his support and valuable suggestions.

I am heartily thankful to my course coordinator, Mr. Vishal Kaushik, for the precise evaluation of the milestone activities during the project timeline and the qualitative and timely feedback towards the improvement of the project.

I sincerely thank to our respected Program Head of the Department, Dr. Amit Agarwal, for his great support in doing our project in Computer Vision & Image Processing at CIT.

I am also grateful to Dr. Manish Prateek, Associate Dean and Dr. Kamal Bansal Dean CoES, UPES for giving us the necessary facilities to carry out our project work successfully.

We would like to thank all our friends for their help and constructive criticism during our project work. Finally we have no words to express our sincere gratitude to our parents who have shown us this world and for every support they have given us.

Ayush Purohit

R102214002

ABSTRACT

Human Action Recognition is one of the imperative exploration regions in computer vision and image processing field which can be seen as a scaffold for machines to comprehend human non-verbal communication. The aim of the action recognition is to perform automated analysis of human events from video data. Automatic interpretation of human actions are required when developing vision based systems for visual navigation, monitoring systems, crowd management, and systems that implicate interactions between humans and machine interfaces. In this work, the problem of action representation to detect and recognize action activities is addressed. A local feature based model is discussed which uses multi-channel spatio-temporal key point features as primitives when representing and recognizing actions. For photometric representation of image data, we implemented HOG descriptor commonly known as Histogram of Oriented Gradient, with STIP feature detectors. STIP effectively captures the local structure in spatio temporal dimensions of the video sequence. We used '*bag-of-visual words*' for representing video sequences. Experiments were performed on WEIZMANN and KTH datasets for eight action categories – Walking, Running, Jumping, Boxing, Waving, Jogging, Clapping, and Cycling.

TABLE OF CONTENTS

S. No.	Contents	Page No.
	<i>Candidate's Declaration</i>	<i>i</i>
	<i>Acknowledgement</i>	<i>ii</i>
	<i>Abstract</i>	<i>iii</i>
	<i>Table Of Contents</i>	<i>iv</i>
	<i>List Of Figures</i>	<i>vii</i>
	<i>List Of Tables</i>	<i>viii</i>
1.	INTRODUCTION	1-8
1.1	What is Action?	1
1.2	Action Recognition in Computer Vision	1
1.3	Historic Overview	1
1.4	Motivation	2
1.5	Modern Applications	4
1.6	Human Action Recognition Framework	5
1.6.1	Feature Detector	6
1.6.2	Feature Descriptor	6
1.6.3	Video Representation	6
1.6.4	Classification	6
1.7	Objective	7
1.8	Requirement Analysis	7
1.8.1	Functional Requirements	7
1.8.2	Non-Functional Requirements	8
2.	LITERATURE SURVEY	9-15
2.1	Background Study	9
2.2	Literature Survey	10
3.	SYSTEM ANALYSIS	16

3.1	Existing System	16
3.2	Proposed System	16
4.	SYSTEM OVERVIEW	17-22
4.1	System Design	17
4.1.1	Block Diagram	17
4.1.2	System Architecture	17
4.1.3	Flowchart	18
4.1.4	Activity Diagram	19
4.1.5	Sequence Diagram	20
4.1.6	Collaboration Diagram	21
5.	COMPUTATIONAL THEORY	22-32
5.1	Low-level Representation	22
5.1.1	Representation of Phase	22
5.2	Preprocessing Steps	23
5.2.1	Median Filtering	23
5.3	Feature Detectors	23
5.3.1	What is a feature? What constitute a feature?	24
5.3.2	Harris Detector	25
5.3.3	Harris 3D Detector	27
5.3.4	Gabor Detector	28
5.4	Feature Descriptors	29
5.4.1	HOG Descriptor	30
5.4.2	HOG 3D Descriptor	32
6.	REPRESENTATION	33-35
6.1	Bag-of-Features	33
6.1.1	Learning the visual vocabulary	34
6.1.2	Mapping the keypoints to visual words	34

7. SUPERVISED LEARNING	36-38
7.1 SVM for linearly separable dataset	36
7.2 SVM for linearly separable dataset	38
8. CONCLUSION	39
References	40-45
Appendix I: Project Guide	46-48

LIST OF FIGURES

S.N. Contents	Page No.
1. INTRODUCTION	
1.1 Fragments from da Vinci's sketchbooks	2
1.2 Etienne-Jules Marley Chronophotographic experiment.	3
1.3 Geometrical Representation of human body	3
1.4 Gunnar Johansson experiment on 2D Motion Perception	4
1.5 Human Action Recognition Framework	5
2. LITERATURE SURVEY	
2.1 Classification of human activity recognition	11
2.2 Computation of SIFT feature descriptor	12
4. SYSTEM OVERVIEW	
4.1 Block Diagram of Human Action Recognition	17
4.2 System Architecture of proposed model	17
4.3 Flowchart	18
4.4 Activity Diagram	19
4.5 Sequence Diagram	20
4.6 Collaboration Diagram	21
5. COMPUTATIONAL THEORY	
5.1 Aperture problems for different image patches	24
5.2 Three auto-correlation surfaces.	25
5.3 Uncertainty ellipse corresponding	27
5.4 HOG Feature Extraction	31
6. REPRESENTATION	
6.1 Bag of Features Model	33
6.2 Representation of Bag-of-Features	35
7. SUPERVISED LEARNING	
7.1 Number of possible linear classification between two classes	36

LIST OF TABLES

S. No.	CONTENTS	Page No.
1.	INTRODUCTION	
1.1	Basic System Requirements	7
2.	LITERATURE SURVEY	
2.1	Comparison of Different Representations.	14
2.2	Comparison of existing Local Feature Based Approach.	15

1. INTRODUCTION

Human has remarkable ability to analyze human activities absolutely from visual data. We can confine individuals and items, track arranged movements and examine human-object communications to comprehend what individuals are doing and even induce their aims. The objective of Human Action Recognition (HAR) is to anticipate the mark of the activity of an individual or a gathering of individuals from a practical situation. This is a complex difficulty in computer vision where numerous issues are currently under research, including the joint demonstrating of behavioral prompts occurring at various time scales, the intrinsic instability of machine noticeable confirmations of human conduct, the common impact of individuals included in connection and essential part of motion in human conduct understanding [1].

1.1 What is Action?

What constitutes an action—is difficult to define. Even though there is a great demand for a specific and established action/activity hierarchy, there is not any recognized action hierarchy in computer vision till now. Lan et al. [2] define action and activity as,

- Activity to indicate a straightforward, singular action performed by a solitary individual.
- Action to allude to a more unpredictable situation that includes a gathering of individuals.

On the other hand, Giagon et. al. [3] decomposes actions into sequence of key atomic action units, each called an *actom*. An actom is a short instant motion, computed by its focal worldly area around what discriminative visual data is available [3].

1.2 Action Recognition in Computer Vision

Recognition is defined by the trial to figure out if or not input information resides or resembles some specific object, feature, or activity. In the prevailing gimmick, recognition of human behavior is imperative, but arduous job. Action recognition is imperative by seeing it as a resilient and visceral approach to promote more human-centered forms of man-machine interaction [4]. However, the effort required to function these recognition varies and are very complex due to wide range of sub goals such as unique identification of body parts, recognizing gestures and classifying them.

Real time human action recognition is an imperative examination study under the field of computer vision and image understanding which has extensive applications such as in man-machine interaction, sociology research, video surveillance, etc. Be that as it may, these applications request frameworks to move in unconstrained situations which require power against enlightenment, perspective, camera edge, movement and so on. This work accentuates on low-level representations for perceiving human activities in video. Low-level activity acknowledgment methodologies are regularly in view of Spatio-Temporal Interest Points (STIPs) [5] [6]. An assortment of photometric representations for STIP discovery and depiction are utilized for upgrading low-level ways to deal with activity acknowledgment [7]. Here, picture arrangements are spoken to by descriptors that are removed locally around STIP recognitions. This dissertation covers a brief study of previous work, Computational Aspects, Local Spatio-Temporal Features, Motion Descriptors, Evaluation & Performance, and Conclusion & Future Work.

1.3 Historic Overview

Human movement concentrates on providing human shape and features and by applying viewpoint geometry [8]. Early research persuaded by human representations in Artistic representations and biomechanics. Leonardo Da Vinci in one of his sketchbooks he stated that [9] – “It is crucial for a painter, to wind up absolutely acquainted with the life systems of nerves, bones, muscles, and ligaments, such that he comprehends for their different movements and hassles, which ligaments or which muscle causes a specific movement ”.

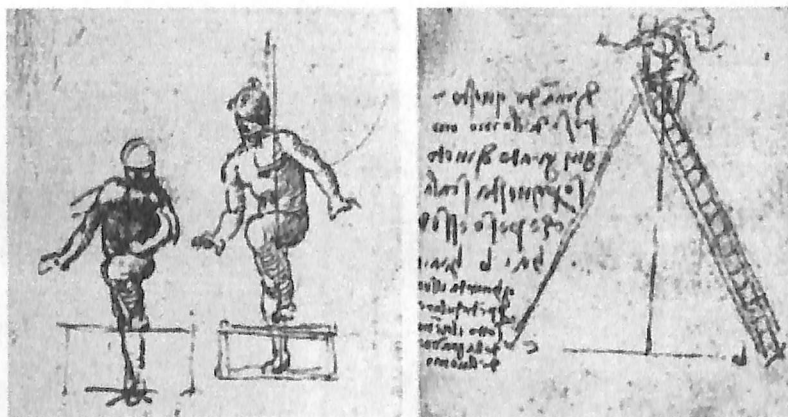


Figure 1.1: Some early sketches from Leonardo da Vinci's. [9]

Galileo Galilei constructed the human model, which was used by G. A. Borelli to analyze geometrical behaviour [10]. He deduces mathematical principles to understand the human body and concluded that bones serve as levers and muscles activities. His study involves muscle examination and a numerical talk of motions, for example, walking or jogging.

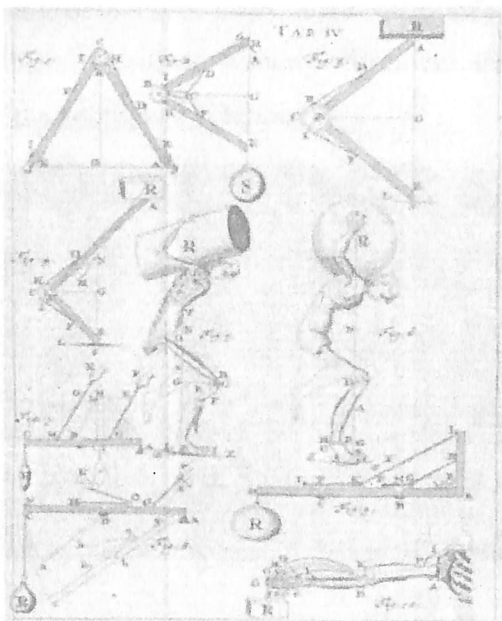


Figure 1.2: Geometrical Representation of human body by Giovanni Alfonso Borelli [10]

In late 18th century, Etienne-Jules Marley [11] made a Chrono photographic tests compelling for the rising field of cinematography. Later on, Eadweard Muybridge [12] designed a model for showing the recorded arrangement of sequence of pictures. He spearheaded movies and connected his system for motion studies.

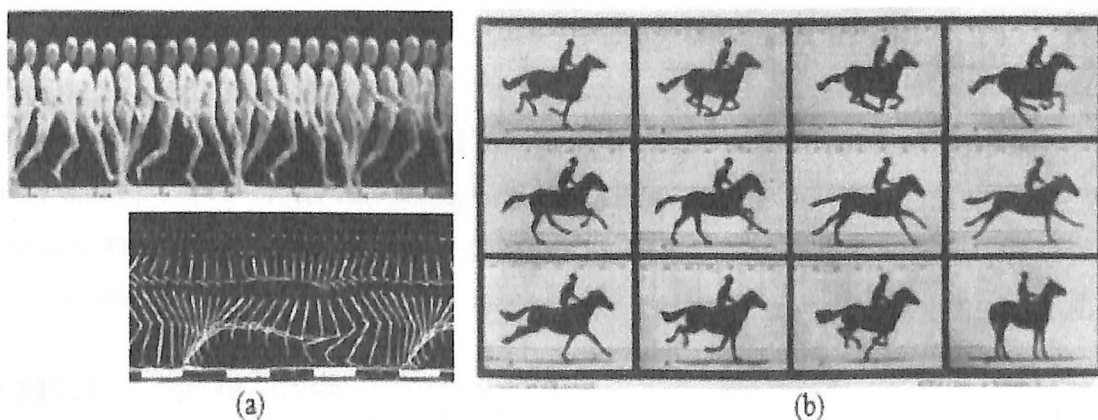


Figure 1.3: a) Etienne-Jules Marley Chronophotographic experiment [11]. b) Eadweard Muybridge's motion pictures for motion study [12].

Gunnar Johansson in [13] pioneered contemplations on the utilization of picture groupings for a modified human movement examination. "Moving Light Displays" (LED) empower distinguishing proof of natural individuals and the sexual orientation and propelled numerous works in computer vision. The figure 4, represents the Gunnar Johansson Experiment's images on 2D motion perception [13] at different time steps. It was deduced that programmed sort of visual information treatment is generally critical. Numerically, these spatio-temporal relations in the proximal jolt design decide the perceptual reaction.

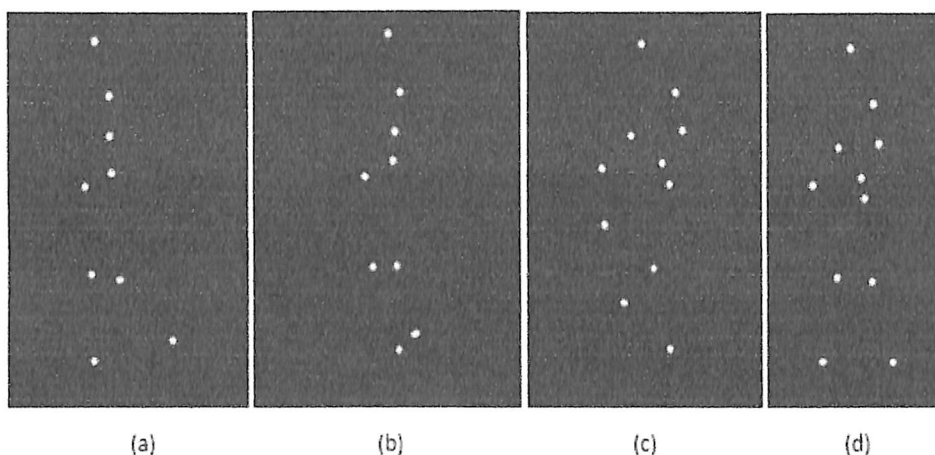


Figure 1.4: Gunnar Johansson experiment on 2D Motion Perception [13]

1.4 Motivation

We ask the question as to whether it is feasible to derive information about a complex scenario of *who is doing what* from just a single video scene. For example, consider any video where a person is doing something. Based on the context of the scenario and person movements, humans can easily interpret what the video is about. As discussed earlier in Gunnar Johansson's Experiment in 1973 on 2-D motion perception, using only few selected key points on the image sequences one can interpret the human motion. Inspired from this work, it was possible to recognize the human motion and activities based on similar key points or interest points using temporal image sequences. Thus, combining both the context and human motion, a system can be developed to derive the scenario from a video or in real time.

