Design of Secure Surveillance Framework

using 5D Chaotic Map Image Encryption Technique

for IoT Systems

A Thesis

Submitted in partial fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science and Engineering

By

Gopal Ghosh

(41800889)

Supervised By

Dr. Divya

Co-Supervised By

Dr. Kavita

LOVELY PROFESSIONAL UNIVERSITY

PUNJAB

2022

Declaration

I hereby declare that the thesis entitled “Design of Secure Surveillance Framework using 5D Chaotic Map Image Encryption Technique for IoT Systems” submitted by me for the Degree of Doctor of Philosophy in Computer Science and Engineering is the result of my original and independent research work carried out under the guidance of Supervisor Dr. Divya and Co-Supervisor Dr. Kavita, and it has not been submitted for to any university or institute for the award of any degree or diploma.

Place: Phagwara

Date: 25/02/2022

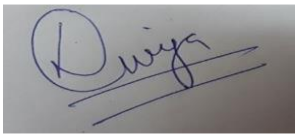
Gopal Ghosh
Signature

Certificate

This is to certify that the thesis entitled "Design of Secure Surveillance Framework using 5D chaotic map image encryption technique for IoT systems" submitted by Gopal Ghosh for the award of the degree of the Doctor of Philosophy in Computer Science and Engineering, Lovely Professional University Punjab, India is entirely based on the work carried out by him under our supervision and guidance, The work recorded, embodies the original work of the candidate and has not been submitted for the award of any degree, or diploma of any university and institute, according to the best of our knowledge.

Place: Phagwara

Date:25 /02/2022



Kavita

Signature of Guide with Date & UID **Signature of Co-Guide with Date & UID**

Acknowledgement

I wish to express my sincere thanks to my supervisor **Dr. Divya**, Assistant Professor, School of Computer Science and Engineering, Lovely Professional University, and my co- supervisor, **Dr. Kavita**, Associate Professor School of Computer Science and Engineering, Chandigarh University, with immense pleasure and a deep sense of gratitude. Both of my supervisors have given me the liberty to follow my passion for research work. The guidance, counseling, and suggestions were beneficial in completing work. It was a great privilege and honor to work and study under their guidance. Without their support, I would not have done this work with dedication. Both were my pillars of strength throughout this journey.

I would like to thank my mentor **Dr. Sahil Verma and Dr. Aman Singh** for all the continuous support and confidence to carry my research correctly. I express my sincere thanks to my colleagues who have encouraged me always to complete this workon time.

I express my gratitude towards my colleague **Mr. Parshotam** for working very hard with me in image processing. Special thanks go to the panel and reviewers of the journals who vetted my submissions and gave valuable comments to improve the work further.

Finally, I wish to express my thanks to my mother, father for their sacrifice for encouraging and inspiring me to do my best.

Gopal · Ghosh

Date:25 /02/2022 **Gopal Ghosh**

LIST OF FIGURES

Figure No	Figure Name	Page Number
1.1	Image encryption using chaotic map	9
1.2	Flow chart of genetic algorithm	13
1.3	Flow chart of differential evolution	14
2.1	Local and temporary representation	24
2.2	HAR approaches in terms of machine learning monitoring	25
2.3	Surveillance camera	32
2.4	Real time person tracking using surveillance system	33
2.5	Face detection	33
2.6	Standard test images	35
3.1	Flowchart of proposed image encryption approach	40
3.2	Proposed methodology flowchart	41
3.3	Flowchart of the proposed model HAR	47
3.4	Block diagram of advance intelligent video surveillance system (IVSS) framework.	48
3.5	Human skeleton and selected angle pairs.	52
3.6	Example of convolutional neural network (CNN) for image classification.	59
3.7	An example of an extended presentation of the recurrent neural network (RNN) with a three-length input system, three hidden units, and one output.	59
3.8	Solution architecture.	62
3.9	Subtracting human bodies by removing empty frame from a work frame.	62
4.1	Chaotic attractor in y_1-y_2 plane	73
4.2	Chaotic attractor in y_1-y_2-y_3 plane	74
4.3	Time series plot for state variable y_1 using system	74
4.4	Input images	75
4.5	Histogram analysis	78

Figure No	Figure Name	Page Number
4.6	Gaussian noise attack based encrypted images	81
4.7	Salt & pepper noise based encrypted images	82
4.8	Enhancement based encrypted images	82
4.9	Sensitivity analysis	84

LIST OF TABLES

Table No	Table Name	Page Number
1.1	Types of chaotic variants	9
2.1	Comparison of competitive image encryption techniques	36
4.1	Entropy analysis of the proposed secure surveillance system	76
4.2	PSNR analysis of the proposed secure surveillance system.	76
4.3	Mean absolute error analysis of the proposed secure surveillance system.	77
4.4	Horizontal correlation analysis	79
4.5	Vertical correlation analysis	79
4.6	Diagonal correlation analysis	79
4.7	NPCR analysis of the proposed secure surveillance system.	80
4.8	UACI analysis of the proposed secure surveillance system.	80
4.9	Error between actual and modified keys based encrypted images	84
4.10	Encryption time analysis	85
4.11	Decryption time analysis	85

Table of Contents

Title	Page No
DECLARATION	**ii**
CERTIFICATE	**iii**
ABSTRACT	**iv**
ACKNOWLEDGEMENT	**viii**
TABLE OF CONTENTS	**ix**
List of Figures	**xiii**
List of Tables	**xv**

CHAPTER 1 Introduction

1.1	Human Event Recognition	2
1.2	An Overview on Person Identification	4
1.3	Human Behavior Activity Analysis	6
1.4	Chaotic Maps	8
1.5	Genetic Algorithm	10
1.6	Differential Evolution Algorithm	11
1.7	Objective of the Thesis	14
1.8	Organization of the Thesis	15

CHAPTER 2 Review of Literature

2.1	Research Problems in Gait Analysis	17
2.2	Template Based Learning	18
2.3	Generative Learning	18
2.4	Discriminative Learning	20
2.5	Initiative Learning	21
2.6	Simulation	21
2.7	HAR Approaches in Terms of Feature Removal Process	23
2.8	HAR Approaches in Terms of Recognition Categories	24
2.9	HAR Approaches in Terms of Machine Learning Monitoring	25
	2.9.1 Supervised Methods	26

2.9.2 Unsupervised Learning 26
2.9.3 Semi-Supervised Learning 27
2.10 Image Input Versus Video Input 27
2.11 Single View Compared to Multiple View Acquisition 28
2.12 Security and Surveillance Applications 28
2.13 TI Method 28
2.14 AR System and Approaches 29
2.15 Multimodal Sensors for AR Approaches 29
2.16 Activity Recognition Methods and Techniques 29
2.17 Dynamic Time Warping 30
2.18 Finite-State Machine (FSM) 30
2.19 Hidden Markov Model 31
2.20 Artificial Neural Networks 31
2.21 Some Recent Work and Application 31
2.21.1 Surveillance 31
2.21.2 Crowd Count and Behavior Detection 32
2.21.3 Face Detection 33
2.21.4 Human Pose and Gait Estimation 34
2.21.5 Loitering 34
2.22 Previous Methodologies Used in HAR Model 34
2.23 Comparison of Competitive Image Encryption Techniques 35

CHAPTER 3 Methodologies
3.1 Proposed Image Encryption Technique 39
3.2 Human Activity Recognition (HAR) Model Using Deep Learning 46
3.3 Intelligent Video Surveillance System (IVSS) Framework 47
3.4 Types of Human Interaction Recognition 49
3.4.1 Simple Interaction Recognition 49
3.4.2 Concurrent Interaction Recognition 49
3.4.3 Recursive Activity Recognition 50
3.5 Stream Analysis of the Proposed Algorithm 50

3.5.1 Point Between Joints 50
3.6 Mechanism of Sliding Kernel 52
3.7 Discrete Haar Wavelet Transform (DWT) Algorithm 53
3.8 Dimension Reduction with Averaging 55
3.9 Thresholding with Inverse HWT Methodology 55
3.10 Feature Extraction 56
3.11 Types of Classifiers 57
3.12 Rule Tree Classifier 57
3.13 Bayesian Classifier 58
3.14 In-depth Learning Using Human Recognition 58
3.15 Sensors 59
3.16 Data 60
3.17 Inertial Sensors Data 60
3.18 Physiological Sensors Data 60
3.19 Environmental Sensors Data 61
3.20 System Description 61
3.21 Preprocessing 62
3.22 1-D Convolutional Neural Network 63
3.23 Classification of Moving Human using KNN (k nearest neighbors) 63
3.24 Feature Fusion Algorithm 63
3.25 Infrastructure 64
3.26 Non-Sub Sampled Contourlet Transform Domain 72
3.27 Non-Dominated Sorting Genetic Algorithm (NSGA-3) 72
3.28 Beta Chaotic Map 72

CHAPTER 4 Experimental results and considerations

4.1. Secret Keys 74
4.2. Fitness Evaluation 74
4.3. Comparative Analysis 75
4.3.1 Visual Analysis 75
4.3.2 Entropy 76

	4.3.3	Mean Absolute Error	77
4.4.	Security Analysis		77
	4.4.1	Histogram Analysis	77
	4.4.2	Correlation Analysis	78
	4.4.3	Differential Analysis	80
4.5	Robustness Against Various Attacks		81
	4.5.1	Noise Attack	81
	4.5.2.	Enhancement Attack	82
4.6.	Key Analysis		83
	4.6.1	Secret Key Space	83
	4.6.2	Secret Key Sensitivity Analysis	83
	4.6.3	Execution Time	84

CHAPTER 5 Conclusions and Future Scope

5.1	Conclusions	86
5.2.	Future Scope	87

References **88**

Dissemination of Work **100**

Abstract

The huge sending of data and information using Internet of Things (IoT) is really empowering keen city projects everywhere on the world. The major concern while transmitting sensitive data is security and privacy protection across the IoT network. One of the known mechanism Human Activity Recognition (HAR) is a PC vision tool for video surveillance applications. The biggest advantage in convolutional neural networks is used to gain straight forwardly from crude time arrangement information, killing the requirement for area skill to physically design input highlights. Despite extensive research in the field of object detection and tracking, there are very few existing visual tracking systems that can automatically and robustly track a target while correctly maintaining their identities. Many of these target tracking systems do not handle background clutter, illumination variations within the scene, background motion (flowing water, moving leaves), camera motion, and frequent occlusion between targets well.

This thesis presents a work that addresses all of these surveillance system problems while improving tracking efficiency and requiring less computational time to process. One of the fastest growing markets in the security industry is video surveillance systems. The quantity of surveillance cameras mounted in public and private locations has currently improved significantly. Video cameras are used in surveillance structures to reveal the sports of gadgets (people, cars, and so on.) in a scene. Surveillance video statistics evaluation may be used for a selection of purposes, which includes crook investigations, navy intelligence, purchaser traffic pattern discovery, and so forth. However, existing video surveillance structures are not yet able to provide a comprehensive safety tracking system. The model can get acquainted with an indoors portrayal of the time association statistics and, in an ideal world, carry out correspondingly to fashions healthy on a designed variation of the dataset. The identification of a person in the event of occlusion, as well as at different walking speeds, remains a challenge. This issue has been addressed by the development a mechanized framework to distinguish the individual, which utilizes easier and all the more remarkable component extraction procedures to uniquely identify the person using his traits and contributes to an overall great acknowledgment

rate even on account of impediment. Concurrent interactions are interactions that occur between multiple people at the same time. The current challenge to be addressed in this field is concurrent interaction modelling.

In surveillance camera because of the complexity of body part movement, recognizing concurrent interactions becomes difficult to find the suspicious activity. So, Intelligent Video Surveillance System (IVSS) has been arisen to identify different exercises continuously video. HAR model consist of joint angles which are acquired utilizing an RGB-Depth sensor. An averaging calculation is implemented to diminish the computational cost, which brings about better execution in keeping up high precision. K-Nearest Neighbor is the most straightforward AI calculation which is completely founded on Administered Learning method. Nevertheless, the K-nearest neighbor (k-NN) calculation is use to perceive every one of the human exercises as for the given information. Following that, by composing all elemental actions/interactions, the entire scenario can be identified. Recursive Activity is a new attempt to define behavior in this thesis. This is accomplished by using a semantic descriptor that combines local and global motion features. In this model, all of the interactions between people are captured in a video clip. The interactions define new recursive activities such as greeting, fighting, loitering, and purchasing. The semantic features of location, velocity, and time sequence of motion attributes are used to describe person-to-person interactions. The semantics of the spatial distance between the people are defined using relations such as overlap, above, below, next to (right), and next to (left).

In video reconnaissance frameworks, the yield of foundation deduction will be utilized as a contribution to the following more significant level cycles. The significant examination difficulties of the foundation deduction measure are commotion, dynamic foundation, shadows, and changes in light. To resolve these issues, we utilized foundation deduction dependent on Versatile Neuro Fluffy Surmising Framework, Modified k-means algorithm, and flexible modified Adaptive Gaussian mix (GMM) model with three Temporal Differencing (TTD) methods in our work.

Accurate moving target detection remains a challenge in research on the detection and tracking of moving people. When there is a small difference in brightness between the background and moving people, partial occlusions, and cross targets, segmentation becomes more difficult. This study proposes an adaptive particle filter-based detection and tracking algorithm for moving people detection by taking these issues into account. We proposed a work that remembers three significant stages for moving individual identification and following: pre-preparing, discovery of moving individuals, and following of moving individuals. A versatile middle channel removes noises and blurs from input video frames. A Modified Watershed segmentation process is used in the second phase to detect moving people in each pre-processed video frame. In the tracking phase, an adaptive particle filter based on a hybrid algorithm Artificial Bee Colony-Particle Swarm Optimization (ABC-PSO) is used to track the detected moving people. By overcoming existing limitations, the proposed techniques are capable of carrying out the detection and tracking process of moving people effectively.

Shadow reveals the geometrical relationship between the object and the source of illumination. It creates issues by reconstructing an object's shape, which causes object segmentation to be inaccurate. As a result, shadow expulsion is a significant and required system for diminishing the adverse consequences of shadow while tracking moving objects in video sequences. We discussed Fuzzy based shadow removal in surveillance video in this thesis. Our trial results show that the proposed scalable people tracking system is capable of removing shadows and accurately tracking people. In today's world, public safety is a critical issue. Automatic threat detection assists security personnel in efficiently maintaining public safety. One such automatic threat detection technology in public places video surveillance to find the lost or unclaimed objects.

In thesis, we describe a unique context for abandoned and eliminated object detection in complicated surveillance video sequences using heritage subtraction and foreground analysis. It is also supplemented by tracking in order to reduce false positives. Our framework incorporates a versatile calculation adaptive Abandoned Object Detection (AOD) algorithm related with GMM for detecting abandoned and removed objects in a real-world environment. This algorithm relies upon on the type of static locales and customer characterized boundaries, for example, object length and abandoned time. An adaptive algorithm is one which adjustments its behavior based on the records to be had on the time the algorithm is administered. Along with this algorithm, a process that handles occlusion to detect unattended or removed objects in complex crowded environments and a process that distinguishes static person from static object have been integrated. A video surveillance framework is separated into two sections: client and server. A surveillance camera on the client side first captures video. At the server, captured video data is used for analysis and video processing. Video data must be highly compressed in order to be sent over very low bit-rate networks. Channels for mobile devices to reduce the size of the video, video compression techniques can be used. Even though we have used chaotic map to exhibit the ability to produce dynamic systems to produce sequence of numbers that are random in nature. This approach is used to encrypt messages.

For decryption, the sequence of random numbers is highly dependent on the initial condition used for generating the sequence. We attended moving people detection and tracking in compressed domain in this thesis by using a new parallel image cryptosystem is presented using 5D hyper-chaotic system. The 5D hyper-chaotic system is combined with the Logistic map to generate pseudo-random sequences of better properties in plain image. In this way, the provided 5D hyper-chaotic based parallel image cryptosystem could achieve excellent plaintextsensitivity and relatively satisfactory speed performance. Numerous simulations are conducted, and security analyses based on the simulation results are exhibited in support of the security and availability of the proposed 5D hyper-chaotic based parallel image cryptosystem. Even though 5D hyper-chaotic map provides more complex chaotic behaviors. The 5D hyperchaotic map has five attributes, better Lyapunov exponent (LE) and a cubic nonlinear product. The large no of attributes will provide good resistance, improve the key space and provides a dynamic chaotic behavior.