1.5 Modern Applications

Understanding actions or activities from temporal images is an important however arduous to carry out. The field of human model representation and recognition is moderately old, yet there

are only limited and countable number of real-life applications. Perceiving the character of people and the activities, exercises and practices performed by one or more persons in video arrangements is vital for different applications. Some of the achieved goals and applications of action recognition include:

- Motion Capture and Animation
- Face Analysis
- Surveillance System
- Search and Indexing
- Obstacle Avoidance
- Mixed Reality
- Unusual Activity Detection
- Medical Applications
- Consumer Electronics and Social Applications

1.6 Human Action Recognition Framework

This area depicts the activity recognition system that is utilized as a part of this postulation. The system concentrates on a subset of the strategies introduced in the previous segments. The representation used is local keypoint features and classified using Bag of Features (BoF) framework, as it is more promising bearing for portraying the more progressed datasets. Regulated and unsupervised characterization will be looked at, to perceive how preparing information influences the acknowledgment rate.

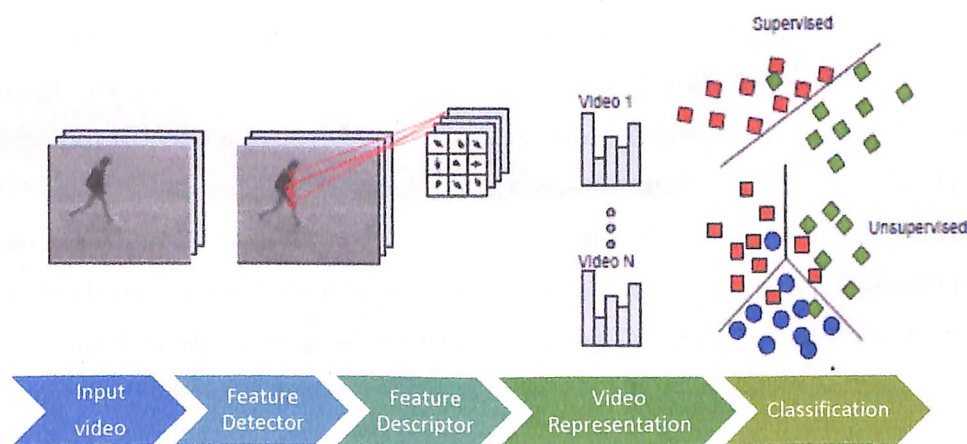


Figure 1.5: Human Action Recognition Framework

The system comprises of four primary parts: feature detector, feature descriptor, video representation and classification. The diverse parts of this structure can be supplanted with various calculations to shape distinctive mixes and in this way deliver distinctive results.

1.6.1 Feature Detector

The initial step is to distinguish interest points in the video, which are the positions where the component descriptors are figured. These focuses ought to in a perfect world be situated at spots in the video where the move is making place.

1.6.2 Feature Descriptor

The feature descriptor encodes the data in the region of an interest point into a representation suited for representing the activity. The feature descriptor ought to preferably be invariant to changes like orientation, scale and enlightening to be able to match features across different kinds of videos.

1.6.3 Video representation

The arrangement of local feature descriptors in a image sequence must be combined into a representation that empowers the correlation with different recordings. The most well-known strategy is the Bag-of-Features representation, where the spatial and temporal regions of the features are disregarded. Different strategies tries to consider the relationship between the feature components.

1.6.4 Classification

The classification step can be: unsupervised, semi-directed or supervised. In unsupervised methodology, we expect that we don't have the foggiest idea about the names of any of the recordings. The videos are gathered in a cluster in view of their resemblances. The amount of gatherings used for unsupervised learning is still a growing area in research, or can be dynamic where unmistakable fragments of recordings thinks about to different semantically ramifications. In semi-controlled gathering there is some earlier learning about the recordings, which can involve a couple stamped examples, e. g. an impediment saying that case an and b are from different classes without providing the mark.

1.7 Objective

The existing STIP systems maneuvers on luminance illustrations of the spatial information. As results, information may be lost when either luminance or chromatic representations are considered in isolation. The procedure utilizing movement directions includes following and thick multi-scale optical stream calculation for which the related computational multifaceted nature included is much higher than STIP-based methodologies.

- The aim of this paper is to address behavior and representation of human activities in assorted and real-time datasets.
- To reformulate Space-Time Interest Point detectors to join various photometric diverts notwithstanding picture intensities.

1.8 Requirement Analysis

The Project requires some basic minimum resources so that it can function properly. The Human Action Recognition System has Functional as well as non-functional requirements that have to be full filled in order to make the application run properly. The program has to full fill the nonfunctional requirements to maintain the quality of its output and overall. All these requirements have been listed below briefly.

Table 1-1: Basic System Requirements

Operating Systems (Windows Based)	Windows 7 (x86 & x64 architecture) Windows 8 (x86 & x64 architecture) Windows 10 (x64 architecture)
Software Used	MATLAB 2010a
Suggested Architectures	32-bit (x86) 64-bit (x64)
Other Obligations	1.6 GHz of processor or more. 2 Giga Bytes of RAM Minimum 10 GB of available HD space.

1.8.1 Functional Requirements

The different useful necessities of the framework can be abridged as follows:

- MATLAB 2010a version or above is required in order to run the system. However, with change in updated versions might lead to bugs and errors due to unavailability of inbuilt functions.

- For using the short keys, the operating system support is required.

1.8.2 Non Functional Requirements

Performance:

- The inserted picture created ought not to contain any distortion. Additionally the application ought to be secure to measurable and examination investigation.

Reliability:

- The capacity of the created framework is to carry on reliably in a user acceptable way when working within the environment for which the framework was expected.
- The product should not crash under any circumstance such as client entering not valid arguments, client attempting to stack unsupported documents and so forth. It ought to demonstrate fitting message for each client produced message.
- Infant mortality:
 - Given a large complex video size as an input, system might fail due to limited processing capabilities.

Portability:

- The extent to which programming running on one stage can without much of a stretch be changed over to keep running on another stage. E.g., number of target proclamations (e.g., from Unix to Windows).

2. LITERATURE SURVEY

This thesis covers various fields in computer vision which comprised of action-based recognition, representation of human actions, and recognition in terms of neighboring features. In this chapter, we discussed some of the main contributions in these areas with emphasis on action representation and keypoint-based motion recognition. In the next chapter we present action representation and describe methods for motion estimation with close relation to the methods used later in this work.

2.1 Background

High level approaches for unconstrained action recognition desires for demonstrating video frame series to recognize high level motion features, which can be developed on local features. These features typically contemplate novel video representations depends on neighbor illuminant and chrominance keypoints. High-level approaches often require high computing power, which uses computationally exhausting video processing processes but are superior to low-level recognition access with respect to recognition ratio. Low-level approaches are reasonably straightforward, moderately simple to actualize and possibly scanty and proficient. As local keypoints are inherently robust against noise and fluctuating background for example, impediment and disorder. Therefore in this dissertation, we thrust on low-level representations for human action recognition in such video stream.

Low-level activity recognition methodologies are generally based on space-time keypoints or spatio-temporal feature points. Here, frame arrangements are depicted by descriptors which are derived regionally around interest points. The descriptors are vector quantized (VQ) in light of visual vocabulary, training & perception operates on these quantized descriptors, involving the bag-of-features framework. In a static frame, color descriptors surpass magnitude based descriptors in an assortment of image coordinating and object recognition tasks [45], [46]. In the spatial domain, multi-channel illuminance independent expression of keypoint detectors are stated in [47][48][49]. For descriptors, multichannel expression proposed in [45], [46] presented diversified color interest points variants conspicuously, OpponentSIFT significantly boost the overall outcome. However, using such methodologies might led to information loss if intensity or chromatic representations are used separately. STIP feature extraction require low computational power than using motion trajectories which includes tracing and clustered multi-scale optical flow

calculation. To overcome this issue, an invariant photometric description is proposed which is integrated in multi-channel STIPs feature extraction.

2.2 Literature Survey

Action recognition has been explored in many different ways. At present there are various different strategies defined to perform action recognition. In [14], the actions are classified based on full body actions. In this, the classification of action is studied on the basis of spatial and temporal structure of body movements. In [15][16], the actions are classified based on hierarchal approach to differentiate the human action recognition problems. This classification human action are classified into two ways i.e., single layered approach and hierarchal approach. Based on the different approaches, different strategies are followed to recognize human actions which are based on different representation and learning methodologies. Volker Kruger in [17], categorized and human actions on the basis of scene, full body and with/without using body parts.

In [40] [41], hierarchal approach is reviewed to differentiate the human action recognition problems. In the review, human action are classified into two ways i.e., single layered approach and hierarchal approach. Based on the different approaches, there are further classification to recognize human actions which are based on different representation and learning methodologies as shown in figure 1.

Actions can be categorized based on human body model based method, holistic methods and local feature method. The *human body model based methodology* of action recognition makes utilization of the bits of data which is extracted from the body parts of the human beings. The model mainly comprises of the two necessary principles [18]: The appearance of the body part of the human being in the image or video [19]. With the assistance of moving light display, the people can perceive the motions by depicting the movements of the fundamental joints of individuals. In the study performed by Johansson [42], established that the range between ten to twelve of the moving light displays which are in movement mixes in proximal boost have evoked the impression of human body motions. An extended approach of epipolar geometry that is the geometry of dynamic scenes is indicated in [43]. The system has been implemented in variety of sequences. The theory of chaotic systems in [44] have been used for action recognition. A complete set of new features are implemented in order to classify human body gestures that are dynamic in nature.

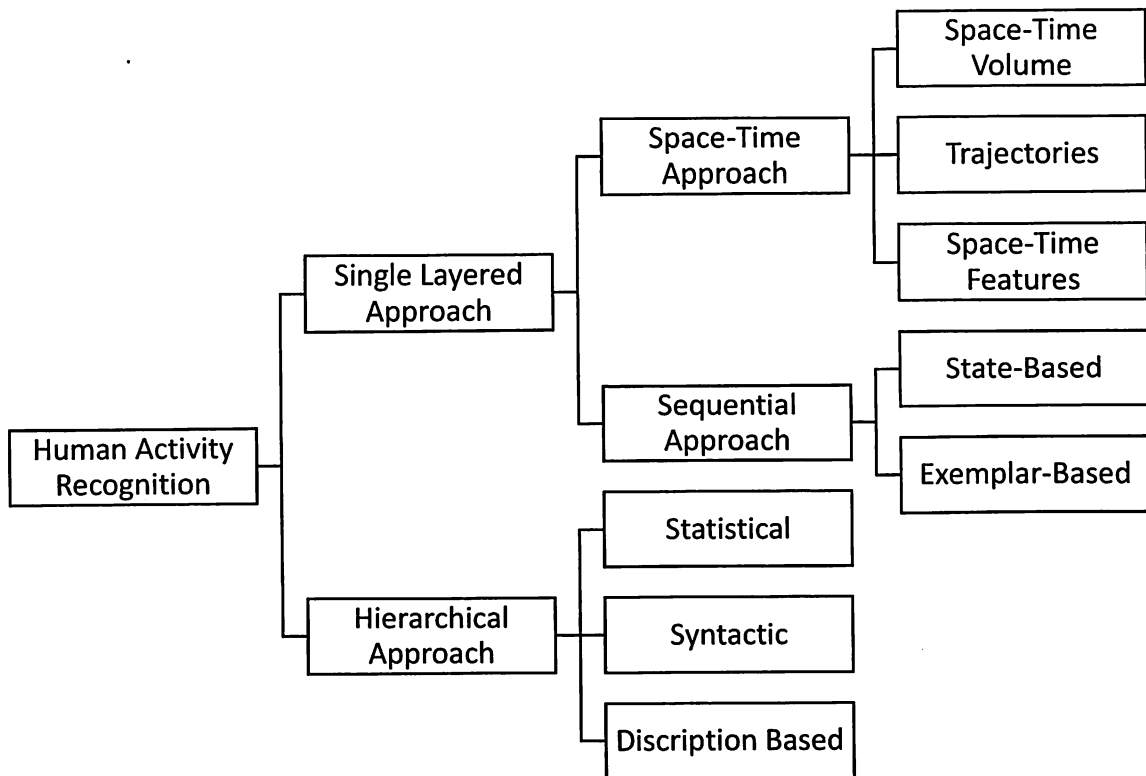


Figure 2.1: Classification of human activity recognition [40].

The *holistic based technique* for action recognition uses the data retrieved on individual's localization in real time datasets [20]. The silhouette and in addition the shape based components serves as one of the principle characteristic that have been utilized to represent the human dynamics and the body structure for recognition of actions in videos [21][22]. The fundamental point of the holistic based strategy is not to utilize any of the data provided by the body parts of the people. The *local feature based approach* implements the usage of local features for the process of action recognition [23]. The most important principle involved is that there is no need of the data on localization of people or data of the model of human body. This technique has been a standout amongst the most exceptionally examined area in the field of action recognition.

Interest Points provide a brief representation of image content by portraying nearby parts [24] within the context thus offers robustness to the clusters and intra-class variations. A low level action recognition approach are often based on STIPs, however due to insufficient photometric invariance of intensity channel [25], these systems are highly susceptible to illuminance. In a spatial domain, color descriptors outclass illuminance factors in diverse object recognition errands [25] [26] because of its superior balance between photometric invariance and discerning power.

In a non-temporal domain, multi-channel photometric invariant formulation of feature detectors are proposed in [27][28][29]. These papers are concluded by expanded repeatability, entropy and object recognition results when contrasted with intensity based detections.