Chapter 1

Introduction

The industrial revolution occurred before 200 years prior to expand usefulness, which prompted the speed increase of the economy. Mechanical force replaces human and creature muscle power. The Web unrest started in 1950, and it depends on systems administration, processing, and correspondence to build usefulness and lift the worldwide economy significantly more than the mechanical unrest. The Internet of Things (IoT) is the following insurgency that will drive another string to build usefulness. The IoT is an amalgamation of savvy gadgets, shrewd frameworks, and wise dynamic capacities. Everything in the actual world can be associated and speak with one another thanks to IoT. IoT devices can communicate with one another without the need for human intervention. The IoT employs unique addressing schemes to interact with other objects/things and collaborate with objects to develop new applications or services [1]. Health monitoring, smart environment, smart homes, smart cities, wearables, industrial internet, smart retail, smart farming, and other applications are examples of IoTapplications. As the use of IoT systems grows, so do the number of issues that arise. Among many other concerns, the security of IoT can't be ignored. IoT gadgets can be gotten to from anyplace through an untrusted organization like the web. Accordingly, IoT frameworks are helpless against a wide scope of vindictive assaults.

The IoT is a quickly developing advancement that has a significant impact on all aspects of life. It is regarded as the next step in Internet technology. The effect of the attack and threat surrounding the Internet and its connected systems is considered in the variation of the operating environment in IoT. The IoT is a diverse system that includes a variety of sensors, nodes, and devices, as well as a variety of technology at each layer. However, because IPv4 address space is limited, an IoT object uses IPv6 to fill the void in the Internet. IoT objects can be smart metering devices, sensory devices, or health care sensors. Because providing the best insight for controlling the elements in a network is a major concern for network users, IoT occurs when a threat to data occurs in the network. There is a summary of various security breaches, with an output of around 10 million records per day / 420,000 per hour [2]. As new connected devices enter the market and security breaches are monitored, researchers have considered the cause for exposing the vulnerabilities by making the world aware of the potential damage that occurs in connecting devices that lack proper security features.

1.1 Human Event Recognition – An Overview

With the dramatic increase in advent of smart technologies, the deployment of video surveillance system exists in day-to-day life. Understanding the concept of human activities from videos are a challenging part of the computer vision community, which involves observing and acknowledgment of human movement designs assembled by movement sensor is a moving exploration theme [1]. Gestures, actions, interactions, and group activities have been classified into four levels dependent on their intricacy. Motions are simple movements made by a person's body parts. The actions of two or more people are defined as interaction. Interaction can occur between people or between people and objects. Concurrent interactions are those that are carried out by multiple people at the same time. Recurrent activity is defined as an activity that occurs on a regular basis for a set period of time. The interest in behavior analysis has raised drastically throughout the entire period, where video summarization and video retrieval are still hot topics of the research [2].

Human activity recognition is used in various aspects of domains like robotics, computer engineering, healthcare and natural sciences. It intends to give precise outcomes from discovery of human conduct by executing on the predefined model [3]. A video surveillance system/CCTV is composed of system of cameras, monitor/display units and recorders. These systems can be placed in interior and exterior areas of buildings. The real time footage can be monitored by security guard, or can simply be recorded and stored in the Digital video Recorder (DVR) or Network video Recorder (NVR) [4]. Human monitoring of surveillance videos is a difficult task to recognize the abnormal behavior or activities of human. Intelligent video surveillance system (IVSS) has been emerged to detect human activities in order to reduce the workload of video analyst and to extract accurate information from the surveillance video. In addition, police work cameras defend against property stealing, and trickiness. It's horribly hard to prompt away with taking a certain something, if their unit cameras film photography you. Even though GHMS (Global Mobile Home Security) can be used to achieve high security [5].

Different sorts of sensors are utilized in HAR framework like advanced cameras, profundity sensors, wearable sensors and gyro sensors [6]. There are mainly two steps involved in the sensor-based system. The first step is to calculate the applicable highlights dependent on the sensor's information. In the subsequent stage, a picked arrangement calculation characterizes the action as per the highlights acquired in the initial step. A few highlights are removed from signal examination, wavelet investigation and recurrence area examination. Human behavior is modelled as a random sequence of events. Previously seen actions from image feature databases are used to recognize actions. The location of the person can be used to represent behavior. In different situations, the same action sequence can represent different behaviors. Person interaction with a static object, individual cooperation with someone else, or communications between gatherings of people can all be used to determine human behavior [7]. His behavior could also be modelled using a suitable learning algorithm to detect anomalies. Global features are those that are extracted directly from videos as a whole and encoded as a whole feature. The personal title is localized and separated from this reflection by using basic drawing techniques to plan the framework or shape (e.g., Region of Interest (ROI)). Other global approaches, such as angle perception, edges, or visual distribution, are viewed as global definitions. Other outline based worldwide portrayal strategies incorporate 3D space-time volumes, which are then used to address the outline picture along the time pivot. Discrete Fourier Transform (DFT) is a worldwide methodology that gives recurrence space data of return for money invested for acknowledgment. Because of their sensitivity to noise and occlusions, global feature descriptors are gradually becoming obsolete [8]. A novel biometric framework which uses different angles between the skeleton joints to detect the human activities and is totally based on the RGB depth sensor data, this approach helps the surveillance system in many aspects. The sliding kernel method used to store the particle joint angle pairs. The ubiquity of communication devices such as smartphones has led to the emergence of context-aware services that are able to respond to specific user activities or contexts [9]. By and large, the handling structure of an Intelligent Video Surveillance System (IVSS) comprise of the few phases: Movement/Object recognition, Person identification using gait, Activity Analysis, and Behavior Understanding. A proposed thresh holding methodology utilizing the opposite of HAAR wavelet change (HWT) is used to support better definition signals. HWT is commonly used for channel audio, to reduce information size, and to detect quantity. In particular, the fundamental objective is on the Individual Recognizable proof and exercises that include connections between individual to individual inside the climate. Hence, HWT is extremely successful in measurement handling. This thesis is compiled as; In section 1 provides a quick introduction to the proposed lesson. In section 2, HAR applications including the previous system are reviewed [10]. In section 3, the proposed calculation of the acceptance of the action is actually tested. In section 4, the individual test status is discussed. within the latter area, it discusses the benefits and disadvantages of the proposed research as well as improving some of the best practices. A robust object tracking algorithm that takes advantage of both holistic templates and local representations to account for drastic changes in appearance [11].

1.2 An Overview on Person Identification

The most important stage in video surveillance is to recognize the authenticated user from the progressing recordings [12]. A video examiner analyzes the recording to determine who is in the entire scene and what they are doing on your own. However, this approach must be automated. In order to automate this in surveillance applications, people can be identified either by their presence or by their behavior**.** As a result, in this thesis, biometrics-based person identification from videos has been introduced in order to uniquely identify the person. People are identified according to their physical characteristics. Some documents have been proposed in the field of safety frameworks. Each security framework application is utilized for an alternate reason, and there are a couple of studies in the composing that utilization a camera in IoTs and sensors. The Web of Things could help expert and understudies learn and exhibit different data by social affair information and deciding the best course to prepare. Subsequently, they can assist with securing the rights and individuals of urban areas. Map making, or the creation or investigation of guides, is another utilization of PC Vision. The most notable illustration of map making is Bing Maps or Google Maps, the two of which are helpful administrations for individuals. Biometrics refers to both physical and behavioral characteristics. Biometrics had advanced to the highest level of security in a variety of situations. Physical biometrics include physical characteristics. The individual can be distinguished by the manner in which the person strolls, for example the individual's walk. As a result, gait recognition from video plays a critical role to verify the individual in order to tighten security. Furthermore, walk acknowledgment can be utilized in one or the other check (validation) depending on the situation [13].

This Person Identification using Gait System has a separate training and testing phase. During the preparing stage, highlights are extricated, intertwined, and they study training images [14]. The same procedure was used to test images during the testing phase. In addition to the metric learning step, the distance between training and image testing is measured to determine similarities between people, leading to a unique personal ID in the final stage. When there is an occlusion and the person is walking at different speeds, identifying the person becomes difficult. Similarly, PC Vision is modified to perceive explicit pieces of guides, like streets, water, structures, or fields. The illustration of how Computer Vision can be applied in Smart earth, showed a savvy checking framework that utilized a PIR sensor, a Raspberry Pi, and a cell phone. They ought to likewise utilize a smoke alarm to recognize fires. After sending the picture to the client's email through Wi-Fi, the client will be advised of the fire or hoodlum. For movement recognition, they utilized the smoke disclosure calculation and the foundation deduction calculation. The knees are not apparent in the recordings during impediment, which may decrease the acknowledgment rate unexpectedly. In such cases, considering novel static and dynamic highlights can bring about further developed personality. Observing assumes a critical part in guaranteeing our security. Security framework sensors assume a basic part around here. There are two sorts of sensors: indoor sensors that guarantee and secure the inside of the home and outside sensors that guarantee and ensure the outside of the home. Microwave sensors, Passive Infrared Sensors (PIR), photoelectric sensors, commotion identifiers, and ultrasonic sensors are the most well-known. The remote framework gives a fit, a la mode, and amazing answer for the issue of far-off house access, security, and checking with face discovery. To affirm or dismiss the individual, the check mode utilizes a 1:1 match between the put away format (exhibition) and the test layout. Following the Person Identification, the action taken by the person in the specific scene is critical. Activity recognition/analysis of the specific person is required in order to analyze the behavior of humans from videos.

1.3. Human Behavior Activity Analysis

A high-level activity analysis is required for the video surveillance application, which consists of multiple atomic actions of individuals. Gestures, actions, interactions, and activity can all be able to be utilized to recognize an individual's behavior. Human Behavior Analysis (HBA) is a significant part of a video observation framework. Programmed location of strange exercises in an observation climate can be utilized to caution the applicable authority of expected crook or hazardous practices, for example, programmed announcing of an individual with a sack dallying at an air terminal or station [15]. Moreover, in sports, watchers need a more expanded visual portrayal of the game for better arrangement, while mentors and players need to separate important data for execution and strategy improvement. With the expanding interest for exactness in choices made by refs/umpires/authorities, smart video handling frameworks are turning into a distinct advantage in sports. object tracking is more difficult and trickier to surveillance in real time [16]. The objective of human movement discovery is to investigate continuous action from an obscure video. Single-person actions, such as walking, kicking, standing, and so on, have a meaningful interpretation. Interactions are activities in which multiple actors and/or entities participate at the same time Interaction Recognition also recognizes interactions that occurred concurrently between people as well as activities that occurred recursively.

The two types of interactions are considered for detection are Human-human and Object-human interactions. Handshakes, hugging, and kicking are examples of simple interactions, whereas someone attempting to approach the other person is an example of a more complicated interaction. To analyses the scene at various levels of abstraction, a number of processing steps are required, beginning with the behaviors of objects of interest [17]. The initial step is to recognize and follow the subject(s) of premium to produce movement portrayals (e.g., movement direction or mix of nearby movements), which are then prepared to distinguish activities or collaborations. Position data can be enhanced with different descriptors, for example, body joint directions or changes in head present, contingent upon the nature of the camera see. A gathering action is characterized collectively of individuals or potentially protests acting in similar setting to accomplish a few objectives; it's anything but restricted to a solitary one and every one can have its own objective. So, the objective of visual reconnaissance isn't just to supplant natural eyes with cameras, yet additionally to mechanize the whole observation task however much as could reasonably be expected. HBA incorporates stages, for example, preprocessing, following of people, include extraction, assessing body posture, and understanding human practices from a grouping of pictures. HBA is a significant application in sports reconnaissance and conduct investigation, in addition to other things. Sifting approaches were utilized to finish the following task [18]. The molecule channel- based strategies are fundamentally used to keep up with numerous probabilistic speculations about body act. Posture assessment should be possible utilizing either a stick figure model or movement history picture highlights. Foundation mess, dynamic foundations or conditions (indoor and outside), brightening change, and camera developments, among different difficulties, exist in the pre-handling stage. The most widely recognized difficulties during the following and highlight extraction stages are appropriate component choice, low level element extraction, and impediment examination, among others. Movement investigation, saliency calculation, changes fit as a fiddle and posture, diverse dressing styles, camera see points, and multi-camera combination, in addition to other things, are considered in present assessment. Prior days were centered around perceiving low-level human exercises like hopping, running, and waving hands. In the real world, people interact with one or more other people and objects. Interactions can be between humans or between humans and objects. The difficulty lies in either locating or tracking multiple subjects at the same time, or in recognizing the entire human group's activities as "walking together" rather than "walking." Some two people in the group perform one interaction, while another two people perform another [19]. Computer vision has evolved in the last decade as a key technology for numerous applications replacing human supervision.

The classification of sequential images remains a persistent challenge. As a more explicit test, the exhibition of the framework is vigorously reliant upon the grouping or taking in human conduct from video information, which has been widely examined in this proposal. Learning and comprehension of standards of conduct can be considered as the arrangement of time-changing component data [20]. The crucial issue of conduct understanding is taking in the reference conduct successions from preparing tests and successfully conceiving both preparing and testing strategies with little varieties in the component information inside each class of highlights. Ongoing techniques, like HMM and CRF, endeavor to catch this design and learn activity portrayals. These portrayals are just scholarly at the level of a couple of video casings, and they neglect to demonstrate activities at their full fleeting extent. This motivated us to create a human interaction analysis system as well as a model for concurrent interaction analysis that could predict actions/interactions taking place at the same time.

1.4 Chaotic Maps

In a dynamic setting, chaotic maps are frequently researched because of their chaotic nature. It indicates that even a minor alteration in the starting circumstances may have a large impact on the results. Chaotic maps are used or implemented so that the entropy generated by the map can produce confusion and diffusion. This thesis has been presented by using the technique 5D chaotic map. Higher the dimension is preferred because of their secured nature and complex trajectory i.e, dimensions of the image define initial attributes of the image required for 5D chaotic map. Image encryption using chaotic map is presented in figure 1.1.

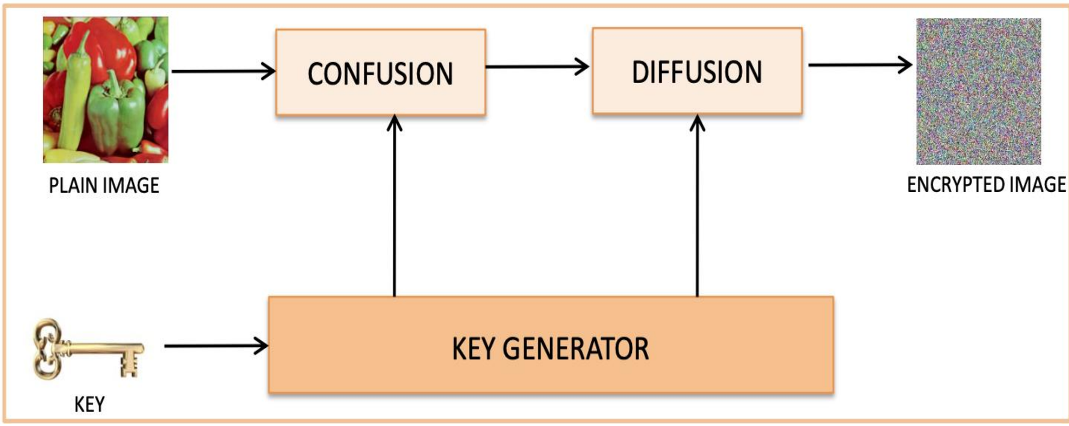


Figure 1.1: Image encryption using chaotic map

There are different types of chaotic variants addressed by the current research as shown in Table 1.1.

Table 1.1: Types of chaotic variants

S. No	Chaotic Variants	Remarks
1	2-D Rational map	Discrete in TD and rational in SD
2	3-cells CNN system	Continuous in TD and real in SD
3	Arnold’s cat map	Discrete in TD and real in SD
4	Baker’s map	Discrete in TD and real in SD
5	Chen attractor Circle map	Continuous in TD and real in SD
6	Exponential map	Discrete in TD and complex in SD
7	Hyper Logistic map	Discrete in TD and real in SD
8	Henon map	Discrete in TD and real in SD
9	Logistic map	Discrete in TD and real in SD

S. No	Chaotic Variants	Remarks
10	Lorenz attractor	Continuous in TD and real in SD
11	Tent map	Discrete in TD and real in SD
12	Tinkerbell map	Discrete in TD and real in SD
13	Zaslavskii map	Discrete in TD and real in SD

1.5 Genetic Algorithm

An evolutionary strategy for addressing non-deterministic polynomial time (NP) complete problems is the genetic algorithm (GA). Picture encryption may be considered an NP-complete problem because of the approach used. The ideal parameter settings for successful encryption are determined using GA. GA basically works on some solutions consisting of population where the size of the population is the no of solutions. Each solution has a chromosome. The feature or parameters that define the individual are referred to as "the chromosome".

The fitness level of the solution is used to select the best individuals for mating pool. The individuals with higher fitness level have a higher probability of being selected in the mating pool. It is predicted that by selecting the best individuals, the mating pool will produce higher-quality offspring. By continuing to select and breed high-quality individuals, there will be a greater probability of maintaining only the good qualities of the individuals and discarding the poor. Finally, the required optimal or acceptable solution will emerge. [21]

However, the offspring who are now being produced using the chosen parents only have the characteristics of their parents. To address this issue, some modifications will be made to each offspring in order to generate new individuals. The new population, which will replace the previously used old population, will be made up of all newly created individuals. A generation is the term used to describe the population that has been generated. The fitness value of a chromosome can be calculated by the process called evaluation [22].

Initilization

After getting how to represent each individual, the population must be initialised by selecting the appropriate number of individuals.

Selection

The next step is to choose a few individuals from the mating pool's population. The best individuals are chosen based on a threshold depending on the previously estimated fitness value [21].

Variation Operators

Parents are chosen for mating based on the individuals in the mating pool. There are a number of variation operators to apply for each pair of parents selected, including:

Crossover (recombination)

In crossover, parents swap genes from their previous generation. The resulting offspring get the genes from both their parents. The resulting generation is called offspring after the parents exchange the genes[21]. The chromosomes that arise are known as offspring. Single-point crossover is the name given to this operator.

Mutation

Following the crossover period, the children are subjected to a mutation phase. When each gene has a random chance of changing, this is known as mutation. This allows for exploitation because the children's solution will not alter as dramatically as it did during the crossover period, but they will still be free to explore around the present solution's neighborhood. The flowchart of GA is shown in figure 1.2.

1.6 Differential Evolution Algorithm

In evolutionary optimization, differential evolution is a well-known method. In this program, the mutation and crossover operators of GA are cleverly coupled. By comparison to a standard GA, it achieves faster convergence. Three fundamental steps are involved in the process, starting with the establishment of the population and ending with selection and terminating it which have described below. Evolution Algorithm uses biological evolution such as mutation, reproduction and selection. Candidate solutions to the optimization problem play an important role in individuals of population, fitness function determines the solutions of quality. Traditional algorithms differ from EAs in that EAs are not static but dynamic, meaning they can change over time [23].

Three fundamental steps are involved in the process for evolutionary algorithms.

Population-Based:

The goal of evolutionary algorithms is to improve a process where current solutions are ineffective in order to develop new, better solutions. The population is the set of existing solutions from which new ones will be generated.

Fitness-Oriented:

Each unique solution has a fitness value that is calculated using a fitness function. This fitness number represents the quality of the solution.

Variation-Driven:

To make new samples from the search space, apply variation operators to the selected samples. Two main variation operators are recombination and mutation.

Using this common structure, the evolutionary algorithm searches the search space for promising solutions [24].

Initialization

Create a population of samples at random from the search space.

Iteration

- **Evaluation**- Create a population of samples at random from the search space.
- **Selection Operator** - Select the samples to be used in the following phase based on the values of the objective function produced for the evaluated samples.
- **Variation Operator** - To make new samples from the search space, apply variation operators to the selected samples.

Termination

Stop the computation if the termination criteria are met.

Figure 1.3 depicts the differential evolution flow chart.

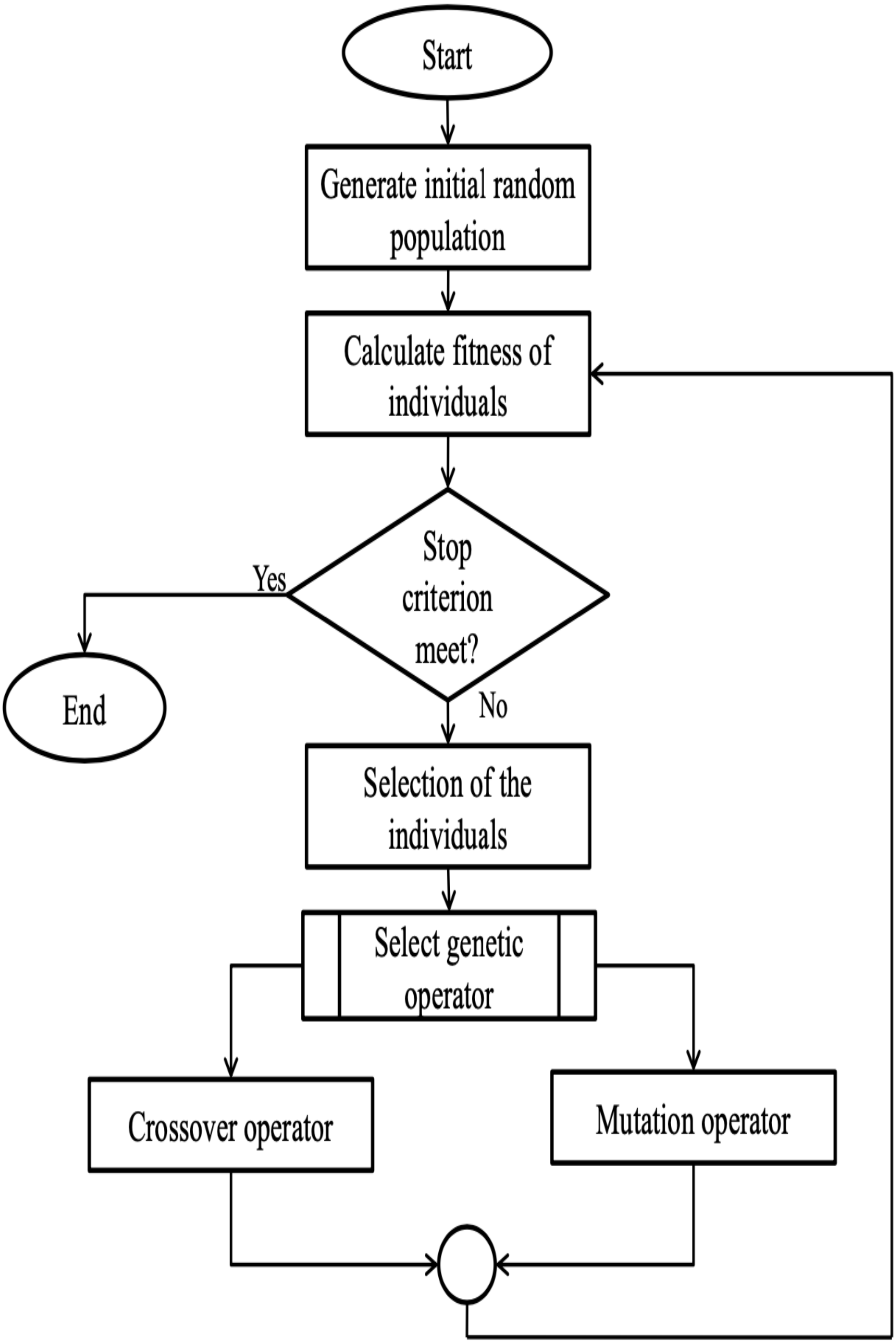


Figure 1.2: Flow chart of genetic algorithm

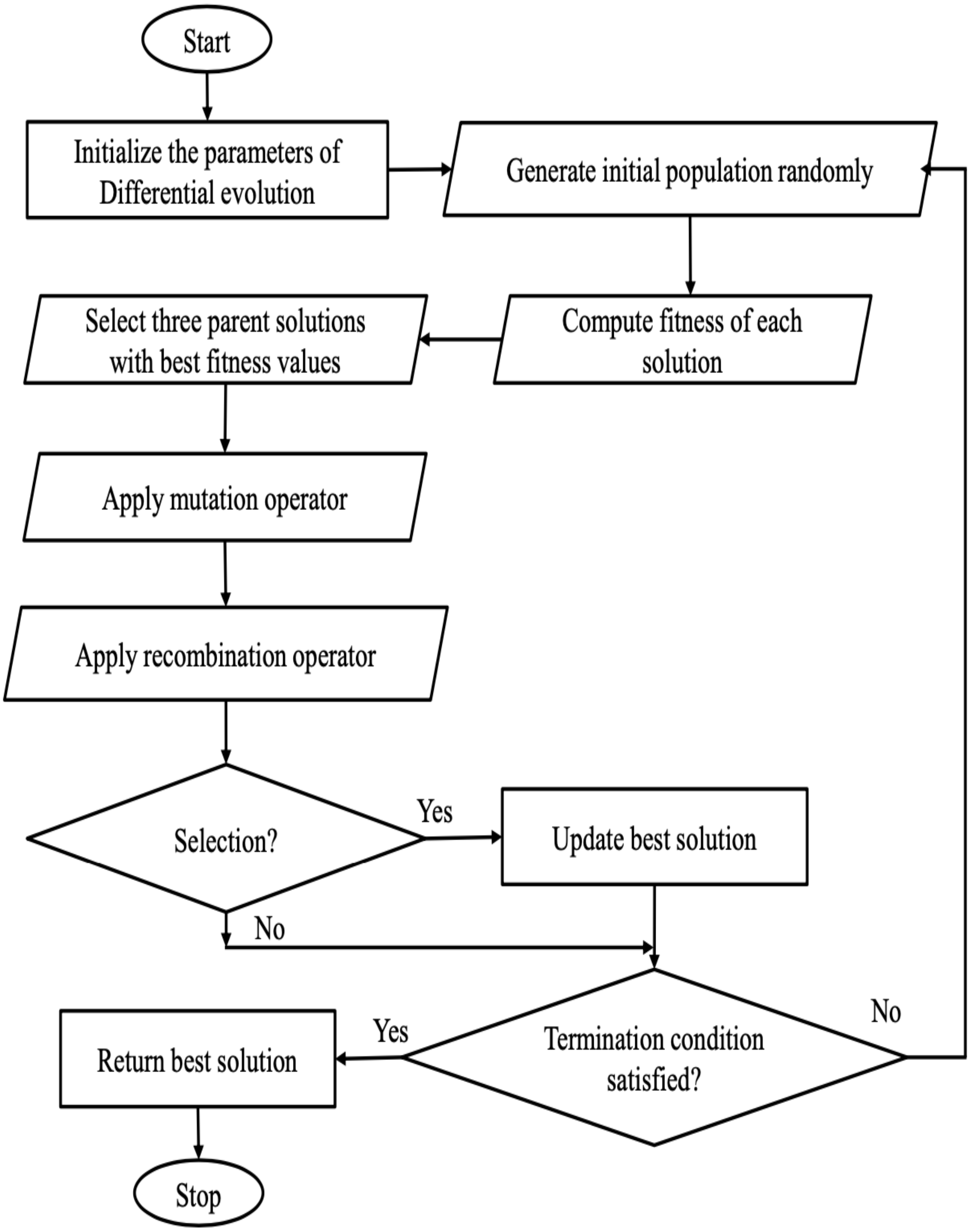


Figure 1.3: Flow chart of differential evolution

1.7 Objectives of the Thesis

By conducting the review of recent research on secure surveillance systems, we have observed various gaps. In order to overcome these gaps following objectives are formulated:

1. To analyse the performance of existing surveillance systems and various image encryption techniques.
2. To design and implement efficient meta-heuristic based 5D chaotic map image encryption technique to secure the images.
3. To design a secure framework to prevent unauthorized access in the existing surveillance frameworks.
4. To compare the proposed technique with existing machine learning models.

1.8 Organization of the Thesis

The rest of the thesis is organized as follows. In chapter 2, the review of literature is discussed from where the comprehensive summary of a previous topic has been enumerated, describe, summarize, objectively evaluate and clarify the previous research. In chapter 3, the research methodologies are discussed where the specific algorithm or technique used to process and analyse the information of the topic. Partial- regeneration based non-dominated optimization (PRNDO) has been used because it will provide optimal solutions considering both best and worst cases. This type of optimization is mainly used to solve the asymmetrical problems. In this, bottom-boosting methodology is utilized to reduce the number of function evaluations. In chapter 4, the experimentation results and discussions are presented where by using the proposed algorithm different analysis results are presented like Secret key, Comparative analysis, Visual analysis, Robustness against various attacks, key analysis, Security analysis, Fitness evaluation. For this analysis different model has been taken like 2D Logistic Sine System (2LSS), Cosine-Transform based Chaotic Map (CTCM), Discrete Fractional Random Transform (DFRT), Compressive Sensing and Chaotic Map (CSCM) etc. In chapter 5, the conclusion and future scope is discussed. An image encryption technique for secure surveillance for IoT systems has been proposed in this thesis.

Chapter 2

Review of Literature

The adoption of non-core functionality is a major problem for AI because it relies heavily on other factors such as feature selection in continuous data allocation training and standard separation methods, and it is difficult to determine the final size data training and machine size [22]. There are two parts on CNN. The first part is to remove the sequence object, which includes convolutional layers and pooling. Each layer input is to remove the layer before it. As a result, the actual signal has been converted into feature carriers. The second part is a fully connected layer, and a fully connected layer separates the vectors of the object. CNN is the most widely used method of in-depth learning in the most common computer field, and in the realization of human activities through clothing [23]-[24]. CNNs tend to have several hidden layers that use convolutional filters to extract anonymous input data presentations. The following sections are included in the full process of monitoring activity: video detection, pre-processing, tracking, feature removal, classification, and recognition [25]. The first stage of preparation is to remove the shadow from the video frames and represent it properly [26]. Object recognition is one of the fundamental challenges in computer vision. To find the moving objects in each frame, a background extraction process is used. The cell phone tracking process is followed, and a blob tracker is used to do so. Feal ad is a critical stage in recognition. Model-free or model-based methods are used to test features. Model-based methods are used as prior knowledge to interpret human strength and predict movement limits. Many motion capture programs are currently available in both 2D and 3D analyzes. The analysis of kinematic gait real-time can be measured in two ways: directly or by contact-based techniques, or optically or by non-contact techniques [27]. Sensors such as accelerometers and goniometers are used for precise measurement. Gait analysis, optical techniques using active or hidden signals [28]. The tag is known as the active tag. A novel gait recognition technique based on the image self-similarity of a walking person. It uses a Light Emitting Diode (LED), and the position of the marker is determined by this signal. The active marker has the opportunity to identify its predefined waves. Markers that reflect the light of events are known as artificial symbols. People can be seen in shopping malls and smart neighborhoods using gait-based methods. Unwanted criminals can be avoided using a gait-based security system [29]. Human pose and motion estimation is formulated as inference in this graphical model and is solved using Particle Message Passing (PAMPAS) [30]. The interaction of the entities in a video is formulated as a graph of geometric relations among space-time interest points [31]. The detection of instances of an object class, such as cars or pedestrians, in natural images [32]. This program can help with security programs, remote monitoring, and so on. Instead of a neurological examination and without clinical impairment, the system should determine whether the person is normal or has a video neurological problem [33]. Currently, object tracking and detection are done in a manner similar to a shallow learning paradigm. This method involves building an object model and then applying various algorithms to estimate its parameterized structure [34]. Fingertip detection is a broad application of this technology. It involves determining the states of a target based on its movements [35].