Amid the feature detection, extracting the frame's keypoints is implemented which are prone to match well in different frames. Some of the widely used STIP detectors include the Harris 3D detector [23], cuboid detector [30], Hessian detector [31], and Gabor Detector. At the feature description process, the area around identified keypoint locations is changed over into a more minimized and stable descriptor that can be coordinated against different descriptors. Such space-time keypoint descriptors include Cuboid descriptor [30], HOGHOF [32], HOG 3D [33], and SURF descriptor [34]. A spatio-temporal interest points based detector proposed in [17], which recognizes human actions to extract human expressions. The system uses cuboid descriptors to detect human activities and perform the classification to accomplish the person behavior in a video.

Chen and Hauptmann in [35], introduced MOSIFT algorithm to compute STIPs in a real time dataset. The MoSIFT descriptor which is an extended version of original SIFT detector (as shown in fig. 2.2) evaluates spatial domain and the temporal domain features independently and developed to be robust via grid aggregation of both histograms. A new STIP detector is proposed in [30] which is used to characterize the human face and their actions. The proposed system is found to work well in varying illuminance conditions to produce desired results.

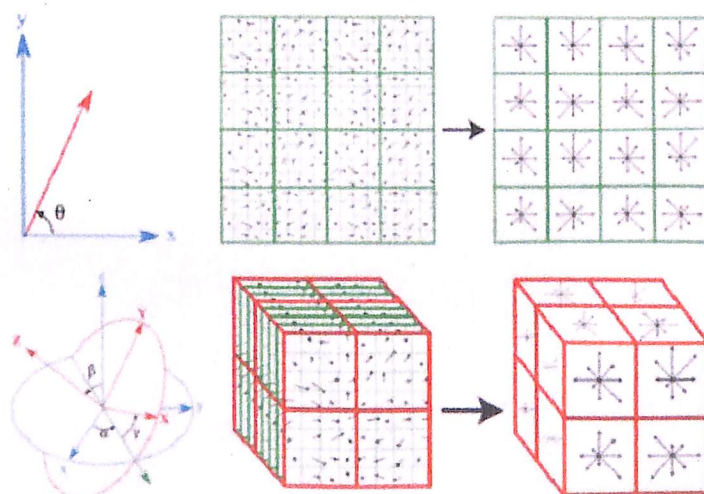


Figure 2.2: Computation of the 2D (Top) and 3D (Below) SIFT feature descriptor. [58]

Cheung et al. [56] proposed video epitomes, a space-time cubes from a specific video by a generative model. The trained framework is a reduced representation, consequently this technique is idle for temporal frames interpolation, yet not for acknowledgment. Plinio Moreno et. al. [57], addressed the issue of selecting the low-level features to describe the interest feature points and proposed a methodology based on Information Diagram concept, by optimizing the filter response in the filter parameter space.

Xinxiao Wu et. al. [36] proposes spatio temporal context distribution feature for recognizing human activity in a temporal framework. To identify the dispersion of spatio-temporal feature set in video frames, multiple GMMs are implemented from different space-time scales obtained from global GMM using MAP. An appearance and motion cues based human action recognition is implemented in [37]. The proposed methodology uses local features to train their VGG – 16 layered CNN model. In [38][39], spatio-temporal features are used as tracking features. In such models the basic idea is to extract the local keypoint features from a sequence of image or video and using them to track the object or human.

Table 2-1: Comparison of Different Representations.

Models	Pros	Cons	Findings
Parametric Representation	Psychological approach. Industrial applications in Medical science and animations.	Finding parts of body, parameter estimation for optimization, depends on tracking.	These approaches are used just in controlled settings and are not applied to realistic actions.
Global Representation	Invariant to color and texture. Representation is easier than parametric models. Suitable for recognition actions within a limited area.	Depend to background subtraction or optical flow computations. Complex and expensive to compute. Sensitive to view point.	
Local Representation	Hybrid of parametric & global representations. Good results can be obtained with low computing cost. Robust to clutter.	Do not model geometrics of action. Heavy feature matching is required.	Used for uncontrolled setting but does not handle camera motion.

Table 2-2: Comparison of existing Local Feature Based Approach.

	Author	Feature Detectors	Feature Descriptors	Recognition Rate
Local Feature Based Action Recognition	Heng Wang et. al. [50]	Harris 3D, Cuboid, Hessian, Dense	HOG 3D, HOG, HOF, ESURF, Cuboid	KTH- 92.1% UCF- 85.6% Hollywood2- 47.4% UCF Sports- 85.6%
	Imran N. Junejo et. al. [51]	-	SSM	CMU MoCap- 95.7% Weizmann- 92.6% IXMAS- 74%
	Ivo Everts et. al. [52]	Multi-channel Color STIPs	UCF11, UCF50	78.6%, 72.9%
	P. Dollar et. al. [30]	Cuboid	Cuboid	KTH- 81.50%
	M. Chen et. al. [35]	SIFT	MoSIFT	KTH- 95.83%
	Ivan Laptev et al. [53]	Harris Operator	HoG, HoF	KTH-91.8%
	Nieble et al. [54]	Cuboid	Cuboid	KTH-83.33%
	Schuldt et al. [55]	Harris (scale space representation)	HistLF, HistSTG	KTH-71.72%

3. SYSTEM ANALYSIS

3.1 Existing System

The recognition of human motion has an extensive variety of real time applications, for example, reconnaissance, perceptual interfaces, understanding of game occasions, and so on. In spite of the fact that there has been much progress on human action recognition in the course of recent decades, action analysis still stays remains a challenging issue. In terms of high level analysis, the recent research concentrates for the most part fall under two noteworthy classes of methodologies. The previous studies, for the most part describes the spatiotemporal conveyance produced by the human movement in its continuum.

Existing STIP-based action recognition approaches uses image intensity to represent motion in a videos. Multichannel expression proposed in [59] presented diversified color interest points variants conspicuously, OpponentSIFT [60] significantly boost the overall performance. However, such methodologies are perceptive to noisy photometric development, for example, shadows and highlights. Additionally, it is not efficient as it might led to information loss if intensity or chromatic representations, if used separately.

Human Action Recognition strategies experience the ill effects of numerous disadvantages, which incorporate

- The failure to adapt to incremental recognition issues;
- The necessity of a concentrated dataset to acquire better recognition ratio;
- The powerlessness to perceive concurrent numerous activities; and
- Trouble in recognizing actions in a complex video sequence.

3.2 Proposed System

To the mentioned issues, we addressed the problem by using color STIPs. These are multichannel formulations of spatio-temporal keypoints detectors & descriptors, we considered independent representations which are produced from color space. To improve the performance, the color STIPs are considered to be reliable and effective even if there are illumination variations and other disturbances in the image.

4. SYSTEM OVERVIEW

4.1 System Design

Vision based study, depends in transit individuals analyze data with respect to the context in their surrounding environment, nonetheless it is presumably the most problematic to device and make it globally acceptable. A few distinct methodologies have been tried as such.

- The process requires to capture the video using some source then extracting the interest point features. These featured structures are given as an input to a classifier for recognizing the human gesture or action.

4.1.1 Block Diagram

A block diagram is a structural representation of any framework in which the fundamental parts or limits are addressed by blocks that exhibit the associations between the two modules. Square diagrams are ordinarily used for bigger sum, less point by point delineations that are proposed to clarify general thoughts without sensitivity toward the unobtrusive components of use. Parity this with the schematic diagrams and configuration traces used as a piece of electrical outlining, which show the execution purposes of enthusiasm of electrical sections and physical improvement.

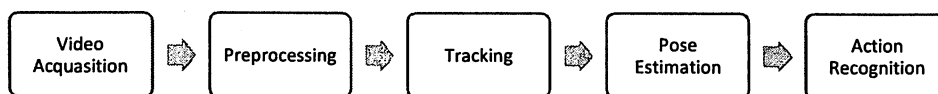


Figure 4.1: An approach towards Human Action Recognition

4.1.2 System Architecture

The system architecture for the proposed work is given in figure 4.2.

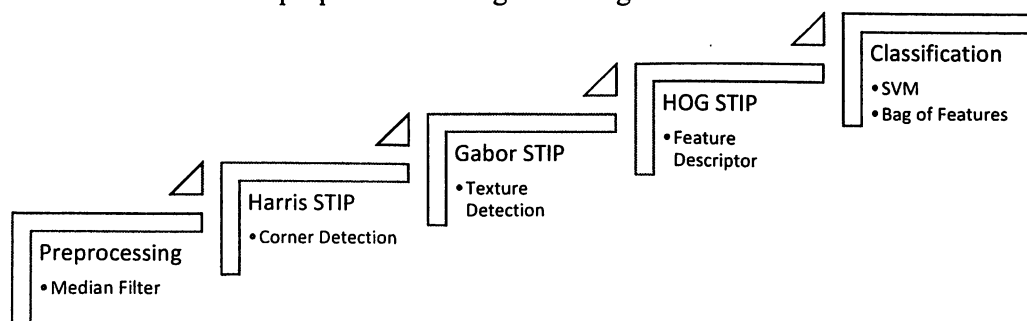


Figure 4.2: System Architecture of proposed model.

4.1.3 Flow Chart

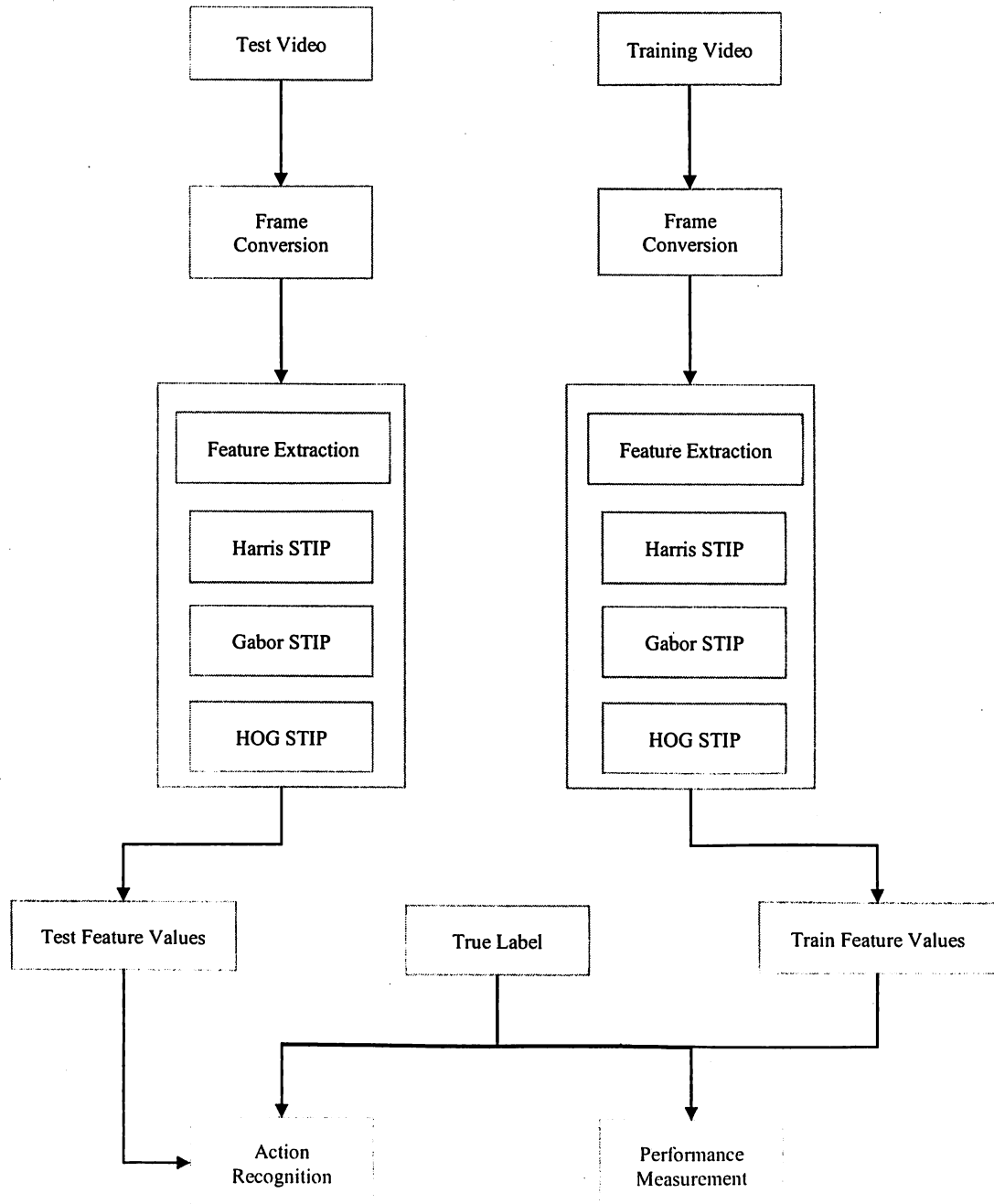


Figure 4.3: Flow Chart for the proposed work on Human Action Recognition

4.1.4 Activity Diagram

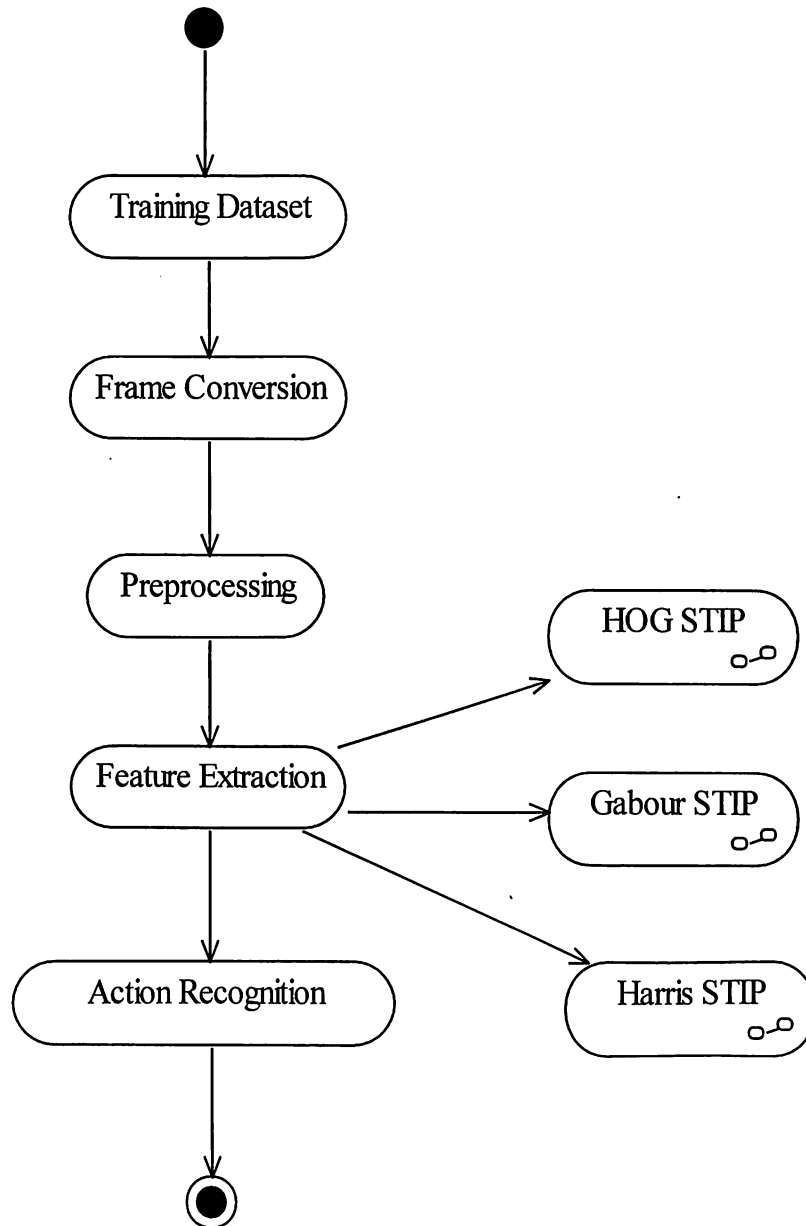


Figure 4.4: Activity diagram for Human Action Recognition

4.1.5 Sequence Diagram

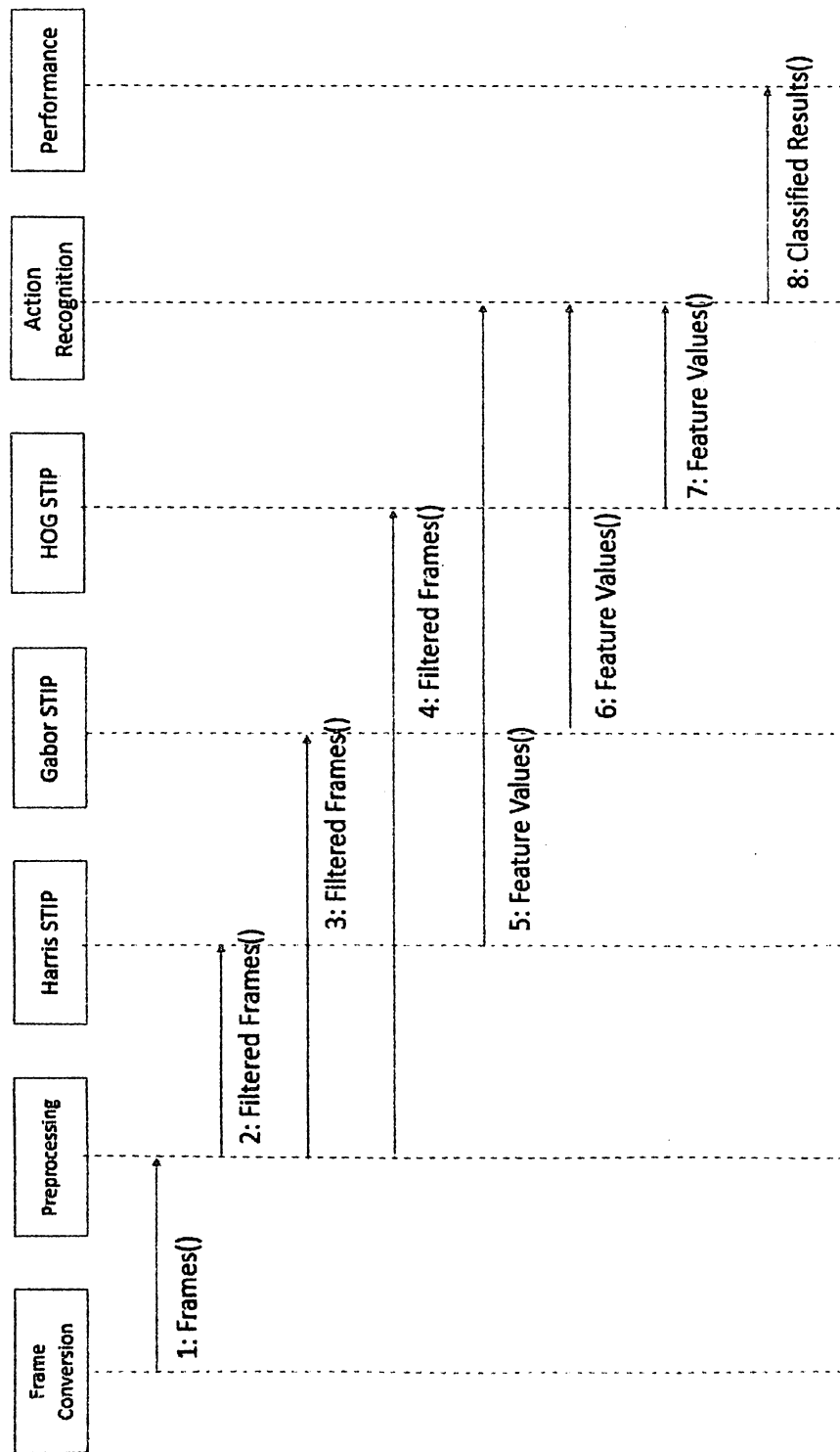


Figure 4.5: Sequence Diagram for Human Action Recognition

4.1.6 Collaboration Diagram

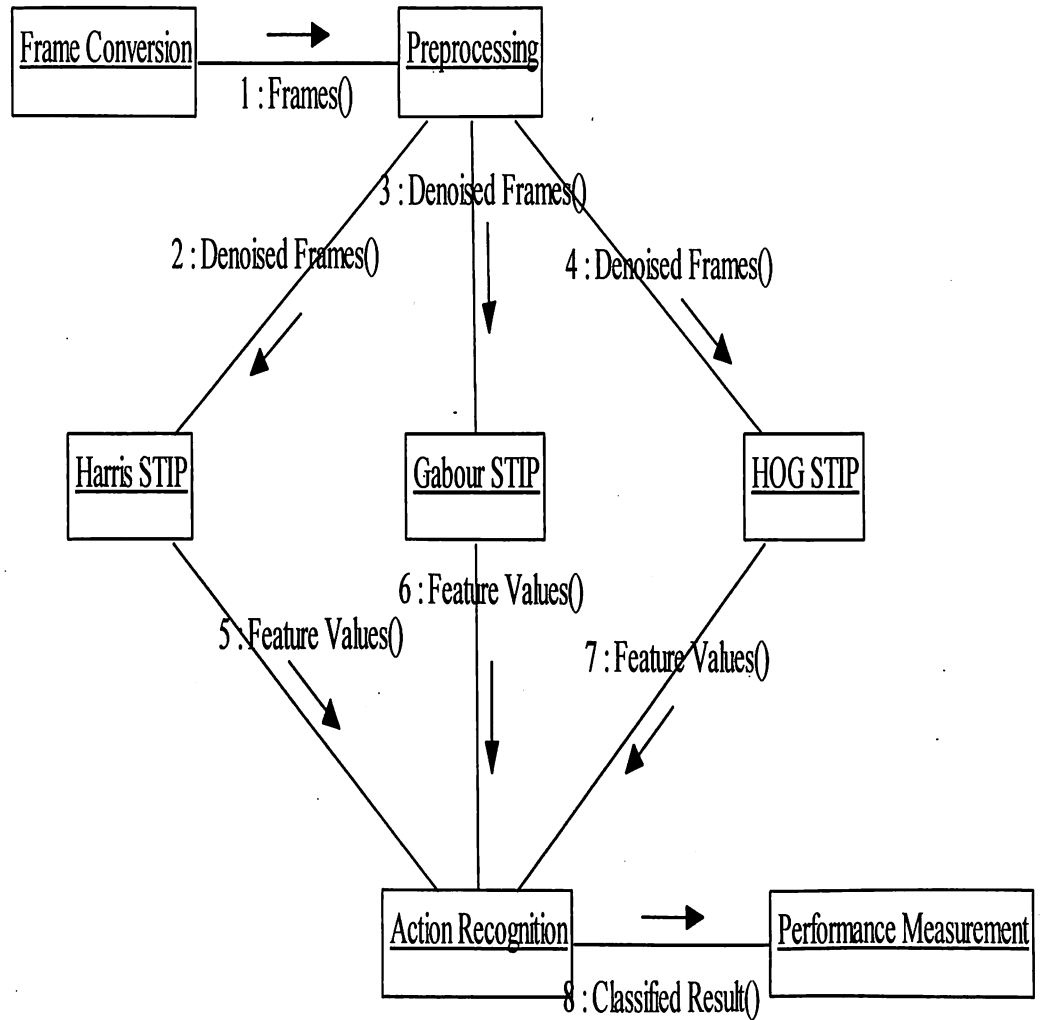


Figure 4.6: Collaboration Diagram for Human Action Recognition

5. COMPUTATIONAL THEORY

Interpretation of visual information involves comparison of images taken under different conditions and at different moments in time. As images representing the same scene or a class of objects might be very different depending on the view, the lighting, etc., there is a need for invariant representations that emphasize the important properties of the image while suppressing irrelevant variations. This section provide a detailed description about keypoint detectors and descriptors implemented in the dissertation.

5.1 Low level representation

To represent the human pose, it is important that the representation should describe some relevant primitive properties of the image. It must be based upon a few estimations or tests upon the picture. The estimation can be seen as a coordination between some suitable features and nearby neighborhoods, or window in the original picture. Contingent upon the level of coordinating, we will appoint some measure to the descriptive change. If the matching is performed with a single template under simple conditions, the measure derived will be a scalar. In general we may want to match with several templates to obtain a better definition of what happens within the window, which leads to a measure in the form of a vector. The matching is usually performed as a convolution between a template and the image. For a particular window, at a particular position of the image having pixel values x_h , inside it, we can write the convolution as a product sum:

$$q = \sum_k f_k h_k$$

Where, h_k are weight coefficients defining the template or kernel.

5.1.1 Representation of phase

Line and edge components for a particular orientation, k , can be combined into a magnitude and phase representation.

$$\begin{cases} q_k = \sqrt{q_{kl}^2 + q_{ke}^2} \\ \theta_k = \arctan(q_{ke}/q_{kl}) \end{cases}$$

We can consider q_k as an appropriate measure for the combination of line and edge, as the corresponding kernels are orthogonal and the individual components can consequently be added geometrically.

The *phase makes it possible to distinguish*, for example, *between a bright line against a dark background and its complement*, or to identify a particular combination of line and edge. If we take into account only the magnitudes, the output from a neighborhood is reduced to 4 components, one for each orientation.

5.2 Preprocessing Steps

Different filtering approaches are available for image processing, e.g., in spatial domain low-pass filters which is often used for image smoothing or blurring, high-pass filters for sharpening the image, averaging filter, median filters, max filter, min filter, box filter, etc.; and in frequency domain. Butterworth low-pass filter, Gaussian low-pass filter, high-pass filter, Laplacian in the frequency domain, etc. In many cases, initially, images are smoothed by employing Gaussian or other low-pass filtering schemes.

5.2.1 Median Filtering

Median filter is a well-known and widely used filtering scheme. We can exploit the nonlinear median filter to filter out noise. Median filtering reduces noise without blurring edges and other sharp details. Rather than basically supplanting the pixel esteem with the average of neighboring pixel values, it substitutes the pixel values with the median of the neighboring pixels. The median is evaluated by first sorting the pixels with respect to their values from the encompassing neighborhood into mathematical order. At that point supplant the pixel being considered by the center or middle pixel of that neighboring window. An image is passed through the median filter to smooth unwanted noisy outlines and thus we can accomplish with a smoothed picture.

5.3 Feature Detectors

A feature or keypoint detector locates the points of interest in an image where elements will be separated. Such keypoints or the regions are known as Spatio-Temporal Interest Points (STIPs) [5]. A keypoint is a STIP in the space (x, y) and time t , that has high measures of changes in its surrounding region. In the spatial area this shows as substantial differentiation changes, yielding

a Spatial Interest Point. Saliency in the space occurs when a point changes after some time, and when this change happens at a spatial interest point, the fact of the matter is then a STIP.

5.3.1 What is a feature? What constitutes a feature?

The perspective of defining feature changes with context and thus cannot be clearly defined, therefore constituents of feature also fluctuate depending on the application and context. In image processing, features could be blobs, edges, interest points, regions of interest, corners etc. are typically considered as image features and therefore, in image processing context, we extract these features for further processing. Note that in varying illuminance conditions, features may not find proper correspondence to the edge locations and the corresponding features.

In what capacity would we be able to discover image areas which can be dependably discovered using different frames of a video, i.e., what could be a possible good keypoints to trail? [61][62] In fig. 5.1, there are three specimen areas to perceive the possible interest point which may be coordinated or followed. It is easy to intercept that, flat surfaces are almost not possible to localize. Areas having substantial differentiation changes are simpler to find, albeit an edge region at a solitary orientation experiences unkind effects of *aperture issue* [63][64][65], i.e., it is only possible to modify the region along the bearing typical to the edge course (fig 5.1-ii). Areas with edges in no under two (in a general sense) different acquaintances are the most easy with restrict, as showed in fig 5.1-i.

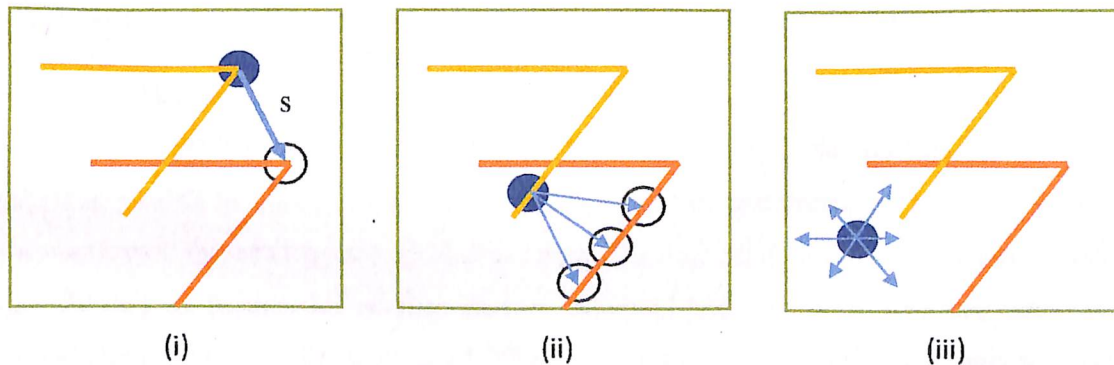


Figure 5.1: Aperture problems for varying feature regions: (a) Corners; (b) B.P. Illusion; (c) flat region.