2.1 Research Problems in Gait Analysis

Research on gait identification is encouraging, however there are a few matters that may lessen the effectiveness of the machine. Some of the downside elements are categorized as troubles or demanding situations within the biometric security system [36]. The factors affecting the visual excellent of a system may be divided into two classes: External and Internal features.

- **External Features**

Factors affecting the outside notion are known as the challenges of gaining reputation. View constant motion recognition, for example - distinctive viewing angles which includes front and facet view, distinctive lighting situations together with day and night time, specific areas such as out of doors / indoor, garb style together with skirts or long coats, special tour conditions along with heavy / soft grass, dry / moist / concrete, ground / stairs, diverse styles of shoes consisting of boots, sandals, individual sporting items which includes backpack, and bag [37].

- **Internal Features**

Factors affecting the internal recognition system and so-called gait recognition challenges [38]. For example, a person's natural behavior may change as a result of foot injury, paralysis, Parkinson's disease, or other physical changes caused by aging, alcoholism, pregnancy, obesity or weight loss, and so on. The tracking accuracy of the proposed algorithm is improved by updating the prediction of the target position with a trusted algorithm. This study is motivated by these internal and external factors to look at approximately of the outside factors that may have negative effects. For example, a fixed and dynamic closure can affect a person's identity and system performance. These are research issues that need to be addressed when developing a credible recognition system. A human action recognition method based on a hidden Markov model (HMM) is proposed [39].

2.2 Template Based Learning

Template-based methods use a variety of presentations to describe the most common aspects of the appearance of a particular task. Templates are common visual objects like 2D/3D static pictures/volumes or succession of visual models [40]. Numerous layout-based techniques produce 2D/3D static formats and contrast picture/video yield layouts with test recordings. Dynamic Time Warping (DTW) is an effective system for the division of large independent segments. In each work template, two categories are created: Motion-Energy Image (MEI), which deals with the presence of motion, and Motion-History Image (MHI), which indicates the arrival of development. In reality, the layout pictures delivered can be considered as models with the heaviness of time. The 3D space video layout utilizing short video preparing time [41]. This layout is little contrasted with all pieces of similar size in 41 test recordings in each of the three measurements. The unique applied technique surveys the level of likeness between the two segments (i.e., a format and a video part of a similar size from the test video) [42]. This strategy is appropriate for various undertakings that happen at the same time. Normal format-based strategies can't create a solitary layout for each errand. In order to avoid the constraints of coplanarity, we consider the cross ratios of frames when constructing invariants.

2.3 Generative Learning

Person-centered actions, space-based actions, sequence of actions, and ultimately behaviors are all used to represent human behavior. The current action sequence is defined by a predefined HMM set and a possible calculation set [43]. Use Bayesian networks and propaganda at the highest levels, as well as nonparametric sample from a previously read database of actions at lower levels. The most likely action was calculated in each category. Human actions are a sequence of different ideas of the human body moving at different times. The authors used only the skeletal features provided by the RGB-D camera in this work. As a distribution tool for many Gaussian countries, Gaussian blends help collect data from different groups. Because of this, individual body composition can be defined as an international distribution group, and HMM can show the dependence of an internal clip between periods. It focused on understanding the sequence of long actions to see people's daily activities. The authors have proposed a method of obtaining video sequences using reduced Fisher Vectors (FVs) in conjunction with structured temporary models. It shows that FV production facilities make them especially ready to integrate production models such as Gaussian mixtures. To reduce the vector size of the feature, using Principal Component Analysis (PCA). The authors have included comparisons of compact videos based on FVs and HMMs to improve the accuracy of job recognition. The HMM is very important in finding vision-based activities. The Layered Hidden Markov Model (LHMM) can acquire various levels of temporal information while detecting human activities [44]. LHMM is a cascade of HMMs, each with different monitoring opportunities considered at different times. The HMM to recognize the function of observing the sequence of symbols in video frames at various times in the context of the Analysis of Human Behavior. Visual tracking, in essence, deals with non- stationary image streams that change over time [45]. When a model is immersed in a particular type of activity, it provides a higher chance of sequencing the characters closer to the task being studied. Extended HMMs to indicate the duration of each condition in the temporary transformation of functions. These types are resistant to time changes and time variations in work performance. This model, however, has no information about local construction [46]. This spatial structure can be important in determining, such as determining whether a movement arises from the upper or lower body, whether the two parts meet or are lost in parallel motion. The clustering of normal activities is detected through a self-tuning spectral algorithm that uses unsupervised model selection.[47] Most of the existing approaches for camera motion detection are based on the affine motion model.[48]

2.4 Discriminative Learning

Methods of discriminating models are guided by past opportunities. With enhanced functionality, this model emphases on computational resources for a precise task. It is very expensive to read, and spelling (Viterbi method) has to be done during training and often. This model is easy to build and only focuses on modeling the background opportunities no matter how the data is distributed. In small data sets, discriminatory models often provide less accuracy. Asset detachment, Support Vector Machine (SVM), close neighbor, Conditional Random Field (CRF), and neural networks are some popular models [49]. A novel method is presented for informing human activities [50]. The author proposes a sequential method of identifying classes of ten tasks. Ten classes were divided into five divisions in this work. To classify different function classes, Last Squares SVM (LS-SVM) and NB algorithms were used as dividers. Functions are identified using factors like as mean, variance, magnitude entropy, and angle between the triaxial accelerometer signals [51]. CRF is a framework for labeling and separating structured data that can handle multiple features in a single integrated model. Job recognition is a structured domain where recognition and labels arrive in sequence over time, and random conditional fields are ready for job recognition.

CRFs, in particular, are non-targeted graphical models where the graph structure forms an independent relationship. In other words, the structure of the graph incorporates the independent relationship between the labels rather than the perceived. As a result of the assumption that the model does not have an independent relationship between perceptions, the tendency to CRFs remains affected even if it incorporates complex and contradictory aspects of observation [52]. CRF was the modeling group activities. For example, in group speaking activities, people should keep in touch with people face-to-face while removing noisy connections like as people looking in the similar direction, which will confuse speaking with lines. Given the picture, the Y activity groups are signified by Y, which are indices for classes such as speaking, queuing, and so on. Actions are signified by the letter h. Use the element level definitions, defined by x, in both activity y and actions h. They used HOG in this case [53]. The model is an MRF model, and includes opportunities for integration as a whole for the weight of clique and strength. The first set of cliques includes the appearance of work and actions. The occurrence of the dual act of action in the next set of cliques. The connection is indicated by a dashed line. This is because the elements of this layer are treated as hidden variables in the method are robotically generated during read and capture.

2.5 Initiative Learning

Simulation learning is just an example of a spontaneous automatic. NB - logistic regression pair, as well as HMM and CRF, popular models. The paired model is defined as a form similar to line dividing, but the parameter is limited indifferent ways [54]. NB estimates the combined possibilities p (x, y) = p (y) * p (x | y) from feature training data x and label y and uses Bayes law to predict p (y | x) in new testconditions. Asset retrieval, on the other hand, measures p (y | x) directly from training data by reducing error function. The discriminatory power of the Artificial Neural Network (ANN) or SVM with the anti-HMM dynamic potential could produce better outcomes for dynamic problems. The proposed model in their work is an integrated HMM method where the HMM status model reflects the temporal data features and ANNis used to measure HMM state distribution. Various types of HMM hybrid systems havebeen used successfully on a large video data set; the most popular method of combining HMMs with ANNs [55]. The author has used the integrated HMM / ANN method to predict time series data, which has led to a model that provides better sequential classification. Hybrid-based ANN variants are also widely used for a variety of recognition functions, including speech recognition, handwriting recognition, sentence recognition, and digital recognition. Online object tracking is a challenging problem due to the complexity of the model selection process. It involves learning an effective model to account for the appearance changes caused by intrinsic and extrinsic factors. [56].

2.6 Simulation

The mimic regression and instability using the HAR data production process (1,5,22) [20]. Inaddition, because this study looks at instability in a specific indicator, we imitate the HAR model (1,2,5,10,22) [56]. Following that, we compared the series with the actual revision ofthe daily SPX index and saw a tense situation. The simulation was based on two-hour volatility and uncertainty, but because we can only obtain daily returns and obtain inconsistencies from Oxford Realized Library, we prefer to simulate daily returns and uncertainty to make appropriate comparisons of known SPX index values 7 and -8 usedto imitate HAR (1,5,22) and HAR (1,2,5,10,22), respectively [58]. We subtract the valuesc, (d), (t), (w), (b), and (m) from the 7th and 8th points of the SPX index during the sample period from January 2000 to May 2020. (d) has a zero definition and a standard deviation equal to the standard deviation of the scale HAR estimates in the order of fluctuations of the SPX indicator in equations 7 & 8. (d) t with the aforementioned distribution. Subsequently, using the standardized error (d) t, we created a one-day variable in advance using the data processing process described in equation 7 and 8. Because self-esteem by definition is not a bad thing, we must ensure that inserting the wrong word (d) does not cause negative inconsistencies the day before [59]. When this happens in imitation, we produce a new error that is already randomly distributed until one day in advance sees the instability as right. Activity recognition fits within the bigger framework of context awareness. While this is not a complete approach, it does notprovide a clear 11th definition of his approach, and we believe that at present, this is thebest way to insert random errors in determining one-day variables. The transitions fromdifferent states can be made based on a position of a person, scene change detection, or an object being tracked [60]. We produce dual variables with a sample size of available SPX data to reduce the influence of startup values. We use only the last half of the estimated values in any HAR data-generating process when comparing the apparent return and variability of the SPX series and both the simulated HAR series [60]. After determining the apparent flexibility, we mimic the return and calculate the refund using the following formula:

$$r(d)t = RV\,(d)t \,¤\, (d) \qquad 2.1$$

In this calculation, r (d) t daily measured return on t, RV (d) t day after day is measured not equity attempts in equations 2.1, and the error name (d) t is the most common method of distribution”.

2.7 HAR Approaches in Terms of Feature Removal Process

Handmade techniques rely on human ingenuity and prior knowledge to end prejudice. These methods include three main steps: (1) pre-acquisition, which is consistent with the classification of action, (2) material selection and professional production, and (3) action planning according to published factors [61]. However, for all applications that require real-time feedback, it is important to build a standalone device that can simultaneously detect multiple magnetic signals over a long period of time and identify tasks performed very quickly. In this regard, the device you want should be lightweight, lightweight, and easy to wear, if possible, equipped with a long-lasting battery, and equipped with a microcontroller with sufficient internal memory and signal storage, and able to support the implementation of job recognition separation. The body of multicamera human action video data with manually annotated silhouette data that has been generated for the purpose of evaluating silhouette-based human action recognition methods [62]. The purpose of this study was to compare two sets of real-time HAR activities: Feat Set A, which incorporates time, frequency, and time limits expressed in the literature, and Feat Set B, which incorporates time zone variables based solely on understanding how a particular activity would affect sensory signals. The most knowledgeable features in each set are identified using the GA feature simultaneously with the parameter efficiency. Feature subsets were obtained compared to the performance tests of machine learning classifiers. This family of methods is less expensive, more flexible, and does not rely on large sets of training samples because the classification is done using standard trained partitions. Because the duration of a person's work varies and the exact parameters of the work are difficult to define, the continuous fragmentation of emotional sequence data is a difficult task. Robust multi-object tracking by-detection requires the correct assignment of noisy detection results to object trajectories. Currently, both HAR detection methods and default HAR methods detect sensor signal from time to time using window technology. The partition algorithm is used to identify the function in each window [63]. Window strategies include custom slide windows, event description windows, and behavior description windows. Window samples may not always have the same label. A very sparse measurement matrix is constructed to efficiently extract the features for the appearance model [64]. To generate a label for all window samples, you use a quick delivery window labeling method. However, because the most visible category has been selected to represent the window, some samples will be labeled negative. This problem is known as a multi-classroom window problem, and is caused by window cracks and labeling. The problem of multiple study windows is a common problem in the HAR timeline, and has a major impact on finding a sequence of temporary tasks [65]. Space- based methods, as discussed below, visual-based methods, and other hand-based methods include: binary LBP pattern, visual dictionary used for text editing and abstract concepts [66].

Human activities can be seen in two ways: static and dynamic. Solid activity is defined by the shape and shape of the organ in space, while dynamic functions are defined by the movement of these vertical functions. As a result, action recognition can be based on spacecraft or temporary signals that define and identify these actions. Both local and temporary representation can be divided into three categories as shown in figure 2.1. [67].

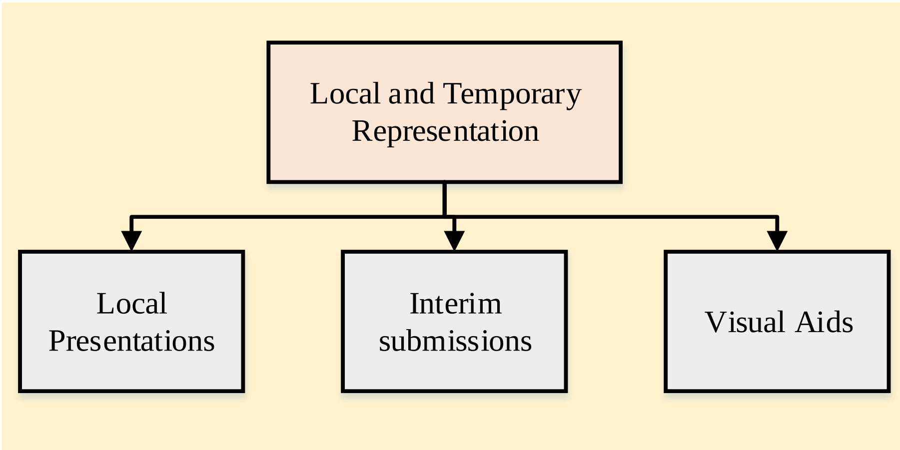


Figure 2.1: Local and temporary representation

2.8 HAR Approaches in Terms of Recognition Categories

In general, human recognition systems have three main steps: Determining the process of determining which part of the body to follow or see; tracking process for connecting successive images; and identifying the process of translating semantics into local practice, position, and performance. This section is provided to represent a range of HAR-based strategies for each phase of research on alternative HAR approaches [68]. This assumes that the optional algorithm is determined by the representation of the selected task.

2.9 HAR Approaches in Terms of Machine Learning Monitoring

In-depth learning methods have been used successfully in programs such as the ubiquitous computer, health and well-being. Flexible and duplicate neural networks, in particular, transform the human data (HAR) field because of their ability to read semantic presentations in green input. To remove the common features, however, a large number of well-chosen details are required, which is a well-known function that is disrupted by private concerns and anchor costs. As a result, learning uncontrolled representation is essential to using a large number of unlabeled data generated by smart devices. In the many public databases available from HAR based on smartphones, ignored, and readable, we are exploring the proposed approach. Our approach is superior to or compared to fully functional networks, and is much higher than autoencoders. Significantly, in the case of observation, self-monitoring features significantly increase the acquisition rate, earning kappa points of 0.7-0.8 with only 10 labeled examples per class. Even though features are transferred from a different data source, we get the same amazing performance[69].

HAR is based on machine learning methods and is divided into three categories below the level of learning monitoring [70]: supervised, observable, and monitoring methods. HAR approaches in terms of machine learning monitoring is presented in figure 2.2.