Impulses can be formalized by looking in any event complex possible planning premise for differentiating two picture regions, i.e., their summed square difference,

$$E_S(s) = \sum_i w(x_i)[F_1(x_i + s) - F_0(x_i)]^2$$

In this, F_0 and F_1 are the two consecutive images which are being matched, $s = (s; v)$ is the *displacement vector*, $w(x)$ can be defined as a window function over the space (u, v) , and the \sum_i is measured over the keypoint region.

To compute the stability of feature points in an image with respect to small changes at some location Δu , can be performed by associating a frame with its own, this is also called as *auto-correlation function*.

$$E_A(\Delta u) = \sum_i w(x_i)[F_0(x_i + \Delta u) - F_0(x_i)]^2$$

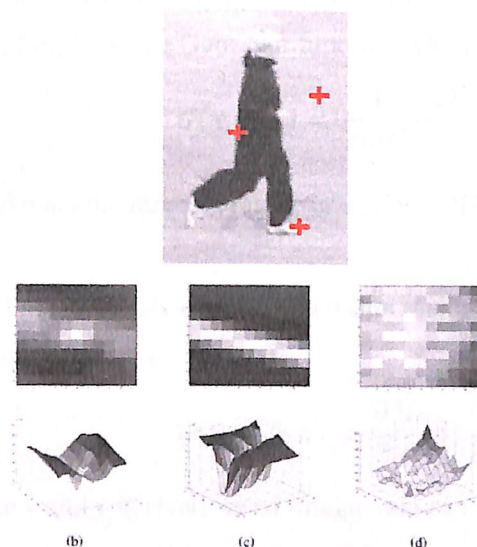


Figure 5.2 The auto-correlation displaying a corner, edge and a textureless surface.

5.3.2 Harris Detector

Corner detection [66] is a methodology utilized in wide area of application in the field of computer vision frameworks to extricate certain sorts of components and construe the points of a frame. These methodologies are broadly utilized as a part of movement identification, picture enrollment, video following, picture mosaicing, display sewing, and 3D demonstrating and question acknowledgment. In this thesis we used 3D Harris corner detection [5] which overlaps with the state-of art.

The calculation evaluates for every pixel in the frame to check whether a corner is available, by considering how comparable a patch focused on the pixel is to adjacent, to a great extent covering patches. The similitude is calculated by measuring the SSD between the two regions. Lower the

SSD will be more similarity could be find between two images. In the event that the pixel is in an area of uniform force, then the adjacent patches will appear to be same. On the off chance that the pixel is on an edge, then close-by regions in a heading opposite to the edge will look entirely changed, yet adjacent regions in a course parallel to the edge will come about just in a little modification. In the event that the pixel is on an element with variety in all headings, then none of the adjacent regions will appear to be same.

Step 1:

Compute the x and y derivatives of the two frames F_x and F_y by by using Laplacian of Gaussian (LoG). Let, G be the Gaussian function which can be written as:

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{y^2+x^2}{2\sigma^2}}$$

Where, (x, y) and σ are the location on the image on a 2D matrix and the standard deviation respectively.

The (undirected) second derivative of a two-dimensional image, known as the Laplacian operator, this can be given as:

$$\nabla^2 G(x, y; \sigma) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}$$

Now, to compute x and y derivatives of image, this can be given as:

$$F_x = G_\sigma^x * F$$

$$F_y = G_\sigma^y * F$$

Step 2:

Calculate the value of F_{x^2}, F_{y^2}, F_{xy} which can be given as:

$$F_{x^2} = F_x \cdot F_x$$

$$F_{y^2} = F_y \cdot F_y$$

$$F_{xy} = F_x \cdot F_y$$

Step 3:

Convolve each of these images with a larger Gaussian. This can be done by computing sum of product of derivatives at each pixel.

$$S_{x^2} = G_{\sigma_1} * F_{x^2}$$

$$S_{y^2} = G_{\sigma_1} * F_{y^2}$$

$$S_{xy} = G_{\sigma_1} \cdot F_{xy}$$

Step 4:

Calculate the scalar interest, A . This is measured by calculating the Eigen values matrix A , which is discussed earlier which is given as:

$$A = w * \begin{bmatrix} F_x^2 & F_x F_y \\ F_x F_y & F_y^2 \end{bmatrix}$$

Where, w is the Gaussian Kernel. Using matrix A , Eigen Values of (λ_0, λ_1) and Eigen direction as shown in fig 5.3. As greater uncertainty depends on smaller eigenvalue to locate good feature keypoints [62].

Step 5:

Calculate final response of the detector in each pixel value, given as:

$$R = \lambda_0 \cdot \lambda_1 - k(\lambda_0 + \lambda_1)^2$$

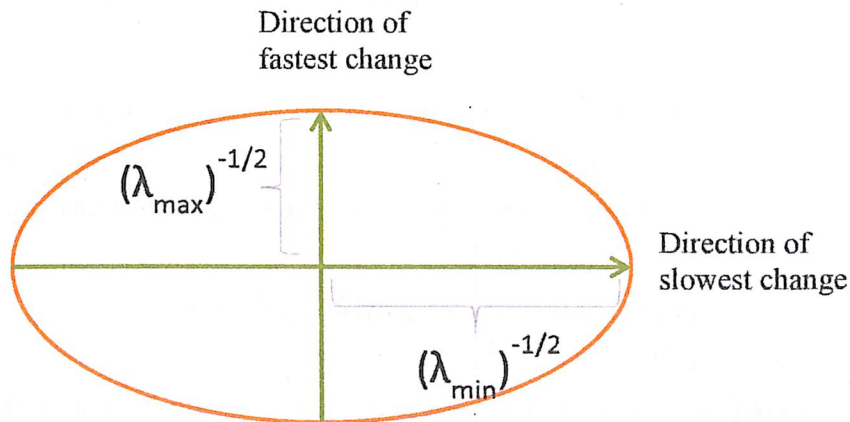


Figure 5.3 Eigenvalue scrutiny of the autocorrelation of the given ellipse.

Step 6:

Fix the threshold value of the local maxima and return these as detected keypoint regions. The eigenvalues of λ conclude if there is a corner, an edge or a *flat* features over the image. This is given as:

$$\begin{cases} \lambda_1, \lambda_2 > 0; & \text{Corner} \\ \lambda_1 \gg \lambda_2 \text{ or } \lambda_1 \ll \lambda_2; & \text{Edge} \\ \lambda_1, \lambda_2 < 0; & \text{Flat} \end{cases}$$

5.3.3 Harris 3D Detector

The Harris three dimensional detector was first used by Laptev and Lindeberg in [5], as an improved version of the original Harris corner detector, proposed by C. Harris [66] in space-time. In this, the video or frame is symbolized in a 3D function of $f(x, y, t)$, where t is the temporal dimension (time) in 2D (x, y) space. All steps of the algorithm are similar to the original Harris point except for now t -dimension is added to the detector. The scale-space representation $G(p; \Sigma)$ obtained by the convolution of the spatio-temporal signal $f(p)$, $p = (x, y, t)^T$ with separable 3D Gaussian kernel G , is given as:

$$G(p; \Sigma) = \frac{1}{(2\pi)^3 \sqrt{\det(\Sigma)}} e^{-(p^T \Sigma^{-1} p)/2}$$

Where,

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \tau^2 \end{pmatrix}$$

$$p = (x, y, t)^T$$

Σ is the Covariance, p denotes image coordinates. σ^2 and τ^2 denote spatial and temporal scale parameters respectively.

The second moment matrix is now spatio-temporal which is given as:

$$A(p; \Sigma) = G(p; \Sigma) * \begin{bmatrix} I_x^2 & I_x I_y & I_x I_t \\ I_x I_y & I_y^2 & I_y I_t \\ I_x I_t & I_y I_t & I_t^2 \end{bmatrix}$$

The Gaussian window function is now spatio-temporal having temporal variance τ_t , the corner function can be written as:

$$R = \text{Det}(A) - k(\text{Trace}(A))^3 = \lambda_1 \lambda_2 \lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3)^3$$

5.3.4 Gabor Detector

The inspiration to utilize Gabor functions is mostly related to life which is more or less related to visual cortex of humans [68]. The function goes about as low-level organized edges & surface differentiators which is variation to wide range of frequencies and scale. These substances elevated noteworthy hobby and motivated experts to extensively research the properties of Gabor functions. For a 2D Gaussian curve with a SD of σ_x, σ_y in x and y direction, response of the filter is given by:

$$h(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\right\} \cos(2\pi u_0 x)$$

Each filter output is smoothed using a Gaussian smoothing function that matches the corresponding filter spatial Gaussian curve:

$$I(x, y) = h(x, y) * g(x, y)$$

where,

$$g(x, y) = \exp\left(-\frac{\sqrt{x^2+y^2}}{2\sigma^2}\right)$$

A complex Gabor channel is described as aftereffect of a Gaussian window times a multifaceted sinusoid, i.e.

$$G(t) = ke^{j\theta}w(at)s(t)$$

Such that,

$$w(t) = e^{-\pi t^2}$$

$$s(t) = e^{j(2\pi f_0 t)}$$

Here k, θ, f_0 are filter parameters. The real and imaginary values if the filter is given as:

$$G_{real}(t) = w(t) \sin(2\pi f_0 t + \theta)$$

$$G_{imaginary}(t) = w(t) \cos(2\pi f_0 t + \theta)$$

Dollar et al. [68] reports that the Harris3D interest point detector does not identify enough keypoint features to perform well. The Cuboid interest point detector is therefore tuned in a way that its outcomes would be more features than Harris3D for the same recordings. It is sensitive to periodic movements which happen frequently in real life recordings thus proving a good results.

The response function for cuboid detector is given as:

$$R = (f * G(x, y; \sigma) * h_{ev})^2 + (f * G(x, y; \sigma) * h_{odd})^2$$

where

$$h_{even}(t; \tau, \omega) = -\cos(2\pi t\omega)e^{-t^2/\tau^2}$$

$$h_{odd}(t; \tau, \omega) = -\sin(2\pi t\omega)e^{-t^2/\tau^2}$$

h_{even} & h_{odd} are the quadratic pair of 1D Gabor filters. The 2D Gaussian filter window given as $G(x, y; \sigma)$ is implemented along the spatial direction.

5.4 Feature Descriptors

In the wake of unique elements or interest focuses, we should match them, i.e., we should figure out which keypoints originate from relating areas in various sequence of frames. Feature

descriptors represents the selected pixels in a way that expands characterization execution and gives a sparse representation. The descriptors should be invariant to issues like scale-space, orientation and illuminance variations. This invariance empowers descriptors to be coordinated crosswise over videos which have contrasts in these parameters.

5.4.1 HOG Descriptor

Histograms of Oriented Gradients [69] is one of the most powerful 2D descriptor, which was originally framed for human detection. HOG, classifies local object keypoint features and structure rather well by distributing local intensity gradients or edge directions. A HOG descriptor is processed utilizing a piece comprising of a matrix of cells where every cell again comprises of a lattice of pixels. The quantity of pixels in a cell and number of cells in a piece can be differed according to the necessity and need. HOG performs best using 3×3 cells in each block with 6×6 pixels size for each cell. Algorithm for HOG descriptor is given below.

Step 1: Gradient Computation

The initial step is the calculation of orientation and magnitude. One of the most well-known technique to implement the one dimensional focused point discrete derivative veil in both the x and y pivot. In particular, this technique requires separating the grayscale picture with the accompanying channel picture with the accompanying channel portions:

$$D_x = [-1 \ 0 \ 1]; D_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

So, being given an image I , x and y derivatives can be calculated using a convolution operation:

$$I_x = I * D_x; I_y = I * D_y$$

The magnitude G , is given as:

$$|G| = \sqrt{I_x^2 + I_y^2}$$

The orientation θ , is given by:

$$\theta = \arctan \frac{I_y}{I_x}$$

Step 2: Orientation Binning

The next phase includes find the cell histograms. Every pixel inside of the cell makes a weighted choice for the histogram channel based with respect to the qualities found in the angle calculation. Concerning the weight, pixel commitment can be the angle greatness itself, or the square foundation of the inclination extent.

Step 3: Descriptor Blocks

In order to account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cell together into larger, spatially-connected blocks. The HOG descriptor is then the vector of the components of the normalized cell histogram from all the block regions. These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor.

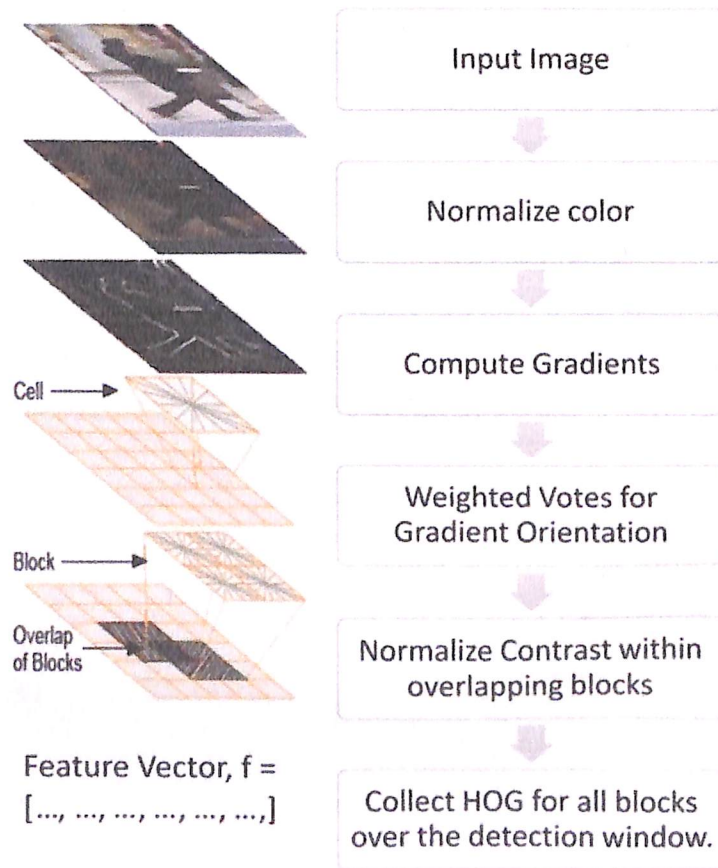


Figure 5.4: HOG Feature Extraction

Step 4: Block Normalization

There are diverse techniques for square standardization. Let, v be a chance to be the non-standardized vector containing every one of the histograms in a given piece, $\|v_k\|$ be its k -standard for $k = 1, 2$ and e be some little steady. At that point standardized vector can be given as:

$$L2 - norm: f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}}$$

5.4.2 HOG 3D Descriptor

The previous methodology of utilizing Histogram of Oriented Gradient in 2D as a spatio-temporal features involved registering gradient in cuboids instead of cells, yet the gradient directions are still 2D. Therefore HOG is incapable to capture the temporal information in the video, which is the reason the expansion of HOF essentially enhances the outcomes.

HOG3D [70] changes the fundamental methodology by considering the 3D gradient rather than the 2D gradient. The gradients are evaluated in 3D, and histograms are quantized into polyhedrons. This rich method for quantization is intuitive on the off chance that we take a gander at the standard method for quantizing 2D slope introductions by guess of a circle with a polygon. Every side of the polygon relates then to a histogram receptacle. By extending this to 3D the polygon turns into a polyhedron. The polyhedron utilized is an icosahedron which has 20 sides, hence bringing about a 20 receptacle histogram of 3D gradient directions. The last descriptor is gotten by connection of histograms into a vector, and standardization by the L2 standard.

6. REPRESENTATION

This segment depicts the representation of video in the proposed framework. The keypoint components extracted from a sequence of image frames which must give, such that it works well for recognizing human action. This generally results in a dimensionality diminishment, the image sequences as often as possible winds up being spoken to as a single vector, which is a profitable data to a classifier.

6.1 Bag of Features

The previous decade has seen the developing ubiquity of Bag of Features (BoF) [71] [72] ways to deal with numerous PC vision assignments, including picture arrangement, video search, robot localization, and composition acknowledgment. Part of the advance is simplicity. BoF strategies depend on orderless accumulations of quantized nearby picture descriptors; they dispose of spatial data and are thusly reasonably and computationally less difficult than numerous option techniques. Regardless of this, or maybe as a result of this, BoF-based frameworks have set new execution gauges on well-known picture characterization benchmarks and have accomplished versatility leaps forward in picture recovery.

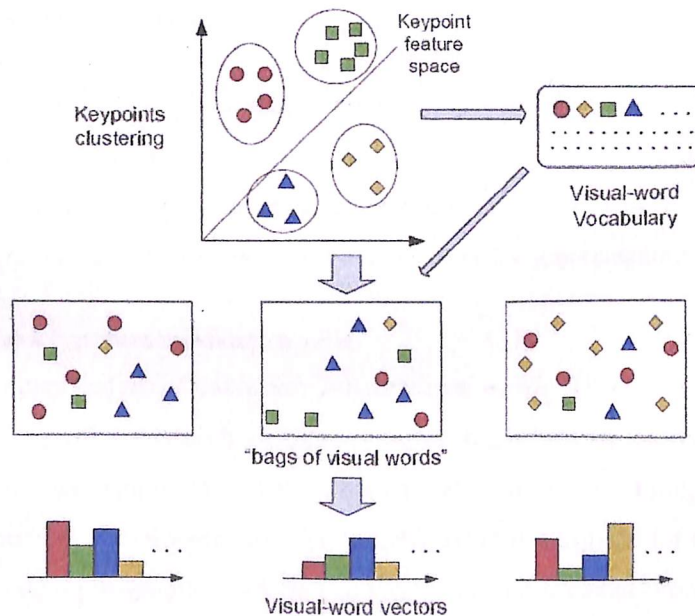


Figure 6.1: Bag of Features Model [73]

Bag-of Features (BoF) is derived from the Bag of words (BoW) [74] representation that is utilized as a part of common dialect processing to represent the text. The musing is that the substance can be spoken to by the occasions of words, ignoring the solicitation in which they appear in the substance, or to use the purposeful anecdote in the name, put each one of the words in a sack and urge them out without knowing the solicitation in which they were put in. The outcome of the count is a histogram of word occasions, which can be more important for taking a gander at compositions than a prompt word-by-word relationship as appeared in fig 6.1. The compartments in the histogram are known as the vocabulary, and can be the completed course of action of an extensive variety of words used as a part of the substance, or only a subset as it might be pertinent to filter through normal words like: the, be, to and of.

6.1.1 Learning the Visual Vocabulary

Utilization of the vector quantization (VQ) method groups utilizes the K -means to cluster the identical keypoint features and encrypts each interest point by the record of the cluster to which it has a place. Every cluster as a visual word indicate to a particular interest point feature class in the cluster. In this manner, the grouping process creates a visual word vocabulary depicting distinctive local patterns in pictures.

K-means clustering:

- Randomly initialize K number of cluster centers.
- Iterate until convergence:
 - Allocate each keypoint to the nearest cluster.
 - Compute every bunch focus as the mean of all focuses relegated to it.

6.1.2 Mapping the keypoints to visual words

We can represent every feature of image as a *Bag of visual words*. This representation is practically equivalent to the Bag-of-words archive representation regarding structure and semantics. Both representations are inadequate and high-dimensional, and pretty much as words pass on implications of a record, visual words uncover nearby examples normal for the entire picture. The pack of-visual-words representation can be changed over into a visual-word vector like the term vector of an archive. The visual-word vector might contain the vicinity or nonattendance data of

each visual word in the picture, the tally of each visual word (i.e., the quantity of keypoints in the relating bunch), or the check weighted by different elements.

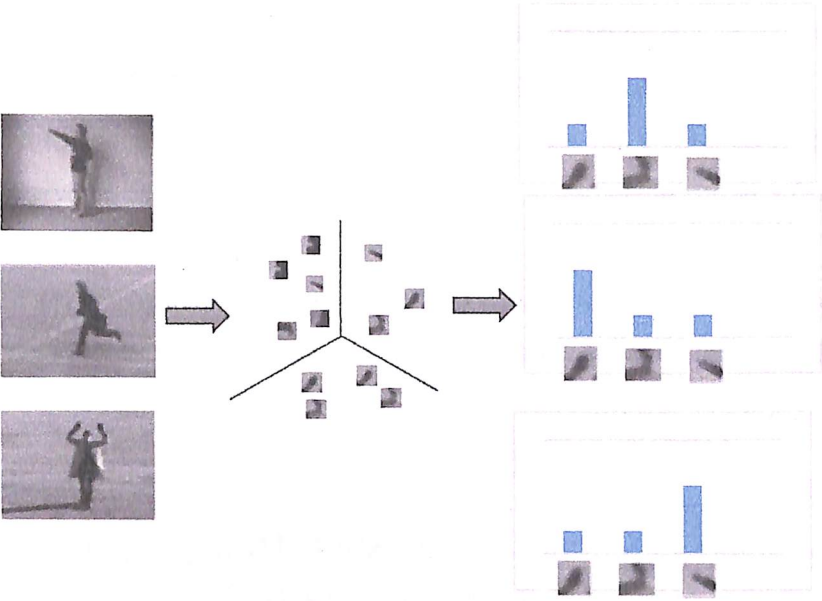


Figure 6.2 Representation of Bag-of-Features on KTH Dataset

7. SUPERVISED LEARNING

From the point of view of the vision group, classifiers are not an end in themselves, but rather a collective methods, so when a system that is simple, dependable and successful gets to be accessible, it has a tendency to be received broadly. The support vector machine [75] is such a method. This ought to be the main classifier you consider when you wish to assemble a classifier from samples. We give an essential prologue to the thoughts, and demonstrate a few cases where the system has demonstrated helpful.

Assume we have a set of N -point x_i that belong to two classes, which we shall indicate by 1 and -1 . These points come with their class labels, which we shall write as y_i , thus, our data set can be written as:

$$\{(x_1, y_1), \dots, (x_N, y_N)\}$$

7.1 SVMs for Linearly Separable Datasets

A Support Vector Machine (SVM) is a linear binary classifier that tries to augment the separation between the points of two classes. The arrangement comprises of a hyperplane that isolates the two classes in the most ideal way. It is conceivable to extend the SVM to accomplish non-straight multi-class order.

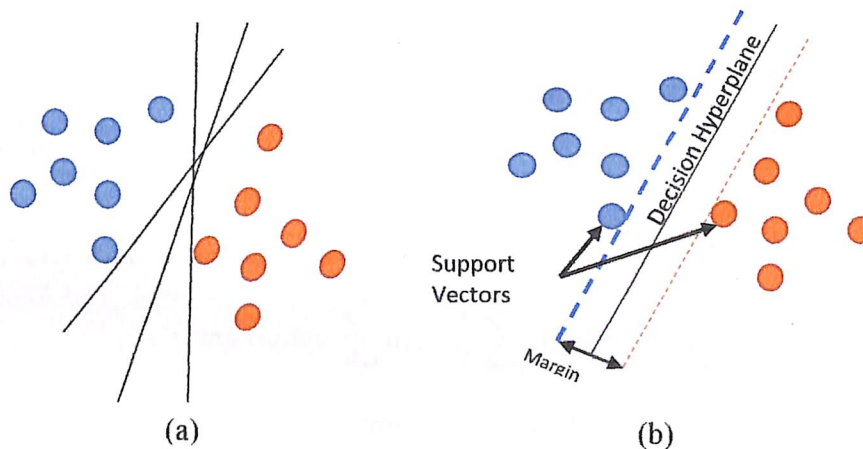


Figure 7.1 a) Number of possible linear classification between two classes; b) Generalised Classification using maximum margin

In a linearly separable data set, there is some choice of w and b (which represent a hyperplane) such that $y_i(w \cdot x_i + b) > 0$ for every example point. There is one of these expressions for each data point, and the set of expressions represents a set of constraints on the choice of w and b . These

constraints express the fact that all examples with a negative y_i should be on one side of the hyperplane and all with a positive y_i should be on the other side.

The most optimal choice of hyperplane is the one that is furthest from both clusters. This is obtained by joining the closest points on the two clusters, and constructing a hyperplane perpendicular to this line, and through its midpoint as shown in fig 7.1b. This hyperplane is as far as possible from each set, in the sense that it maximizes the minimum distance from example points to the hyperplane.

If we have N training points, where each input x_i is a D -dimensional feature vector, and is one of two class $y_i \in \{-1, +1\}$. The hyperplane is defined as:

$$y_i(w * x_i + b) = 0$$

where w is the normal to the hyperplane. The objective of the SVM can then be described as finding w and b such that the two classes are separated.

$$\begin{aligned} y_i(w * x_i + b) &> 0 && \text{for } y_i = +1 \\ y_i(w * x_i + b) &< 0 && \text{for } y_i = -1 \end{aligned}$$

These equations can be combined into

$$y_i(w * x_i + b) - 1 > 0; \quad \text{for } \forall_i$$

SVM for the given problem can be given as:

Notation:

We have a training set of N examples

$$\{(x_1, y_1), \dots, (x_N, y_N)\}$$

where y_i is either 1 or -1.

Solving for the SVM:

Set up and solve the dual optimization problem:

$$\begin{aligned} \text{maximize } & \sum_i^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i (y_i y_j x_i \cdot x_j) \alpha_j \\ \text{Subject to: } & \alpha_i \geq 0 \\ \text{and } & \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned}$$

We can then determine w from

$$w = \sum_{i=1}^N \alpha_i y_i x_i$$

Now for any example point x_i where α_i is non-zero, we have that

$$y_i(w * x_i + b) = 1$$

Classifying a point:

Any new data point can be classified by

$$f(x) = \text{sign}(w * x_i + b)$$

7.2 SVMs for non-Linearly Separable Datasets

In the definition utilized above, it is expected that the two classes are totally distinguishable by a hyperplane, yet this is frequently not the situation. The focuses are frequently caught in a way that makes it difficult to discrete them, yet it is still craved to have a classifier that commits as couple of errors as could be expected under the circumstances.

This is achieved by introducing a positive slack variable $\xi_i, i = 1 \dots L$. This slack allows the points to be located on the “wrong” side of the hyperplane as seen in the modified expressions.

$$\begin{aligned} (w * x_i + b) &> +1 - \xi_i && \text{for } y_1 = +1 \\ (w * x_i + b) &< -1 + \xi_i && \text{for } y_1 = -1 \\ \xi_i &\geq \forall_i \end{aligned}$$

Again, we have

$$w = \sum_{i=1}^N \alpha_i y_i x_i$$

but recovering b from the solution to the dual problem is slightly more interesting. This final expression can be thus given as:

$$\sum_{j=1}^N \alpha_j y_i x_i \cdot x_j + b = y_i$$

8. CONCLUSION

Current activities are for the most part performed in controlled setting for instance without movements in background. Assessment on such information does not help much to find genuine impediments of every technique. Nonetheless it is my opinion that assessment of strategies ought to be moved to practical scenes gradually. Implementing such systems using original sports recordings, motion pictures, and video information from the web, will help us to find the genuine necessities for activity acknowledgment, and it will offer us to move center to other vital issues some assistance with involving in real life action recognition, for example, segmentation of real time activities, managing obscure movements, composite activities, and view invariance, for instance.

In this dissertation, we implemented spatio-temporal interest point detectors and descriptors to fuse numerous photometric channels in addition to image intensities, bringing about color interest points. The upgraded framework of local features results in better performance of the system while recognizing the actions. Color feature points are evaluated and appeared to provide a good recognition rate in KTH and Weizmann datasets.