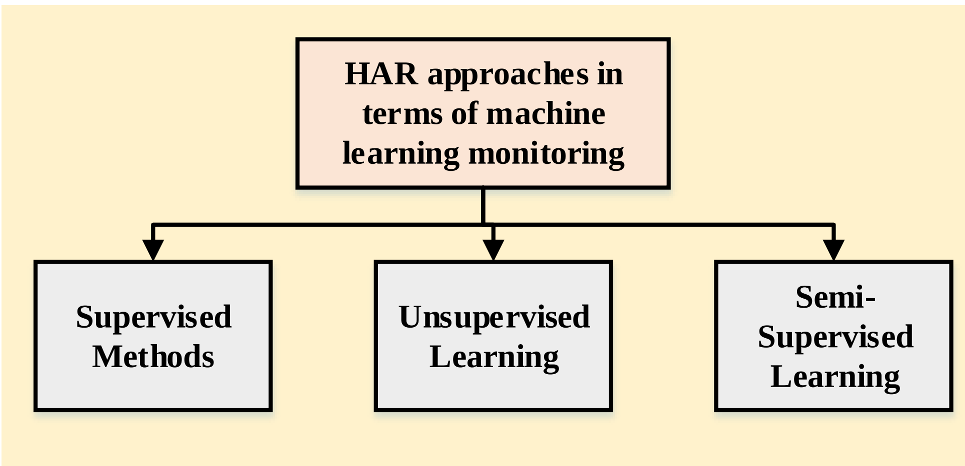


Figure 2.2: HAR approaches in terms of machine learning monitoring

2.9.1 Supervised Methods

In supervised learning, training takes place offline and is automated using labeled data to predict unlabeled data results. Teaching aids are used in HAR to distinguish and identify temporary actions. As the name implies, supervised learning involves the presence of a supervisor as a teacher. The supervised learning algorithm learns from labeled training data and can predict the results of unlabeled data. Accurate machine learning can be built, measured, and sent successfully. It takes time and technical knowledge from a team of highly skilled scientists to create a data science model. In addition, data scientists must rebuild models to ensure that the information provided remains accurate until the data changes [71]. These subclass methods are mathematically simple, highly accurate, and reliable. As examples of supervised learning, consider vector support, direct and indirect programs, random forest, and neural networks.

2.9.2 Unsupervised Learning

This means machine learning methods that do not require any kind of monitoring. These methods allow the model to detect discriminatory aspects of incoming label information using real-time learning strategies such as K-means and neighboring K-means. The timeline is one of the mathematical modes of unsupervised learning. Unknown parameters (of interest) in the model are related to the times of one or more variations at a time, so these unknown parameters can be estimated to be given from time to time. Times are often psychologically measured from samples. First and second time orders are basic. The Vector defined by the first order is a random vector, and the matrix of the covariance is the second minute of the order (where zero means). high orders. The modeling model, which is an example of word mathematical (visual power) in text-based (hidden power) document, is an excellent example of dynamic models in machine learning. When the title of a document is changed, text words are generated according to different mathematical parameters. Under certain assumptions, it is shown that the time method (tensor decay techniques) always returns the boundaries of a large section of flexible hidden models. These methods are computer-generated, accurate, and unreliable, for example, introducing a new algorithm that shows human activities in a completely ignored environment, allowing them to see every day and forgotten tasks. The visual model looks at the short-term and long-term action relationships and gains significant results in action and predictive action [70].

2.9.3 Semi-Supervised Learning

This term refers to integrated and uncontrolled learning methods. When label data is insufficient to produce an accurate model, training is performed with labeled and unlabeled data [71]. These approaches can benefit from the discriminatory power of controlled methods of distinguishing between the characteristics and strength of uncontrolled methods of automatic HAR actions. For example, it describes a hybrid framework for online recognition of everyday life activities. Appearance information is essential for applications such as tracking and people recognition [72]. The authors present a comprehensive picture of human activities through the integration of unsupervised people (representing land movements and local activities) and supervised learning strategies. For some services, it provides limited or ignored HAR frames. The analysis of human activities is one of the most intriguing and important open issues for the automated video surveillance community [73].

2.10 Image Input Versus Video Input

Human performance recognition research can be divided into two sub-categories of inclusion information

i) Man-Made Observations Based on Still Images

where the system can find activities such as accommodation, travel, food, etc. in pictures Functional features make it different from others [74].

ii) **Video-Based Human Activity Recognition**

Some functions cannot be found with a single image at this time. There is a need for access to additional information related to pre- and post-event events, for example, by examining previous and subsequent frames. The videos are more accurate in this case because the relationship between two consecutive frames can be established.

2.11 Single View Compared to Multiple View Acquisition

The One view is compared to the acquisition of multiple views. Digital camera regional view feature. Many researchers are interested in seeing the functions of only one tool within a relevant region, while others assume that many devices are placed in specialized locations to have different views on the work performed [75]. We can divide works of art into two categories:

i) Single view

ii) Multiple views acquisition

2.12 Security and Monitoring Requests

Staff monitor traditional monitoring systems. They should always be aware of human activities seen on camera. The growing number of camera deployments and views puts the work of providers under tremendous pressure, leading to lower productivity. Many traditional methods of identification depend on the structure of the object, but the use of those methods in a comprehensive, comprehensive system, and detailed views are difficult [76].

Insufficient energy is one of the most common problems with CCTV equipment, and it often becomes a source of abnormal or abnormal behavior. A clean, sufficient energy source is needed to run the right system. On hot days, for example, when the ventilation units are overcrowded with power grids, the power can change dramatically. For this reason, agencies should plan ahead and specify electrical installations or backups where necessary.

2.13 TI Method

TI is a method that allows users to measure their presence in a physical environment and interact in real time, as if they were in the same physical location yet geographically [77]. Those applications require a large amount of pc processing power and provide a large amount of information that must be transmitted in real time via the network. Compression techniques are often used to reduce the amount of information that will be transferred. for example, using mobility magnification, a multi-camera TI machine can extract body kinematic parameters in all frames from cloud statistics, greatly reducing public transmission between remote websites. In addition, 3-dimensional (3-D) videoconferencing video applications work well with laptop hardware environments that require 3-d real-time video conferencing as a result of computer algorithms.

2.14 AR System and Approaches

AR systems are now widely used in everyday life thanks to advancements in visual and sensor technology [78]. Over the last decade, scientists have used a variety of approaches to recognize human behavior in a wide range of application fields. In review, we classify AR systems based on their design methodology, with a focus on the data collection process.

2.15 Multimodal Sensors for AR Approaches

In the last decade, multimodal AR approaches have grown in popularity. They understand human sports by the use of both visible and non-visible sensors at an equal time. The type of model and the use of sensors can also vary depending on individual needs [79]. for example, a single camera can cover a large area of a specific context, but it cannot be sufficient to study relevant records including ambient temperature, humidity and consumer records containing heart rate. honestly, systems that use the single modality sensor process will no longer work well in situations that require this kind of wide range of input facts.

2.16 Activity Recognition Methods and Techniques

Following Object Detection, Segregation and Tracking, and Feature Disclosure, we can broadly distinguish research into performance and visual behavior, particularly the detection of abnormal behaviors, as monitored, unsupervised, and under surveillance. According to the labeled data, surveyed methods form models of normal or abnormal behavior [79]. Video segments that do not match the "flagged" models as abnormal. This method of finding unusual events is only effective if the unusual events are well described and there is enough training detail. Unsupervised methods, on the other hand, have learned common and unfamiliar patterns in visual data structure structures. The anomalies are identified as individual collections. Semi- guided trails fall somewhere between the first two. Using data with a small label, they learn a typical / unusual event model. To combine features, a k-means algorithm is used. The integration and learning process of integration acquisition involves combining the same descriptions of images / video clips together to form a complete set of collections with different collective structures and a possible semantic definition. Authorities / unusual behavior is defined as non-existent in large data sets. [80].

It was stated that the merging algorithms can be categorized based on the following process [81]:

1) The most common data cases are in the data collection, and the inconsistencies do not belong to another collection or to the data collection.

2) Normal data conditions are found near the cluster centroid, whereas abnormalities are found far from the cluster centroid.

3) Normal data conditions are assigned to larger clusters, and unusual data is assigned to smaller clusters.

The remainder of this Segment provides an overview of other existing activities and acquisition strategies, both monitored and supervised.

2.17 Dynamic Time Warping

DTW is a powerful algorithm that measures similarity (distance) between two sequences, which is a type of template algorithm [82]. Waving, boxing, and clapping are all examples of human activities seen in DTW. Although the length of time between test sequences and reference sequences is inconsistent, DTW can still stop similarities as long as it meets time constraints. It has the advantage of easy simplicity and robust performance, and has recently been used to match human movement patterns. However, it may require a variety of templates for different scenarios, which has led to higher calculation costs to match these broader templates.

2.18 Finite-State Machine (FSM)

A state-transition function is the most important feature. Provinces are used to determine which sequence of references corresponds to what sequence. Higher level descriptions have also used state machine representations of behaviors [83].

2.19 Hidden Markov Model

There are three terms that define HMM. The first term is the first chance of a hidden state. The transformation matrix, which determines the chances of transition from one hidden location to another, is a second term [84]. The viewing matrix, which specifies the probability of a visual signal given a hidden state, is the third word. The same stage includes the chances that a particular HMM will produce a test signal sequence that matches the visual image features.

2.20 Artificial Neural Networks

Artificial neural networks (ANNs) are computer systems inspired by natural organic networks in the brain. It is made up of simple computer units known as highly connected neurons [85]. Synapses found in dendrites or membranes of natural neurons receive signals. while the signals acquired are inadequate (exceeding a positive restriction), the neuron activates and releases the signal with an axon. This signal may be sent to another synapse, activating other neurons. ANNs are made up of a series of units and are therefore called multiple ANNs. They mainly contain inputs in the first layer extended by weights and calculated mathematical function that determines the function of a neuron. Other work involves the release of an artificial neuron in the final layer. In the ANN structure, all other layers are hidden layers. To process information, ANNs incorporate artificial neurons. As the weight of an artificial neuron increases, the input power increases significantly.

2.21 Some Recent Work and Applications

So far, this paper has discussed a comprehensive review of the Intelligent Video Surveillance System framework. This Section outlines some of the strategies, methods, and applications that have been proposed, tested, and implemented in this field, commonly known as Surveillance, Health and Sport.

2.21.1 Surveillance

The frequency, complexity, and complexity of security breaches and attacks are growing rapidly. Many businesses struggle with security, especially if they have a restaurant or a store. Robbery, theft, shoplifting, and terrorist attacks are on the rise worldwide. As a result, businesses and governments are increasingly using video surveillance cameras. With the advancement of technology and the lack of human capabilities in video surveillance, various smart video rental applications have emerged in the field of security, such as roaming, crowd management, face detection, posture and calibration, and other crime-related activities. Surveillance Camera is presented in figure 2.3. Real Time Person Tracking using surveillance system is presented in figure 2.4 [86].

2.21.2. Crowd Count and Behavior Detection

Proper human identification and conduct can enhance the effectiveness of monitoring and managing public and private transportation and circles by marking and identifying areas of high traffic or signature areas that need more attention than others. One of the key strategies associated with this approach was to involve the operator in adjusting the sensitivity of the uncontrolled detector to the level of threat. Because the operator did not need to be an expert on the algorithm, the modification would be similar to identifying tracks as normal or abnormal, or by raising physical- based features that could be sensitive to threatening behavior.

Figure 2.3: Surveillance camera

Figure 2.4: Real time person tracking using surveillance system.

2.21.3 Face Detection

Identification of a person is a critical aspect of observation and investigation. Various techniques and symbols (e.g., gait, posture, dress) are used to identify a person, but the face has always been the most reliable. The neural network-based detector is trained using three distribution networks with multiple transmission lines that receive inputs from three different face presentations [87]. The total weight of the outputs of the three networks should provide a reliable solution for the presence of facial patterns. Face detection image is presented in figure 2.5.

Figure 2.5: Face detection

2.21.4. Human Pose and Gait Estimation

The measurement of a person's pose in a transport monitoring system refers to the shape of the whole human body, rather than the position associated with a single organ, such as a pose, which can be used for a variety of purposes. While gait means the style of human movement [51]. Both of these factors have aroused the interest of investigators interested in Personal Identification and Employment Awareness.

2.21.5. Loitering

Loitering is defined as the presence of a person or persons in a place longer than usual (fixed time). Although hard work to detect migration has yet to be done, some researchers have offered some practical methods.

2.22 Previous Methodologies Used in HAR Model

Most HAR studies falls into one of two categories: vision-based or auditory-based. To create a tracking pose, a twisted map and descriptive maps are used to show the 3D modification of each bone combination, and the descriptive model used to map and then integrate with a possible model. The correlation between specified and standard Gaussian characters is used to measure the shape of people using a single depth camera. This approach incorporates kinematic bone embedding into Gaussian characters and creates tree-shaped models in quaternion-based rotation. To evaluate the effectiveness of the sample prediction, we made sample samples for the next day on each of the three indicators for each of the data described in the Methodology section. In addition, based on the specifics of the model and forecasting methods, we performed a day-to-day, one week in advance, and a two-week prediction for each sample of indicators. For each measure during the sample period, we look at a rolling window of 1000 lengths [88]. Because data modification causes the longest RV (m) to contain information about the previous 21 days, that rolling window contains 979 views. One-week and two-week forecasts are made up of one-day and ten-day forecasts, respectively. We use the same time travel pattern of 1000 lengths when we predict long-term volatility as we make one-day forecasts. As a result, we made a few forecasts a week in advance and a few more predictions for the coming week. However, unlike RGB-based methods, they are not affected by changes in light. The authors have suggested the use of multiple accelerometers, wearable sensors, and other types of sensors in sensory studies. Explicit data sequences were collected using smart phones while trial participants performed their daily tasks [89]. The time, frequency, and frequency of a wavelet unit of a function unit are then used to summarize it. Custom and customized models are made and job recognition is done using a variety of classification techniques. IoT HAR systems are also recommended. The aim was to use data mining techniques to provide an increasing view of human activities. Because some actions are strongly associated with certain objects and standing conditions, a consistent look can be a useful solution in itself. Long short-term memory (LSTM) networks have recently been used in a variety of ways. Two features of bidirectional short-term memory (BLSTM) to add learning to HAR systems, were enhanced by the success of the survival feature. With improved accuracy, a new method called multicolumn BLSTM (MBLSTM), which tends to incorporate several features of the accelerated signal for human activity recognition. The standard test images for image encryption are presented in figure 2.6.

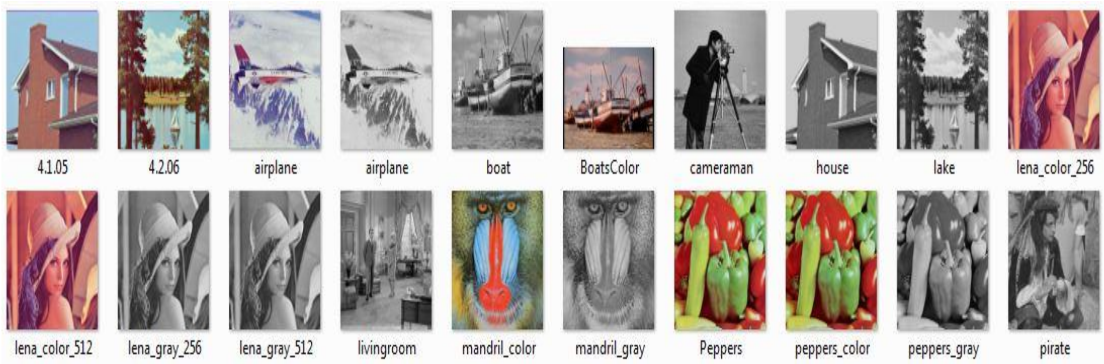


Figure 2.6: Standard test images

2.23 Comparison of Competitive Image Encryption Techniques

Competitive IET is analyzed under different techniques and comparison table is presented in Table 2.1

Table 2.1: Comparison of competitive image encryption techniques

Technique	TD	HPT	PI	CS	PC	SLO
Double random-phase encoding	✓	×	×	G	-	-
Decimal based logistic map	×	×	×	G	-	-
Mixed chaotic maps	×	×	×	G	-	-
Fractal dictionary	×	×	×	G	-	-
DNA and chaotic system	×	×	×	A	-	-
3D chaotic system	×	×	×	G	-	-
Fractio-l chaotic systems	✓	×	×	G	-	-
Cosine-transform-based chaotic system	✓	×	×	G	-	-
Hyper chaotic system	×	×	×	G	-	-
Latin square based block permutation	×	×	×	G	-	-
DNA and chaotic system	×	×	×	G	-	-
Differential evolution	✓	✓	×	G	✓	✓
Genetic algorithm	×	✓	×	A	✓	✓

2.24 Chaotic Image Encryption

Chaotic maps are a possible alternative to traditional cryptosystems because to their deterministic nature, sensitivity to initial values, ergodicity, and unexpected behavior. Chaotic maps are directly linked to cryptographic properties such as confusion and dispersion. As a consequence, chaotic maps have become a common way to encrypt images. Two of the most frequent varieties are one-dimensional and multidimensional. One-dimensional (1-D) systems are simple and efficient, making them straightforward to utilize. On the other hand, they feature a smaller key storage space and a lesser degree of security. Multidimensional systems are characterized by their complexity. These, on the other hand, give more critical space as well as protection from a broader variety of dangers [90].