REFERENCES

- [1] A. Salah et al. (Eds.): Human Behavior Understanding, 2010, LNCS 6219, pp, 1-12, 2010. © Springer-Verlag Berlin, Heidelberg, 2010.
- [2] Lan T, Wang Y, Yang W, Mori G. Beyond actions: Discriminative models for on textual group activities. Neural Information Processing Systems (NIPS), (2010).
- [3] Gaidon A, Harchaoui Z, Schmid C. Actom sequence models for efficient action detection. IEEE Computer Vision and Pattern Recognition. 2011.
- [4] Karl-Friedrich Kraiss, Advanced Man-Machine Interaction Fundamental & Implementation, © Springer-Verlag Berlin Heidelberg, 2006.
- [5] I. Laptev. On space-time interest points. International Journal of Computer Vision 64 (2-3), 107-123.
- [6] P Dollár, V Rabaud, G Cottrell, S Belongie. Behavior recognition via sparse spatio-temporal features. 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005.
- [7] Ivo Everts, Jan C. van Gemert, and Theo Gevers. Evaluation of Color Spatio-Temporal Interest Points for Human Action Recognition. IEEE transactions on image processing, vol. 23, no. 4, April 2014.
- [8] Calter, P.: Geometry in art and architecture. www.math.dartmouth.edu/~matc/math5.geometry/ (visited in Oct. 2006).
- [9] Reinhard Klette and Garry Tee. Understanding Human Motion: A Historic Review. Human Motion: Understanding, Modelling, Capture and Animation. 2008.
- [10] Borelli, G. A.: De Motu Animalium. On the Movement of Animals, translated from Latin to English by P. Maquet, Springer, Berlin, 1989.
- [11] Marey, E.-J. Animal Machine, Locomotion Earth and Air. Germer Bailli'ere, Paris, 1873.
- [12] Marey, E.-J. Le Mouvement. G. Masson, Paris, 1894.
- [13] Johansson, Gunnar. Visual motion perception. Scientific American, Vol 232(6), Jun 1975, 76-88. <http://dx.doi.org/10.1038/scientificamerican0675-76>.
- [14] Daniel Weinlanda, Remi Ronfardb, Edmond Boyerc, A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition, volume 115, Issue 2, February 2011, Pages 224-241.
- [15] Aggarwal, J., Ryoo, M., Human activity analysis: A survey, ACM Computing Surveys 43, 1-43, 2011.

- [16]Guangchun Cheng, Yiwen Wan, Abdullah N. Saudagar, Kamesh, *Advances in Human Action Recognition: A Survey*, arXiv: 1501.05964v1 [cs.CV] 23 Jan 2015.
- [17]Volker Kruger, Danica Kragic, Ales Ude, Christopher Geib, *The Meaning of Action: A review on action recognition and mapping*, Taylor and Francis Online, volume 21, Issue 13, 2007.
- [18]Gunnar Johansson, "Visual perception of biological motion and a model for its analysis", *Perception & Psychophysics*, Springer, vol. 14, no. 2, pages 201–211, 1973.
- [19]A Yilma and Mubarak Shah, "Recognizing human actions in videos acquired by uncalibrated moving cameras", *Tenth International Conference on Computer Vision*, volume 1, pages 150–157, IEEE, 2005.
- [20]J. Yamato, J. Ohya, K. Ishii, "Recognizing human action in time sequential images using hidden Markov model", *Computer Society Conference on Computer Vision and Pattern Recognition*, pages 379–385, IEEE, 1992.
- [21]Di Wu, Ling Shao, "Silhouette Analysis-Based Action Recognition via Exploiting Human Poses", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume 23, Issue 2, 2013.
- [22]Mintao Zhao, William G. Hayward, Isabelle Bühlhoff, "Holistic processing, contact, and the other-race effect in face recognition" *Vision Research*, Volume 105, December 2014, Pages 61–69.
- [23]I. Laptev, On space-time interest points. *International Journal of Computer Vision* 64 (2-3), 107-123.
- [24]G. Willems, T. Tuytelaars, and L. Van Gool. An efficient dense and scale-invariant spatio-temporal interest point detector. In *ECCV*, pages 650–663, 2008.
- [25]K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.
- [26]G. J. Burghouts and J. M. Geusebroek, "Performance evaluation of local color invariants," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 48–62, Jan. 2009.
- [27]J. Stöttinger, A. Hanbury, N. Sebe, and T. Gevers, "Sparse color interest points for image retrieval and object categorization," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2681–2692, May 2012.
- [28]J. van de Weijer, T. Gevers, and J. M. Geusebroek, "Edge and corner detection by photometric quasi-invariants," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 27, no. 4, pp. 625–630, Apr. 2005.
- [29]J. van de Weijer, T. Gevers, and A. W. M. Smeulders, "Robust photometric invariant features from the color tensor," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 118–127, Jan. 2006.
- [30]P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proc. 2nd Joint IEEE Int. Workshop VSPETS*, Oct. 2005, pp. 65–72.

- [31]G. Willems, T. Tuytelaars, and L. Van Gool. An efficient dense and scale-invariant spatio-temporal interest point detector. In ECCV, pages 650–663, 2008.
- [32]I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In CVPR, pages 1–8, 2008.
- [33]A. Klaser and M. Marszalek. A spatio-temporal descriptor based on 3d-gradients. In BMVC, 2008.
- [34]Christian Schuldt Ivan Laptev Barbara Caputo, “Recognizing Human Actions: A Local SVM Approach”, Proceedings of the 17th International Conference on Pattern Recognition (ICPR’04).
- [35]M. Chen and A. Hauptmann, “Mosift: Recognizing human actions in surveillance videos,” Ph.D. dissertation, School Comput. Sci., Carnegie Mellon Univ. Pittsburgh, PA, USA, 2009.
- [36]Xinxiao Wu, Dong Xu, Lixin Duan, Jiebo Luo, “Action recognition using context and appearance distribution features”, IEEE Conference on Computer Vision and Pattern Recognition, 2011.
- [37]Fahimeh Rezazadegan, Sareh Shirazi, Niko Sünderhauf, Michael Milford, Ben Upcroft, “Enhancing Human Action Recognition with Region Proposals”, Journal of Advanced Research, March 2015.
- [38]Edilson de Aguiar, Christian Theobalt, Carsten Stoll, Hans-Peter Seidel. Marker-Less 3D Feature Tracking for Mesh-Based Human Motion Capture. Human Motion –Understanding, Modeling, Capture and Animation. Proceedings Second Workshop, Human Motion. Rio de Janeiro, Brazil, October 2007. ©Springer.
- [39]Gregory Rogez, Ignasi Rius, Jesus Martinez-del-Rincon, Carlos Orrite. Exploiting Spatio-temporal Constraints for Robust 2D Pose Tracking. Human Motion –Understanding, Modeling, Capture and Animation. Proceedings Second Workshop, Human Motion. Rio de Janeiro, Brazil, October 2007. ©Springer.
- [40]Aggarwal, J., Ryoo, M., Human activity analysis: A survey, ACM Computing Surveys 43, 1-43, 2011.
- [41]Guangchun Cheng, Yiwen Wan, Abdullah N. Saudagar, Kamesh, Advances in Human Action Recognition: A Survey, arXiv: 1501.05964v1 [cs.CV] 23 Jan 2015.
- [42]Gunnar Johansson, “Visual perception of biological motion and a model for its analysis”, Perception & Psychophysics, Springer, vol. 14, no. 2, pages 201–211, 1973.
- [43]A Yilma and Mubarak Shah, “Recognizing human actions in videos acquired by uncalibrated moving cameras”, Tenth International Conference on Computer Vision, volume 1, pages 150–157, IEEE, 2005.
- [44]Saad Ali, Arslan Basharat, Mubarak Shah, “Chaotic invariants for human action recognition”, 11th International Conference on Computer Vision, pages 1–8, IEEE, 2007.

- [45]G. J. Burghouts and J. M. Geusebroek, "Performance evaluation of local colour invariants," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 48–62, Jan. 2009.
- [46]K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.
- [47]J. Stöttinger, A. Hanbury, N. Sebe, and T. Gevers, "Sparse color interest points for image retrieval and object categorization," *IEEE Trans. Image Processing*, vol. 21, no. 5, pp. 2681–2692, May 2012.
- [48]J. van de Weijer, T. Gevers, and J. M. Geusebroek, "Edge and corner detection by photometric quasi-invariants," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 27, no. 4, pp. 625–630, Apr. 2005.
- [49]J. van de Weijer, T. Gevers, and A. W. M. Smeulders, "Robust photometric invariant features from the color tensor," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 118–127, Jan. 2006.
- [50]H Wang, MM Ullah, A Klaser, I Laptev, C Schmid. Evaluation of local spatio-temporal features for action recognition. *BMVC 2009-British Machine Vision Conference*, 124.1-124.11.
- [51]Imran N. Junejo, Emilie Dexter, Ivan Laptev, and Patrick Pe' rez. View-Independent Action Recognition from Temporal Self-Similarities. *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 1, January 2011.
- [52]Ivo Everts, Jan C. van Gemert, and Theo Gevers. Evaluation of Color Spatio-Temporal Interest Points for Human Action Recognition. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, VOL. 23, NO. 4, APRIL 2014.
- [53]I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *Computer Vision and Pattern Recognition*. 2008.
- [54]J. C. Nibbles, H. Wang, and L. F.-F. Li. Unsupervised learning of human action categories using spatial-temporal words. In *BMVC*, 2006.
- [55]C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: A local SVM approach. In *ICPR*, 32-36, 2004.
- [56]Cheung, V., Frey, B. J., & Jojic, N. (2005). Video epitomes. In *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (Vol. 1, pp. 42–49)*.
- [57]Plinio Moreno, Alexandre Bernardino, and Jos'e Santos-Victor. Gabor Parameter Selection for Local Feature Detection. *IBPRIA - 2nd Iberian Conference on Pattern Recognition and Image Analysis*, Estoril, Portugal, June 2005.
- [58]Xinghua Sun, Mingyu Chen, Alexander Hauptmann. Action Recognition via Local Descriptors and Holistic Features. In *CVPR*, 2009.

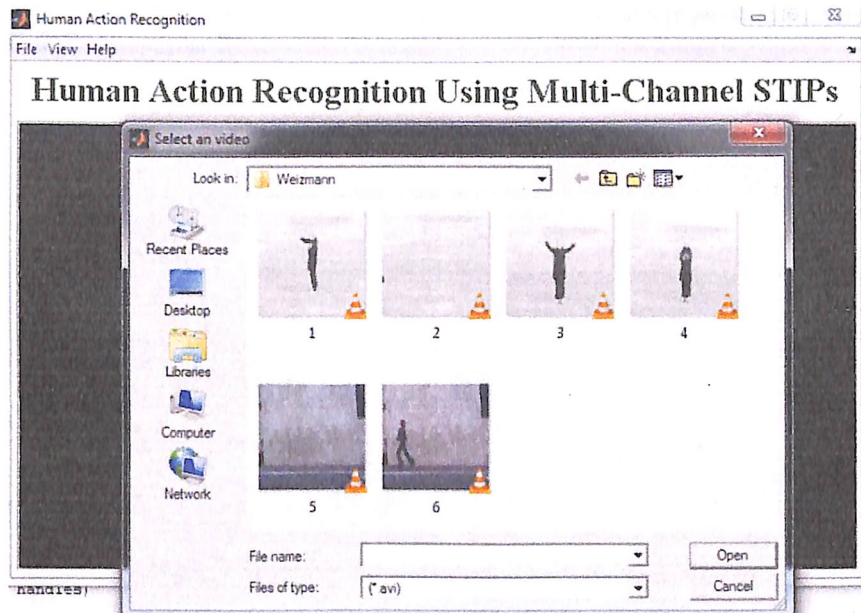
- [59]Koen E.A. van de Sande, Theo Gevers and Cees G.M. Snoek. Color Descriptors for Object Category Recognition. European Conference on Color in Graphics, Imaging and Vision, page 378-381, 2008.
- [60]Koen E. A. van de Sande and Theo Gevers and Cees G. M. Snoek. Evaluation of Color Descriptors for Object and Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 32 (9), page 1582-1596, 2010.
- [61]Malik, J., Belongie, S., Leung, T., and Shi, J. (2001). Contour and texture analysis for image segmentation. International Journal of Computer Vision, 43(1):7–27.
- [62]Shi, J. and Tomasi, C. (1994). Good features to track. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94), pp. 593–600, Seattle.
- [63]Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. Artificial Intelligence, 17:185–203.
- [64]Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application in stereo vision. In Seventh International Joint Conference on Artificial Intelligence (IJCAI-81), pp. 674–679, Vancouver.
- [65]Anandan, P. (1989). A computational framework and an algorithm for the measurement of visual motion. International Journal of Computer Vision, 2(3):283–310.
- [66]Harris, C. and Stephens, M. J. (1988). A combined corner and edge detector. In Alvey Vision Conference, pp. 147–152.
- [67]D. Gabor. Theory of communication. Part 1: The analysis of information. Journal of Institute of Electrical Engineers- Part III: Radio and Communication Engineering. Vol. 33. Issue 2. Jan 2010.
- [68]P Dollár, V Rabaud, G Cottrell, S Belongie. Behavior recognition via sparse spatio-temporal features. 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005.
- [69]N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.
- [70]Alexander Klaser, Marcin Marszałek, and Cordelia Schmid. A spatio-temporal descriptor based on 3d-gradients. In British Machine Vision Conference, pages 995–1004, 2008.
- [71]Eric Nowak, Frédéric Jurie, Bill Triggs. Sampling Strategies for Bag-of-Features Image Classification. Computer Vision ECCV. Volume 3954 of the series Lecture Notes in Computer Science pp 490-503. 2006.

- [72]S. Lazebnik, C. Schmid, J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume: 2. 2006.
- [73]J Yang et. al, Evaluating Bag-of-Visual-Words Representations in Scene Classification, MIR'07
- [74]Joachims, Thorsten. Learning to classify text using support vector machines: Methods, theory and algorithms. Kluwer Academic Publishers, 2002.
- [75] Cortes, Corinna, and Vladimir Vapnik. "Support vector machine." Machine learning 20.3 (1995): 273-297.

Appendix I: Project Guide

Step 1: Click on File Menu and select Open.

Step 2: Select the Dataset (Video File) and click on Open.



Step 3:

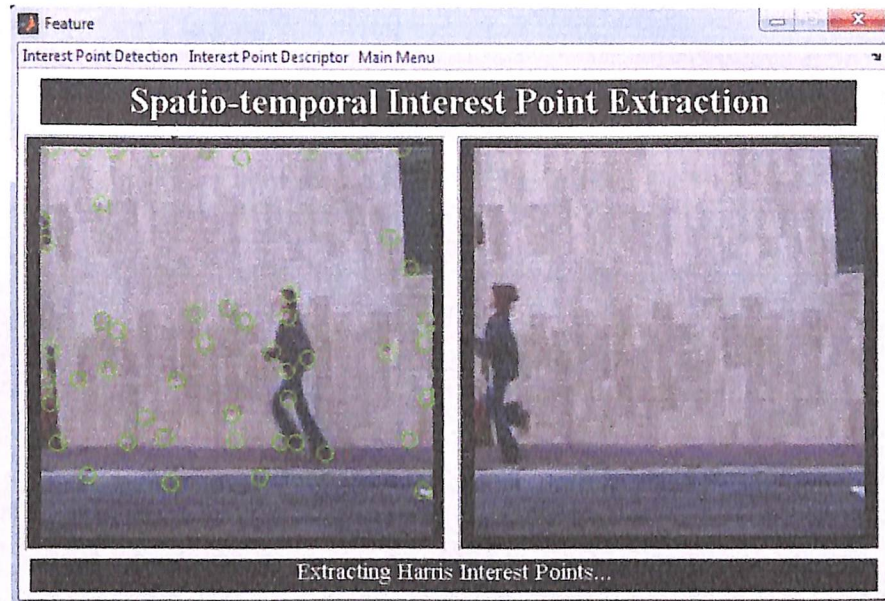
Click on View Menu and select Preprocessing.