The two logistic maps and an external key to encrypt a photograph. The external key establishes the prerequisites for logistic mapping. The secret key is renewed after encrypting a sixteen-pixel block of an image. As a consequence, discovering the secret key will be difficult for an attacker. On the other hand, the input photographs have no bearing on it . This employed a (1-D) chaotic map coupled to a map lattice to encrypt a photograph. You get a lot of key storage and high-security protection with this combination. Although less sensitive to the input image, this approach is nonetheless helpful [91].

The hyper-chaotic maps are used to reduce the prediction time of photo encryption techniques that rely on basic chaotic maps [23, 24]. The pixels in the input picture are scrambled using a matrix shuffling technique in this approach. A hyper- chaotic map is then utilized to dilute the pixel values of the shuffled image. This solution provides much more key space and security [89].

The logistic map is used to encrypt the photographs. In this approach, permutation is done at the bit level. This approach is resistant to a variety of security risks. However, known-plaintext and chosen-plaintext assaults are possible . The proposed permutation and diffusion at the bit level as an image encryption approach. Arnold transformations and logistic maps are utilized to carry out permutation and diffusion operations. When it comes to photo encryption, our solution exceeds the competitors in terms of processing power. Images encoded using permutation-based cyphers are subject to attacks that employ plaintext and ciphertext that already exists [91]. An image is divided into four pieces using this procedure. The blocks are then jumbled using a chaotic map. Finally, the pixel values of each block are changed at the same time to generate an encrypted block. This approach has a rather quick computation speed [92].

This employed a three-dimensional (3-D) Arnold transform to encrypt the pictures. The three essential components of this method are the shuffling, scrambling, and masking steps. The quality of the image used as an input has a big impact on this procedure. It is more resistant to plaintext assaults that are either previously known or explicitly chosen [93]. An image encryption method that selectively encrypts pictures using spatial chaotic maps. Each pixel's initial four bits are encrypted. According to the research, encrypting just half of an image's pixels resulted in better encryption results [94].

Chapter 3

Methodologies

3.1 Proposed Image Encryption Technique

An image encryption technique for secure surveillance frameworks for IoT systems is shown in the figure 3.1. The obtained images from surveillance cameras are initially encrypted by the proposed technique then transmitted over the network. The obtained image is firstly decomposed into red (R), green (G), and blue (B) channels. $B^{'}$. $G^{'}$ and $B^{'}$ are further diffused using $\oplus$ operation using δ_R and produce δ_G and $B^{''}$. Two more keys $d^{'}6$ and $d^{'}7$ are120 generated. $d^{'}6$ is generated from $d^{'}1$ and $d^{'}2$ using $\oplus$ operation. $d^{'}7$ is generated from $d^{'}3$ and $d^{'}4$ using $\oplus$ operation. $d^{'}4$ is then utilized to update $d^{'}5$ and $d^{'}6$. Encrypted color channels i.e., ϵ_R, ϵ_G, and ϵ_B are obtained by diffusing δ_R, δ_G, and $B^{''}$ with $d^{'}5^{'}$, $d^{'}6^{'}$, and $d^{'}7$ and η operator, respectively. Finally, the encrypted image (*EM)* is obtained by concatenating the ϵ_R, ϵ_G, and ϵ_B.

Partial-Regeneration Based Non-Dominated Optimization (PRNDO)

Partial-regeneration based non-dominated optimization (PRNDO) has been used because it will provide optimal solutions considering both best and worst cases. This type of optimization is mainly used to solve the asymmetrical problems. In this, bottom-boosting methodology is utilized to reduce the number of functions evaluations. It also provides a balance between better solutions and computational speed by developing a novel partial regeneration strategy and mutation operator. It contains population initialization, min heap construction, bottom-boosting, termination condition and partial regeneration strategy. Firstly, a population is generated randomly with the pair of scenario and solution. The other parameters are also initialized such as control attributes, maximum number of generations, solution and scenario dimensions. The working of PRNDO has been shown in figure 3.2.

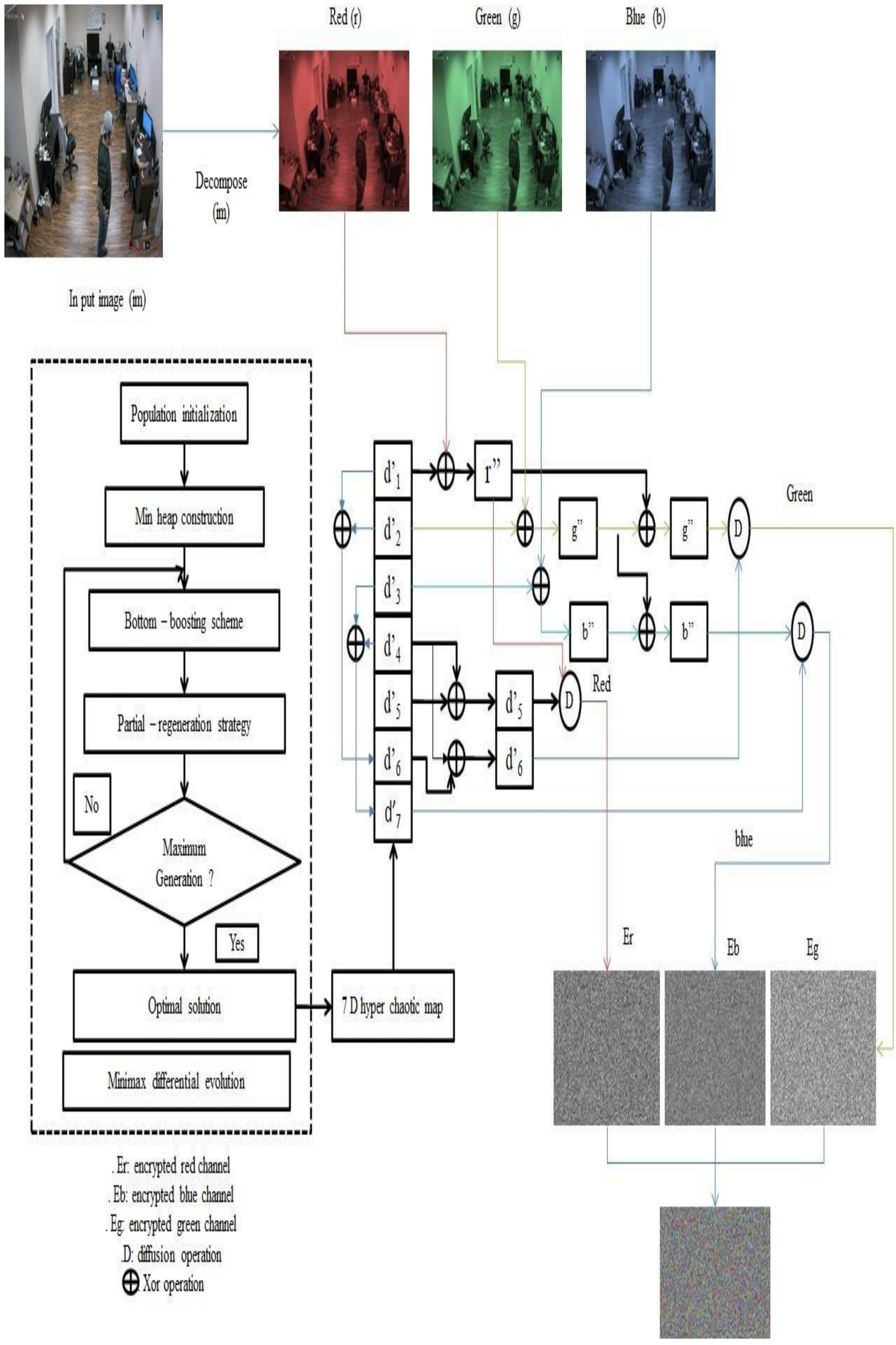


Figure 3.1: Flowchart of proposed image encryption approach

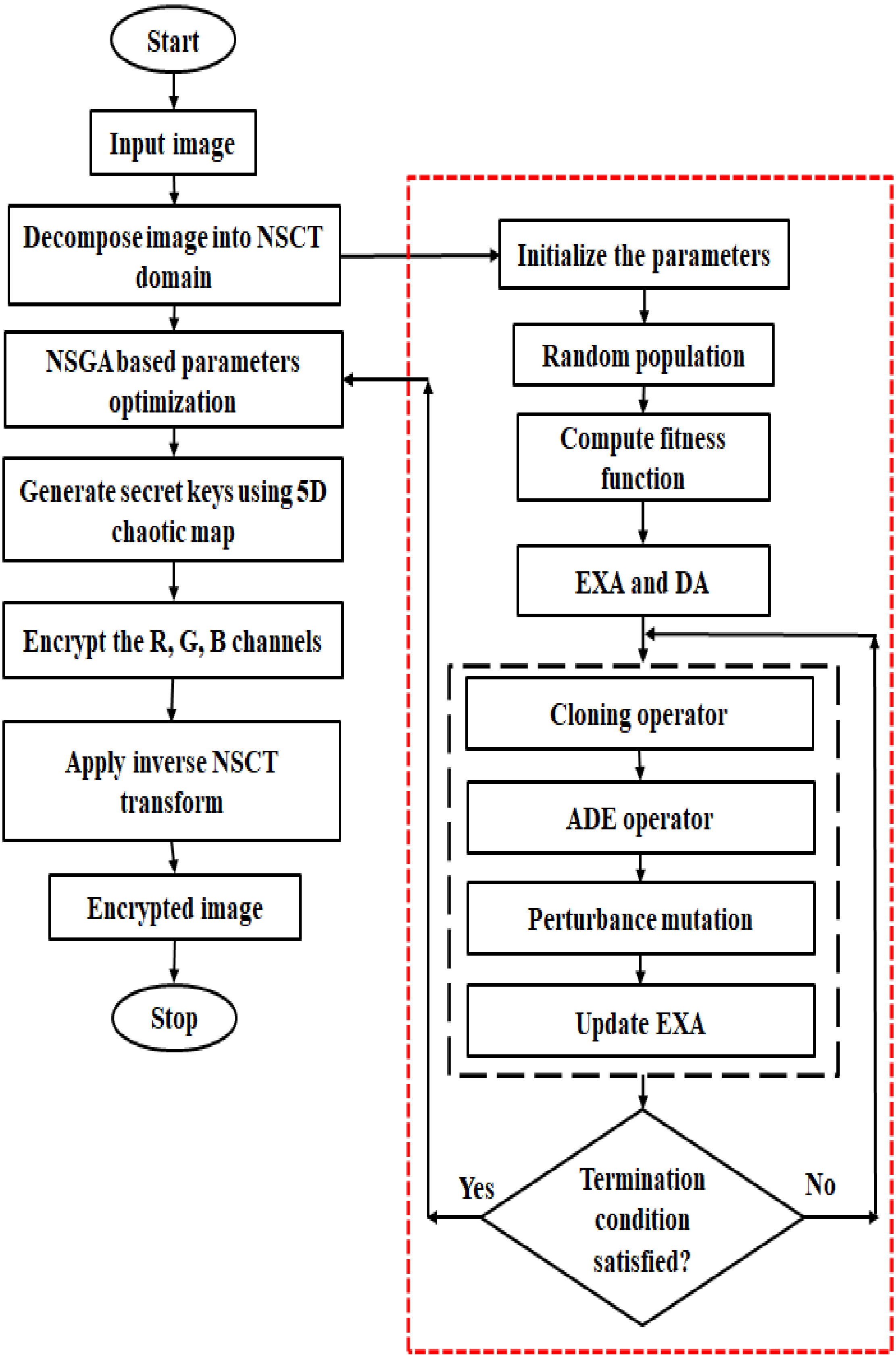


Figure 3.2: Proposed methodology flowchart

"Algorithm 1: Image encryption technique for surveillance systems

Input: Input image *im*

Output: Encrypted image *EM*

/ Divide im into red (R), green (G), and blue (B) channels. */*

$R = im(:, :, 1);$

$G = im(:, :, 2);$

$B = im(:, :, 3);$

Permute R, G, and B using Arnold transform.

obtain optimal parameters ($z_k(k = 1 : Sol_{\hat{D}})$) by using **Algo.** 2; generate secret keys $d'u$, where $u = 1, . . ., 7$ using **Algo. 5**;

// Diffuse the color channels using $d'1$, $d'2$, and $d'3$ $\delta_R=$

$\mathrm{mod}\,(R \oplus d'_1, 256);$

$G' = \mathrm{mod}(G \oplus d'_2, 256); \delta_G = \mathrm{mod}(G' \oplus R'', 256); B' =$

$\mathrm{mod}(B \oplus d'_3, 256); B'' = \mathrm{mod}(B' \oplus G', 256);$

//update the d_5' and d_6' using d_4'

$d_5'' = d'5 \oplus d'4;$

$d_6'' = d'6 \oplus d'4;$

//Encrypt the R'', G'', and B'' using $d'5'$, d_6'', and d_7' $\eta =$

u_{lM} ;

$\epsilon_R = \mathrm{mod}(d_5'' \times R'' + (1 - \eta) \times d_5'', 256);$

$\epsilon_G = \mathrm{mod}(d_6" \times G" + (1-\eta) \times d_6", 256)$;

$\epsilon_B = \mathrm{mod}(d_7' \times B" + (1-\eta) \times d_7', 256)$;

//Final encrypted image is generated by combining the encrypted channels

$EM = \mathrm{cat}(\epsilon_R, \epsilon_G, \epsilon_B)$; return

EM ;"

Algorithm 2: PRNDO based optimal parameters generation

Input: Solution dimension Sol_o, scenario dimension S_o, maximum generation N_h, and population size M_l

Output: Optimized population: Set generation $u = 0$;

//population initialization

$P_a = \{(z_{1,u}, I_{1,u}), (z_{2,u}, I_{2,u}), \ldots, (z_{M,u}, I_{M,u})\}$,

where $z_{k,u} = \{z^1, S^2, \ldots, S^{Sol}d\}$

for $y = 1$ *to* M_s **do**

fitness of individual ($z_{k,u}$, $I_{k,u}$) is evaluated using Algorithm **6**; **end**

for $u = 1$ *to* N_h **do**

Min heap is built for each individual in q_u;

Use (Algorithm 3) to implement the bottom-boosting approach on min-heap;

for $k = 1$ *to* M_l **do**

//Individual (z_e, I_e) stored as a root node in the min-heap

zk,u = ze, Ik,u = Ie;

Extract root node from min-heap;

5 End

$zbest = z1, u$;

Use partial-regeneration strategy for D_u updating (Algorithm 4);

$D_{u+1} = D_u$;

End

In Algo. 2, firstly, the parameters solution dimension Sol_o, scenario dimension s_o, maximum generation N_h, and population size M_l are initialized. Using normal distribution, initial population D_u is generated randomly. Each individual ($z_{k,u}$, $I_{k,u}$) represent as a set of solution and scenario. Fitness of ($z_{k,u}$, $I_{k,u}$) is evaluated using Algo. 6. A min-heap is built using D_u. The scenarios are updated using Algo. 3. The current population is updated by extracting the root node of min-heap. Then, the nodes are arranged in increasing order on the basis of fitness. The initial solution of given population is utilized as global best solution. Thereafter, a partial-regeneration strategy is utilized to update the current population further (see Algo. 4). The entire steps are iteratively implemented till the stopping criteria not met.

Algorithm 3: Bottom boosting approach

Input: Scaling factor F, a min-heap constructed using D_u, crossover rate Ie, and number of fitness evaluations

H_w, **Output:** Min-heap $EE = 0$;

while $EE < H_w$ **do**

//Consider each ($z_{k,u}$, $I_{k,u}$) stored in the root, visit root of min-heap;

Mutant is developed as

$Iw_{k,u} = I_e1, u + H(I_e2 ,u–I_e3 , u)$;//where e_1, e_2, and e_3 are elected on a random basis from $\{1, . . . , M_l\}$

Trial scenario $IV_{k,u}$ is produced using binomial recombination on $I_{k,u}$ and $IU_{k,u}$;

if F ($z_{k,u}$, $IV_{k,u}$) > F ($z_{k,u}$, $I_{k,u}$) **then** extract the root node of min-heap; $Ik,u = IVk,u$;

Update min-heap using ($z_{k,u}$, $I_{k,u}$);

5 End

$EE = EE + 1$;

End

Algorithm 4: Partial-regeneration approach

Input: Sorted population D_u, scaling factor F, crossover rate Ie, and number of updated individuals β

Output: Updated population D_u $y = 1$;

while $y \leq \beta$ **do**

$X_{k,u} = z_{k,u} + H(Y_r1, a - z_l2, u)$;//Here, $k \in \{1, 2, \ldots, \beta\}$

Trail solution $\phi_{k,u}$ is produced using binomial recombination on $z_{k,u}$ and $X_{k,u}$; $zM_l + 1 - y,u = \phi\ k,u$;

Reinitialize $I_Ml + 1 - y,u$, randomly;

Evaluate the fitness of ($z_Ml + 1 - y,u$, $z_Ml + 1 - y,u$) using Algorithm 6;

$y = y + 1$;

End

Algorithm 5: Generation of secret keys using 5DHCM

Input: Optimized parameters $z_k(k = 1{:}\ SolD\hat{})$

Output: Secret keys $d'1$, $d'2$, $d'3$, $d'4$, $d'5$, $d'6$, and $d'7$

//In $z_k(k = 1{:}\ SolD\hat{} - 1)$, $SolD\hat{} = 18$.

//$z_k(k = 1{:}\ SolD\hat{} - 1)$ denotes d_1, d_2, d_3, d_4, d_5, q, r, s, t, u, v, and w, respectively.

$d'1 = z_6(z_2 - z_1) + z_2z_3z_4$;

$d'2 = z_7(z_1 + z_2) + z_5 - z_1z_3z_4$;

$d'3 = -z_8z_2 - z_9z_3 - z_{10}z_4 + z_1z_2z_4$; $d'4 = -z_{11}z_4 + z_2z_3z_1$;

$d'5 = -z_{12}(z_1 + z_2)$; $d'6 = z_1 \oplus z_2$;

$d'7 = z_3 \oplus z_4$;

return $d'1$, $d'2$, $d'3$, $d'4$, $d'5$, $d'6$, and $d'7$

Algorithm 6: Fitness function

Input: solution $z_k(k = 1{:}\ SolD\hat{}\ )$, input image, *im*

Output: Fitness value, F

Encrypted image, *EM,* is obtained using Algorithm 1

II = correlation (EM);

ϵ_x = entropy (EM);

if $\epsilon_x > 7.9990\ \&\& -0.05 \le CC \le 0.05$ **then**

End

return $F(\phi)$

3.2 Human Activity Recognition (HAR) Model Using Deep Learning

It includes foreseeing human development dependent on tangible information and customarily fuses profound area aptitude and sign preparing strategies to enhance the highlights of crude information to coordinate with the AI model. Inside and out learning strategies, for example, convolutional neural organizations and copy neural organizations, have as of late shown strength and surprisingly accomplished excellent outcomes by taking in highlights from crude sensor information. Sensor information, like video, radar, or other remote modes, can be recorded distantly. Information can likewise be gathered straightforwardly regarding the matter, like dealing with custom hardware or smart phones equipped with accelerometers and gyroscopes. in their sensory data. The main motivation for this study is to create a low-cost, efficient HAR system [94]. The flowchart of the proposed HAR model is shown in figure 3.3.

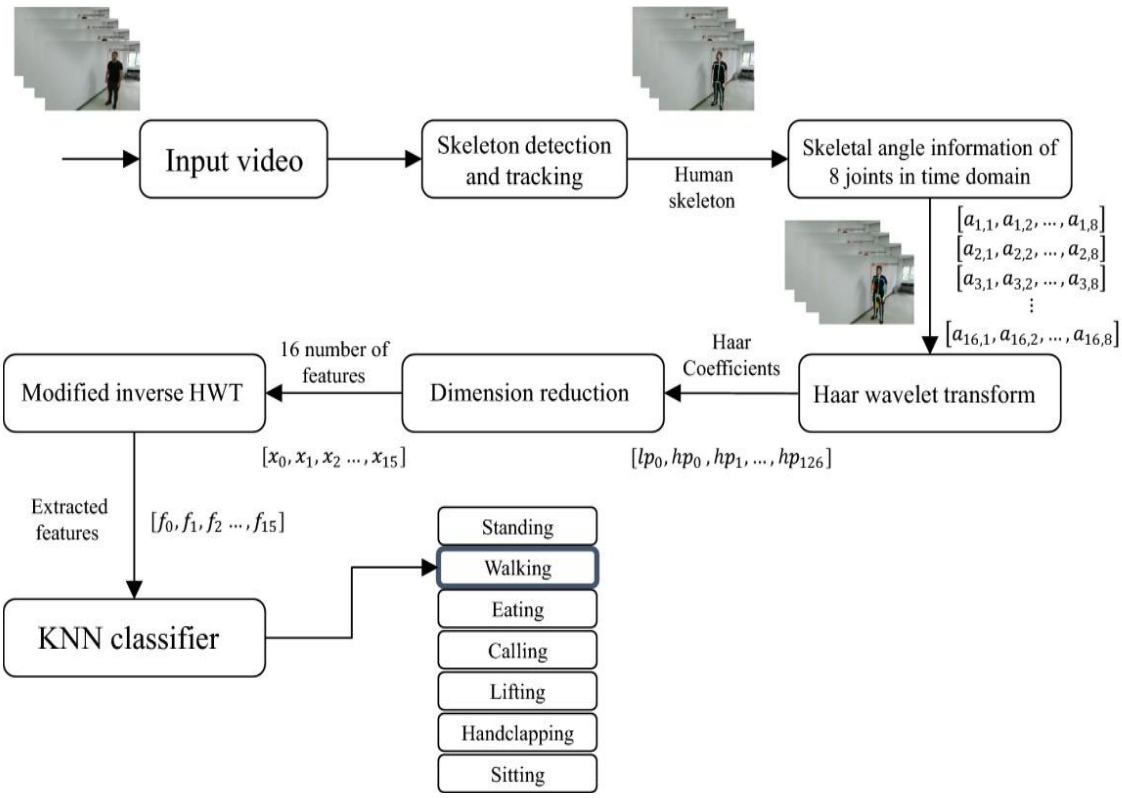


Figure 3.3: Flowchart of the proposed model HAR

3.3 Intelligent Video Surveillance System (IVSS) Framework

Intelligent Video Surveillance System (IVSS) improves the performance of surveillance systems using automated video analytics. The program maintains a home security environment, which reduces burglary rates and improves social stability. Typically, the framework of Intelligent Video Surveillance System (IVSS processing) includes Motion/object detection, object tracking, Personal Identification, Performance Analysis, and Behavioral Understanding of the categories concerned. The ability to find people and understand how they behave spontaneously is an important function of this study. Figure 3.4 shows the proposed block diagram of the proposed Advance IVSS system in this sense.

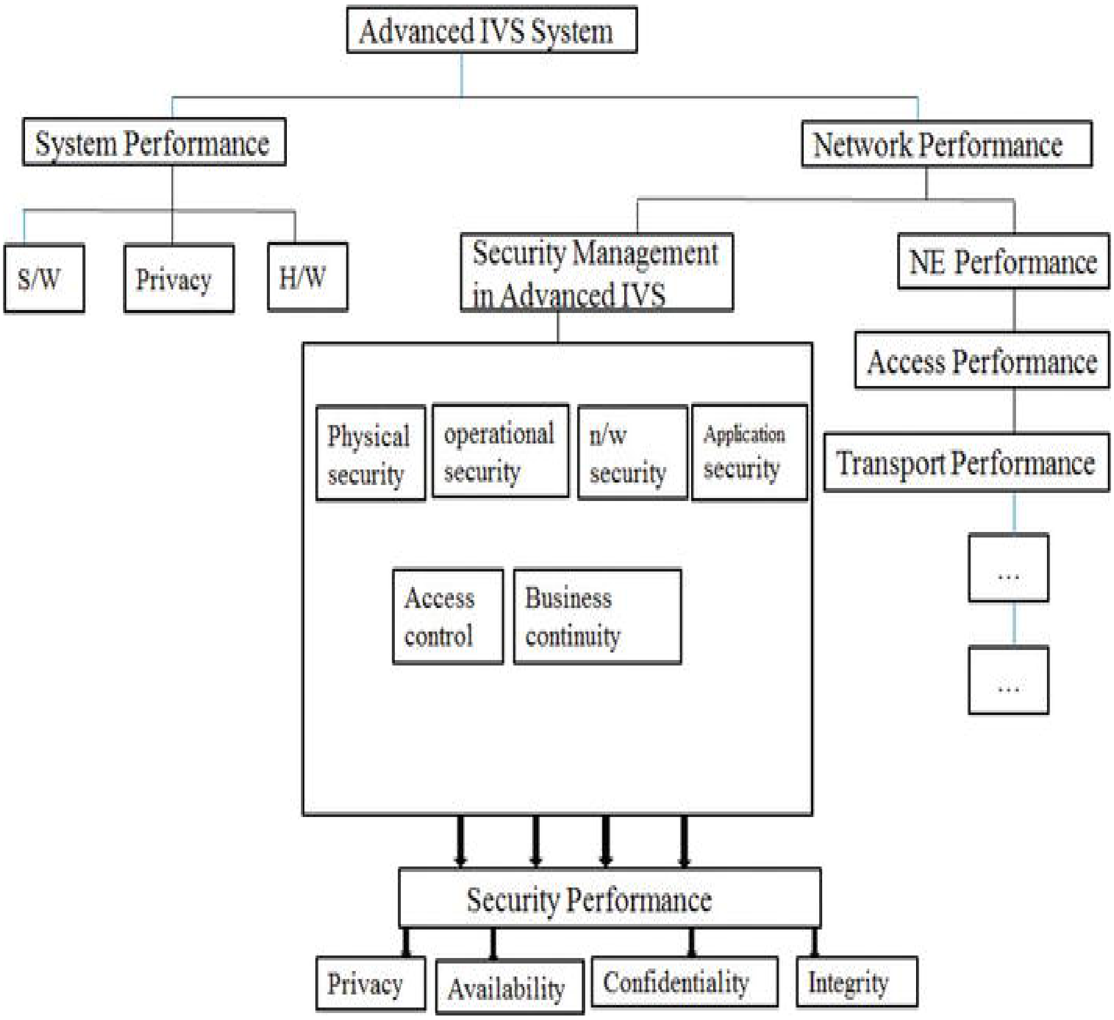


Figure 3.4: Block diagram of advance intelligent video surveillance system (IVSS) framework

Surveillance cameras have been installed in all commercial and public areas to improve security in terrorist activities. In this case, Automatic Person Identification (API) from suspicious images plays a very important role in security programs. Identify a person from a large number of videos without human help. In addition to identifying a person, communication between a person and his neighbor is necessary in job analysis. HAR, joint angles obtained using an RGB-depth sensor are used as features [95].

In addition, many studies have attempted to identify practices and strategies for collaborative recognition, but none of these studies have focused on a comprehensive participatory identification system that includes a few factors.

3.4. Types of Human Interaction Recognition

3.4.1. Simple Interaction Recognition

Person to Person Interaction Recognition (P2P-IR), in addition to person identification, is critical in development of smart video monitoring systems. Human behavior is discussed in the context of computer vision considered as a series of actions carried out by the people in an image sequence Recognize interactions, identify body parts, locations, and complex movement with good detection. These challenges prompted us to propose a simple P2P interaction recognition technique that uses improved spatio-temporal feature extraction techniques and structured learning techniques to recognize the interaction more effectively. Global features are extracted from the entire frame of the video, whereas local features are removed from the object/person in the video. Interaction modelling was carried out by deriving semantics from the detection window based on spatial relationships such as overlap, above, below, next to (right), and next to (left). (On the left) These semantics have been used in peer-to-peer interactions. When making predictions, the relationships between body parts were taken into account. These dependencies were incorporated as semantics using the distance function to obtain simple interactions such as handshakes, hugging, kicking, and pushing [96].

3.4.2. Concurrent Interaction Recognition

Concurrent interaction refers to interactions between multiple people in a complex environment. In this case, the interactions between each pair of people are identified. For example, in the scene, there are four people: the first two are interacting with each other, and the other two are interacting separately. Two interactions in the video may occur concurrently or simultaneously at any given point in time. Hug and Handshake, for example, are concurrent interactions that can occur at the same time, i.e., the first two people hugging each other and the other two people handshaking each other. To identify the interaction, the actions of the two people were modelled independently. Body part modelling and location-based trajectory features are extracted to help with people localization in frames. Based on the learning, the system finds the corresponding sequence of poses by observing the sequence from the feature vector. The pose sequence produced is the one that achieves the best balance of chances by standing. As a result, the visual pose is inserted into the HMM of the next layer. The deep learning has achieved great success in visual tracking tasks, particularly in single-object tracking [96]. The program tries to find the corresponding action sequence in the second layer of HMM. The maximum probability of a sequence of action is used to identify each action1.

3.4.3. Recursive Activity Recognition

Human behavior has been recognized in terms of activity, in addition to concurrent interactions. The major challenge is the long-term and continuous modelling of activity and interaction. A High-Level Semantic Descriptor (HLSD) is proposed in this work. HLSD is derived from motion attributes because the motion of each pixel in the video can provide more semantics. Interaction is identified using a low-level action descriptor and a high-level semantic descriptor. For each interaction pattern, the semantic motion pattern has been identified. Whenever the interaction is repeated for a specific time, the motion pattern remains the same. Initially, the person in the video was identified, and an action descriptor was derived using visual words. The video is used to extract the global and local motion fields. The semantic terms are the motion attributes' location, velocity, and time sequence. The semantic gap was filled with a high-level semantic descriptor, and the descriptor output was fed directly into the Conditional Random Field (CRF). Recent changes are movement symptoms such as arm extension, arm withdrawal, leg extension, leg withdrawal, and body contact. To recognize the interactions, the latent variables are introduced into CRF, and the descriptor is learned using a fully integrated discriminative model. The semantic motion pattern similarity was calculated using the motion pattern's peak [97].

3.5 Stream Analysis of the Proposed Algorithm

3.5.1. Points Between Joints

The Kinect v2 sensor can recognize and follow 25 bone joints [98]. An important part of this method is to select the elements that will form the vectors of the element after the bone has been detected and tracked. Gait recognition algorithms are broadly classified into two types: (1) model-based and (2) holistic approaches1. Model-based approaches extract parameters by analyzing gait sequences. Because parameters are identified through various gait sequences, model-based approaches generally necessitate a rich quality gait sequence. Holistic solutions operate directly on gait sequences, making no assumptions about the walking human. These cameras were capable of recording RGB videos, and posture classification was accomplished through the use of various image processing techniques. Following the introduction of the Microsoft Kinect, a new revolution was born. Microsoft Kinect is a peripheral device that, like a webcam, can be connected to a computer.

In addition to providing an RGB image, it also measures the distance of each pixel from the camera, making background removal, blob extraction, and other tasks easier. Posture recognition, fall detection, face recognition, and other applications The IoT is gaining clout. IoT and fog computing were integrated with basic recognition algorithms 15-17 in a number of research studies [98]. An IoT-Based Fall Detection System with Energy Efficient Sensor Nodes was proposed in an exploration method 15. They have incorporated the idea of mist figuring just as the web of things in this strategy. Kinect Camera records the profundity and shade of pictures at 30 frames each second (FPS), making a three-dimensional (3D) cloud from a telephone design showed at the scene. The right and left knees, just as the elbow, are the most informative provisions of human bones when standing and strolling. Furthermore, the point sets of hipbones are additionally remembered for the proposed approach [99]. The picture shows the point sets utilized in this examination.