The image on the left is from Original video. The image on the right is preprocessed video.

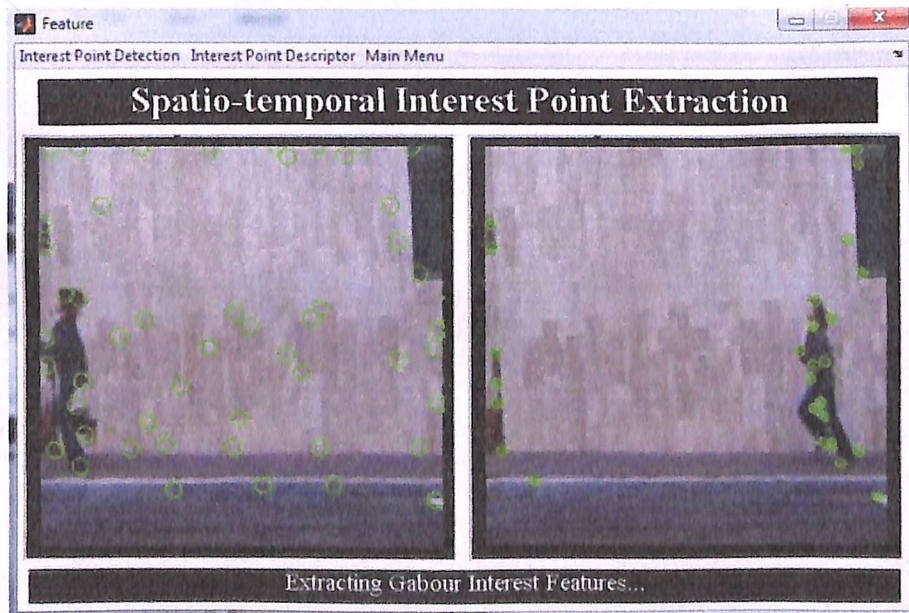


Step 4: Click on View Menu and select Feature Extraction

Step 5: Click on *Interest Point Detection* and select *Harris Point Feature Detection*. Harris Features are displayed on the left side.



Step 6: Click on *Interest Point Detection* and select *Gabor Point Feature Detection*. Gabor features are shown in the right.

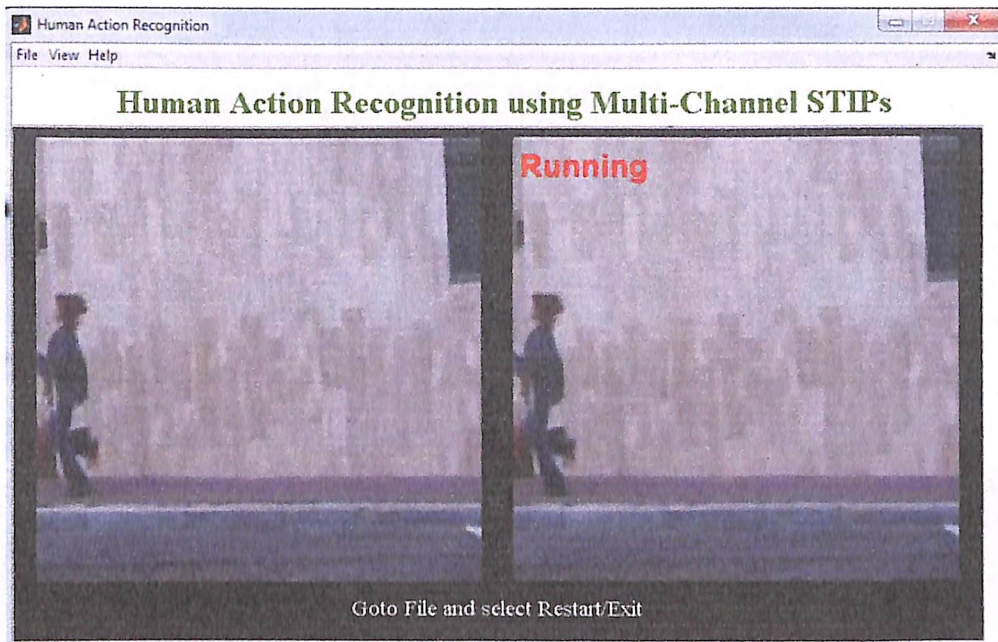


Step 7: Click on *Interest Point Descriptor Menu* and select *HOG Transform Descriptor*.

Step 8: Click on Main Menu.

Step 9: Click on View and select classification.

Step 10: To restart/Exit: Click on File Menu and select Restart/Exit.



Human Action Recognition using Multi-Channel Spatio-Temporal Interest Points

by Ayush Purohit

FILE

TIME SUBMITTED

SUBMISSION ID

10000000000000000000

10000000000000000000

10000000000000000000

WORD COUNT

CHARACTER COUNT

5710

136

Human Action Recognition using Multi-Channel Spatio-Temporal Interest Points

ORIGINALITY REPORT

23%

SIMILARITY INDEX

19%

INTERNET SOURCES

20%

PUBLICATIONS

15%

STUDENT PAPERS

PRIMARY SOURCES

- 1 Everts, Ivo, Jan C. van Gemert, and Theo Gevers. "Evaluation of Color Spatio-Temporal Interest Points for Human Action Recognition", IEEE Transactions on Image Processing, 2014. Publication 2%
- 2 www.ifp.illinois.edu Internet Source 1%
- 3 www.maia.ub.es Internet Source 1%
- 4 ijecs.in Internet Source 1%
- 5 Bebars, Amira Ali, and Elsayed E. Hemayed. "Comparative study for feature detectors in human activity recognition", 2013 9th International Computer Engineering Conference (ICENCO), 2013. Publication 1%
- 6 citr.auckland.ac.nz Internet Source 1%

7	clef2010.org Internet Source	<1%
8	Nie, Siqu, and Qiang Ji. "Capturing Global and Local Dynamics for Human Action Recognition", 2014 22nd International Conference on Pattern Recognition, 2014. Publication	<1%
9	www.cv-foundation.org Internet Source	<1%
10	Liang, Pengpeng, Erik Blasch, and Haibin Ling. "Encoding Color Information for Visual Tracking: Algorithms and Benchmark", IEEE Transactions on Image Processing, 2015. Publication	<1%
11	vision.stanford.edu Internet Source	<1%
12	Submitted to iGroup Student Paper	<1%
13	research.microsoft.com Internet Source	<1%
14	Submitted to Universiti Kebangsaan Malaysia Student Paper	<1%
15	users.utcluj.ro Internet Source	<1%

- | | | |
|----|--|-----|
| 16 | Gall, J., A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky. "Hough Forests for Object Detection, Tracking, and Action Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011.
Publication | <1% |
| 17 | www.jatit.org
Internet Source | <1% |
| 18 | Submitted to Shri Guru Gobind Singhji Institute of Engineering and Technology
Student Paper | <1% |
| 19 | www.eumetsat.int
Internet Source | <1% |
| 20 | www.vertexsoft.co.in
Internet Source | <1% |
| 21 | portal.acm.org
Internet Source | <1% |
| 22 | Submitted to Central Queensland University
Student Paper | <1% |
| 23 | Lecture Notes in Computer Science, 2013.
Publication | <1% |
| 24 | Submitted to Deakin University
Student Paper | <1% |
| 25 | Submitted to University of Pretoria
Student Paper | <1% |

-
- | | | |
|----|--|-----|
| 26 | domino.mpi-inf.mpg.de
Internet Source | <1% |
| 27 | graphics.cs.cmu.edu
Internet Source | <1% |
| 28 | www.allpsych.uni-giessen.de
Internet Source | <1% |
| 29 | Maqueda, Ana I., Carlos R. del-Blanco, Fernando Jaureguizar, and Narciso Garcia. "Human-action recognition module for the new generation of augmented reality applications", 2015 International Symposium on Consumer Electronics (ISCE), 2015.
Publication | <1% |
| 30 | ftp.informatik.rwth-aachen.de
Internet Source | <1% |
| 31 | Yan, Shengye, Xinxing Xu, and Qingshan Liu. "Learning the object location, scale and view for image categorization with adapted classifier", Information Sciences, 2014.
Publication | <1% |
| 32 | Ingo Mierswa. "Controlling overfitting with multi-objective support vector machines", Proceedings of the 9th annual conference on Genetic and evolutionary computation - GECCO 07 GECCO 07, 2007
Publication | <1% |
-

33	Lecture Notes in Computer Science, 2012. Publication	<1%
34	dspace.jaist.ac.jp Internet Source	<1%
35	ijera.com Internet Source	<1%
36	cms.brookes.ac.uk Internet Source	<1%
37	Heiko Fuser. "High-precision THz spectrum analyzer based on an unstabilized frequency comb", 2011 International Conference on Infrared Millimeter and Terahertz Waves, 10/2011 Publication	<1%
38	www.ijee.org Internet Source	<1%
39	robotics.csie.ncku.edu.tw Internet Source	<1%
40	welcome.isr.ist.utl.pt Internet Source	<1%
41	www.citeulike.org Internet Source	<1%
42	Lecture Notes in Computer Science, 2015. Publication	<1%

43	joserivera.org Internet Source	<1 %
44	www.journals.elsevier.com Internet Source	<1 %
45	Studies in Computational Intelligence, 2014. Publication	<1 %
46	Submitted to University of Sheffield Student Paper	<1 %
47	www.utdallas.edu Internet Source	<1 %
48	Submitted to Hanlym University Student Paper	<1 %
49	www.oqtans.org Internet Source	<1 %
50	svr-www.eng.cam.ac.uk Internet Source	<1 %
51	portal2.acm.org Internet Source	<1 %
52	eprints.eemcs.utwente.nl Internet Source	<1 %
53	Alain Simac-Lejeune. "Relevance of Interest Points for Eye Position Prediction on Videos", Lecture Notes in Computer Science, 2009 Publication	<1 %

54	Submitted to Higher Education Commission Pakistan Student Paper	<1%
55	Human Activity Recognition and Prediction, 2016. Publication	<1%
56	top25.sciencedirect.com Internet Source	<1%
57	Lecture Notes in Computer Science, 2010. Publication	<1%
58	www.int-arch-photogramm-remote-sens- spatial-inf-sci.net Internet Source	<1%
59	Wang, Bin, Yu Liu, Wenhua Xiao, Wei Xu, and Maojun Zhang. "Position and locality constrained soft coding for human action recognition", Journal of Electronic Imaging, 2013. Publication	<1%
60	www.cs.cmu.edu Internet Source	<1%
61	eprints-phd.biblio.unitn.it Internet Source	<1%
62	arxiv.org Internet Source	<1%

63	Submitted to Queen's University of Belfast Student Paper	<1%
64	www.lri.fr Internet Source	<1%
65	Noel E. O'Connor. "MyPlaces", Proceedings of the 2008 international conference on Content-based image and video retrieval - CIVR 08 CIVR 08, 2008 Publication	<1%
66	Submitted to So. Orange County Community College District Student Paper	<1%
67	Submitted to CTI Education Group Student Paper	<1%
68	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	<1%
69	Submitted to University of Glasgow Student Paper	<1%
70	Wang, Jiang, Zicheng Liu, and Ying Wu. "Introduction", SpringerBriefs in Computer Science, 2014. Publication	<1%
71	www.joics.com Internet Source	<1%

- | | | |
|----|---|-----|
| 72 | www-nlpir.nist.gov
Internet Source | <1% |
| 73 | Submitted to University of Michigan, Dearborn
Student Paper | <1% |
| 74 | Jie Xu. "Incremental EM for Probabilistic Latent Semantic Analysis on Human Action Recognition", 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 09/2009
Publication | <1% |
| 75 | Studies in Computational Intelligence, 2016.
Publication | <1% |
| 76 | Ta, Anh-Phuong, Christian Wolf, Guillaume Lavoue, Atilla Baskurt, and Jean-Michel Jolion. "Pairwise Features for Human Action Recognition", 2010 20th International Conference on Pattern Recognition, 2010.
Publication | <1% |
| 77 | Rydell, Nils W.. "Forces Acting on the Femoral Head-Prosthesis: A Study on Strain Gauge Supplied Prostheses in Living Persons", Acta Orthopaedica Scandinavica, 1966.
Publication | <1% |
| 78 | Agustí, Pau, V. Javier Traver, and Filiberto Pla. "Bag-of-words with aggregated temporal pairwise word co-occurrence for human action | <1% |

recognition", Pattern Recognition Letters, 2014.

Publication

79	ojs.academypublisher.com Internet Source	<1%
80	www.jdl.ac.cn Internet Source	<1%
81	www.znu.ac.ir Internet Source	<1%
82	www.solustan.com Internet Source	<1%
83	textarchive.ru Internet Source	<1%
84	alpha.science.unitn.it Internet Source	<1%
85	rizoiu.eu Internet Source	<1%
86	Bellamine, Insaf, and Hamid Tairi. "Motion detection using color structure-texture image decomposition", 2015 Intelligent Systems and Computer Vision (ISCV), 2015. Publication	<1%
87	www.dam.brown.edu Internet Source	<1%
88	crcv.ucf.edu Internet Source	<1%

89	new.isr.ist.utl.pt Internet Source	<1 %
90	Minhas, R.. "Human action recognition using extreme learning machine based on visual vocabularies", Neurocomputing, 201006 Publication	<1 %
91	www.yugangjiang.info Internet Source	<1 %
92	www.db-thueringen.de Internet Source	<1 %
93	www.first-mm.eu Internet Source	<1 %
94	Ji, Lei, Zheng Qin, Kai Chen, and Huan Li. "Visual Recognition Using Density Adaptive Clustering", 2011 Fifth FTRA International Conference on Multimedia and Ubiquitous Engineering, 2011. Publication	<1 %
95	Mubarak Shah. "Learning human actions via information maximization", 2008 IEEE Conference on Computer Vision and Pattern Recognition, 06/2008 Publication	<1 %
96	Plinio Moreno. "Gabor Parameter Selection for Local Feature Detection", Lecture Notes in	<1 %