The point between two unique individuals j1 and j2 can be determined by isolating the places of the two individuals in the 3D space in relation to the combined reference r. The following formula:

$$aj_1, j_2 = cos^{-1}\left(\frac{\overrightarrow{rj_1} \cdot \overrightarrow{rj_2}}{||rj_1|| \cdot ||rj_2}\right) \quad 3.1$$

In this calculation, $\vec{rj}1$ it represents the distance in the 3D space between the joint $j1$ and the reference joint r. Similarly, $\vec{rj}2$ it represents the distance in the 3D space between the joint $j2$ and the reference joint r. The product of the dots is indicated by a dot between the vectors. Finally, $||rj1||$ and $||rj2||$ represent the lengths of these vectors.

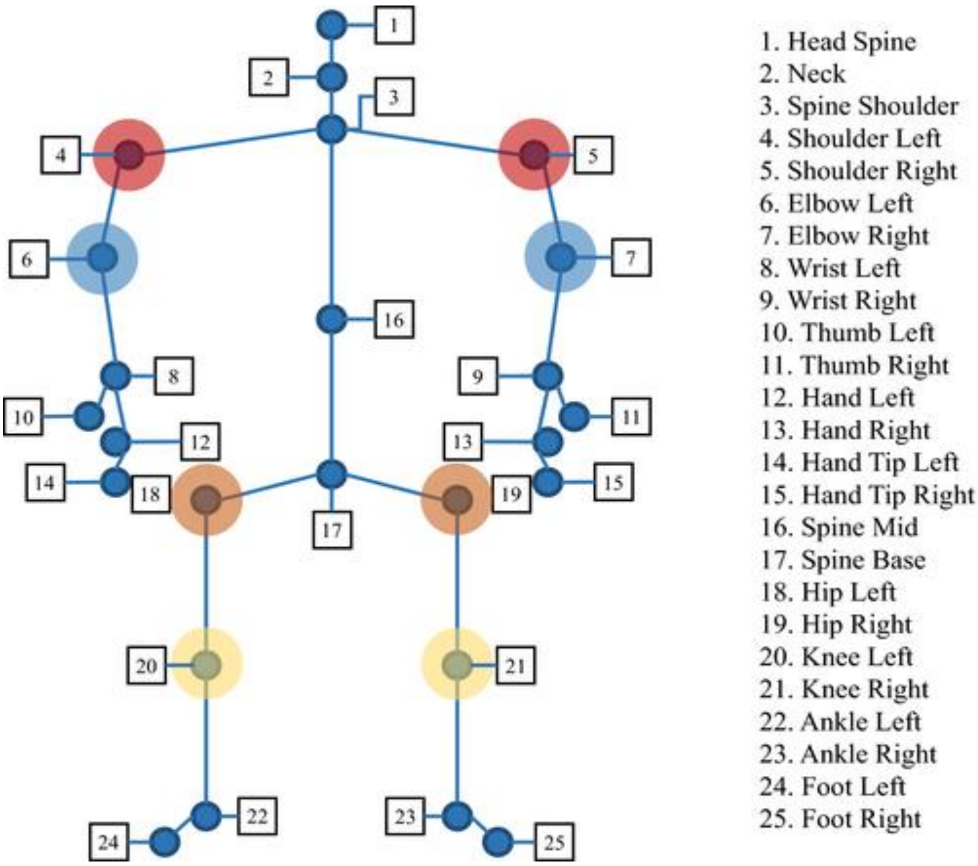


Figure 3.5: Human skeleton and selected angle pairs

3.6 Mechanism of Sliding Kernel

Since human movement happens in the domain of time, it very well may be considered as a two-dimensional issue. To see the activity, it is important to take a gander at the point examples of each point associated with the time region. For this reason, keeping angle swings on a time zone in the form of a loading kernel is advantageous. In the time domain, angle data for each joint can be denoted as follows:

$$k = \begin{pmatrix} a_{1,1} & \cdots & a_{1,8} \\ \vdots & \ddots & \vdots \\ a_{16,1} & \cdots & a_{16,8} \end{pmatrix} \rightarrow k' = [a_{1,1}, a_{1,2}, \ldots, a_{16,8}] \qquad 3.2$$

$$[x_0, x_1, \ldots, x_{127}]$$

The sliding kernel is implemented in this paper using a queue structure. The Kinect camera obtains eight joint angles in each frame and stores them in the queue structure. The Kinect camera detects eight combined angles in each frame and keeps them in line form. The line structure capacity is another program parameter that shows the kernel size in this case. If the queue is equal to 16, the queue retains 16 * 8 = 128 features. Because queue constructing is the first of the first data (FIFO), it behaves in a timeline like the sliding kernel [91]. In our study, the size of the kernel (or line) is set to 16, which produces the best results depending on the difficulty of time and performance.

3.7 Discrete Haar Wavelet Transform (DWT) Algorithm

A discrete wavelet transform (DWT) is a variation that separates a given signal into multiple sets, each of which is a series of time coefficients that define the occurrence of a signal in the corresponding frequency band. The WAVCNS technology is now built into the Neuro SENSE Monitors. The WAVCNS, like the BIS, uses a 0–100 scale to represent the patient's level of hypnosis. WAVCNS technology was developed with special care to avoid nonlinearities and the use of non-minimum phase elements. WAVCNS technology adds no additional computational delays, making it virtually delay free. Haar wavelets are square shaped functions that were created to organize data based on frequency. Wavelet conversion converts data from a local domain to a common domain, and each component on the corresponding resolution scale is saved. The following relationships define the basic Haar function $\psi(t)$ with scaling function $\phi(t)$:

$$\psi(t) = \begin{cases} 1\ t \in [0,0.5), \\ -1\ t \in [0.5,1), \\ 0\ otherwise, \end{cases}$$

$$\psi_1^j\,(t) = \sqrt{2^j \psi(2^j\,t - i)},$$

$$j = 0,1,\ldots,\infty\ and\ i = 0,1,\ldots,2^j - 1,$$

$$\phi(t) = \begin{cases} 1\ t\,\epsilon\,[0,0.5), \\ -1\ t\,\epsilon\,[0.5,1), \\ 0\ otherwise, \end{cases} \qquad 3.3$$

The functional relationship among the wavelet function and the scaling function is as trails:

$$\phi(\mathrm{t}) = \phi(\mathrm{t}) + \phi(2\mathrm{t} - 1) \qquad 3.4$$

and

$$(t) = \psi(t) - \psi(2t-1).$$

The Haar wavelet version consists mainly of measurement and variation. HWT simply takes a two-dimensional combination of data and places it in the first part of the series (lower band) of the input data it has 2^n elements. The other part of the new series (high pass band) is made up of a common difference between the two. This process is repeated until there is 2^{n-1} details of the coefficients and the minimum amount of lowpass values. Assume x is the vector of length N, $x = (x_1, x_2, \ldots, x_N)$ and that N is a power of two. The number of steps required is as follows:

$$\rho = \log_2 N. \qquad 3.5$$

The following is the procedure for calculating lowpass lp_k and high pass band coefficients hp_k :

$$lp_{k = \frac{(x_2k + x_2k+1)}{2} for\ k = 0,1,\ldots,\frac{N}{2}-1} \qquad 3.6$$

and

$$hp_{k = \frac{(x_2k + x_2k+1)}{2} for\ k = 0,1,\ldots,\frac{N}{2}-1} \qquad 3.7$$

This procedure is also reconstructive because:

$$\frac{(x_{2k} + x_{2k}+1)}{2} + \frac{(x_{2k} - x_{2k}+1}{2} = x_{2k} \qquad 3.8$$

and

$$\frac{(x_{2k} + x_{2k}+1)}{2} + \frac{(x_{2k} - x_{2k}+1)}{2} = x_{2k-1} \qquad 3.9$$

The total number of items for each input signal is 128, which means that the step number is $\log_2 128 = 7$. Reducing the measurement of the back-based size is applied after obtaining the Haar coefficients of the factor vectors.

3.8 Dimension Reduction with Averaging

Following the making of 128 Haar coefficients, a technique-based decrease strategy is utilized to diminish computational expenses and kill destruction [100]. The proportion of two HWT contenders to the lowpass or high pass modules will be in accordance with the new vector components in the new set. For instance, the first and second things compare to the main marker of another vector, the third and fourth things relate to the second pointer of the new vector, etc. Coming up next is the condition for lessening the measurements:

$$f_n = \frac{(x_n,2k+x_n,2k+1}{2}, \text{for } k = 0, 1, \ldots, \frac{N}{2} - 1 \quad 3.10$$

and

$$n = 1, 2, \ldots, \rho, \quad 3.11$$

where f stands for the element vector in each subtraction. In addition, it represents the step number. On a regular basis, we use a scale to reduce the size. We use the measurement method after obtaining the Haar coefficients to reduce the size of the graph of the frequency of the Haar coefficients. We use the opposite converter to find feature coefficients in the time zone after the size reduction in the frequency domain. Environmental background sound is a rich information source for identifying individual and social behaviors [99].

3.9 Thresholding with Inverse HWT Methodology

As referenced in the Haar Wavelet Transform area, the first sign can be handily reproduced without losing data using repeated measurement and separation functionality. However, by using a non-contradictory limit, a higher level of pressure can be obtained λ. This is known as loss reduction. Using the limit causes any lower band objects in the signal to be converted to zero if their magnitude is not so great or equivalent as far as possible worth. Subsequently, the number of eggs in the changed over sign will increment, giving a more significant level of pressure.

The threshold value λ of missing pressure is set to 0. The original signal ratio is created when the lost pressure is required or beneficial. Setting a limit to λ it requires caution because there is a trade between the limit value and the quality of the pressed signal. The standard limit figure is

$$x \in [1, N-1] \quad 3.12$$

and

$$T(\lambda, x = \begin{cases} 0 \; if \; |x| < \lambda, \\ x \; otherwise \end{cases} \quad 3.13$$

where x means any Haar accomplice in the high pass band. The proposed impeding technique is a valuable method to work on signal quality during the inverse HWT. This is a make-up

$$x \in [1, N-1] \quad 3.14$$

$$\mu = (lp_0 + hp_0 + hp_1 + \ldots + hp_{N-2}) / N, \quad 3.15$$

$$T(\lambda, x) = \begin{cases} x \;\; \text{if} \, |x| > \mu, \\ \text{sign}(x) \, \mu \; \text{otherwise} \end{cases} \quad 3.16$$

where μ is the definition of Haar coefficients for both primary and secondary grades.

3.10 Feature Extraction

The exclusion algorithm sets the coded data, thereby decreasing the measure of information to be handled Analyzes preset data using techniques such as arbitration detection. Translation and mathematical functions are examples of transformations and mathematical operations. This enables data to be organized into standard formats. This interaction lessens crude information into a restricted and substantial arrangement of sign components, reducing the complexity of the recognition model. There can be a number of factors that can be detected by each available sensor, and if not properly implemented, the problem of segregation can be an extremely determined system. There are many options to lessen the quantity of components utilized in the segment to stay away from unhindered issues. To determine which features and mitigation measures are appropriate for the problem, it is important to first determine which data is available [100].

3.11 Types of Classifiers

There are a number of classification algorithms that can be used in the HAR system. The following are the common types of separators:

- Bayesian Methods: Bayesian methods calculate event opportunities using conditional opportunities based on a set of features and opportunities. The most common methods of Basaya are the Basesi network and the inexperienced Bay, which ultimately is a special case where all aspects are considered statistically independent [101].

- IT-based learning (IBL): Iter-based learning methods (IBL) learn similarities and relationships between features based on a training data set. These methods require a lot of calculation power, a lot of data, and a lot of storage space.

- Support Vector Machine (SVM): Support Vector Machine (SVM) is widely used in HAR systems [27]. Based on multiple kernel functions that convert all features to a higher level to get information and relationships between features and functions.

- Artificial Neural Networks (ANN or NN): Artificial neural networks (ANN or NN) use a weighted network to analyze the features and functions performed. It clearly identifies patterns and models by mimicking the neural synapses found in living things. In this project, it could be the tree of domination or the end of the Basesi method used for performance analysis, one algorithm for each type of partition was used on the mobile device, and the best was used in the system.

3.12 Rule Tree Classifier

The C4.5 algorithms is selected for the law tree categories. This algorithm uses a variety of rules based on a set of training data, which classifies data into different categories (benefits) based on the relationship of the feature to the function to be known (attribute). Decision branches are a feature of local factors, identified as binary variants.

3.13 Bayesian Classifier

The Naive Bayes algorithm was selected for Bayesian classification. For each set of features provided, this algorithm is based on the conditional possibilities of the work being done. It is assumed that all factors are statistically independent. The algorithm starts by calculating and identifying all available features (attributes) and known function (classes) [92]. The conditional opportunities of a given category are used to produce rules. This is done for all possible classes (Equation (3.17)):

$$P(C \mid x) = P(C)\, p(x \mid C)\, P(x). \quad 3.17$$

This translates into time signs for classes that repeat the number of rules in one problem. Following the calculation of all the rules, a conditional legal calculation of a given set of symbols and category is performed.

3.14 In-depth Learning using Human Activity Recognition

On the other hand, due to their high performance, in-depth learning algorithms have gained popularity in various domains in recent years. Because DL is based on the concept of data representation, such strategies can automatically create positive features from raw input data without human intervention, allowing the identification of previously unknown patterns to remain hidden [101]. However, as mentioned earlier, DL models have some limitations:

1. The translation of black box models is difficult and natural.

2. Large data sets are required for training.

3. Computer costs are high.

Because of these limitations, CML methods are still popular in some areas, especially when the training database is limited or when urgent training is required. Convolutional Neural Network (CNN), Recurrent Neural Networks (RNNs), Short-Term Memory Networks (LSTMs), Gated Recurrent Unit (GRU), Integrated Autoencoders, and other standard DL algorithms.

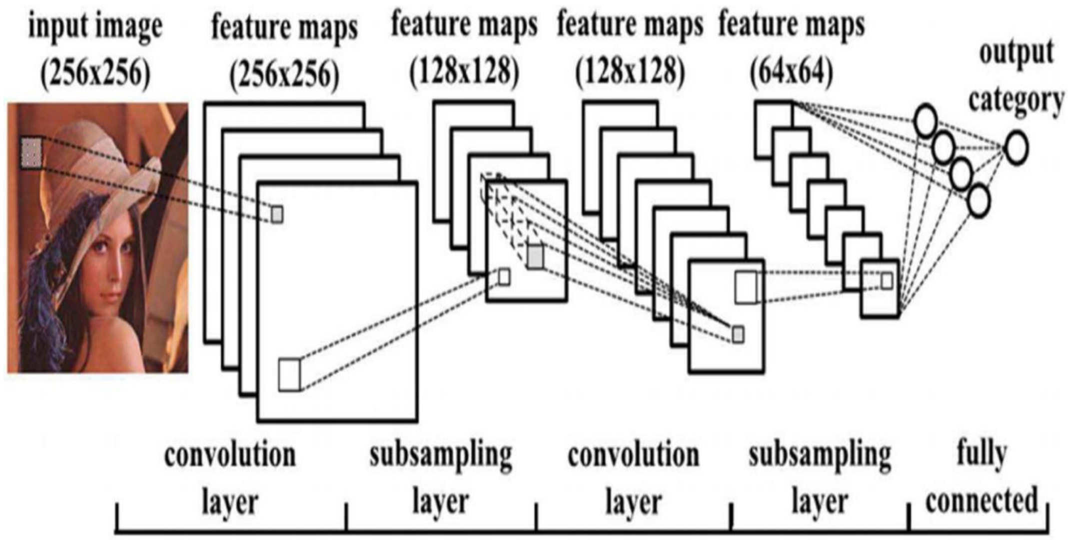


Figure 3.6: Example of convolutional neural network (CNN) for image classification

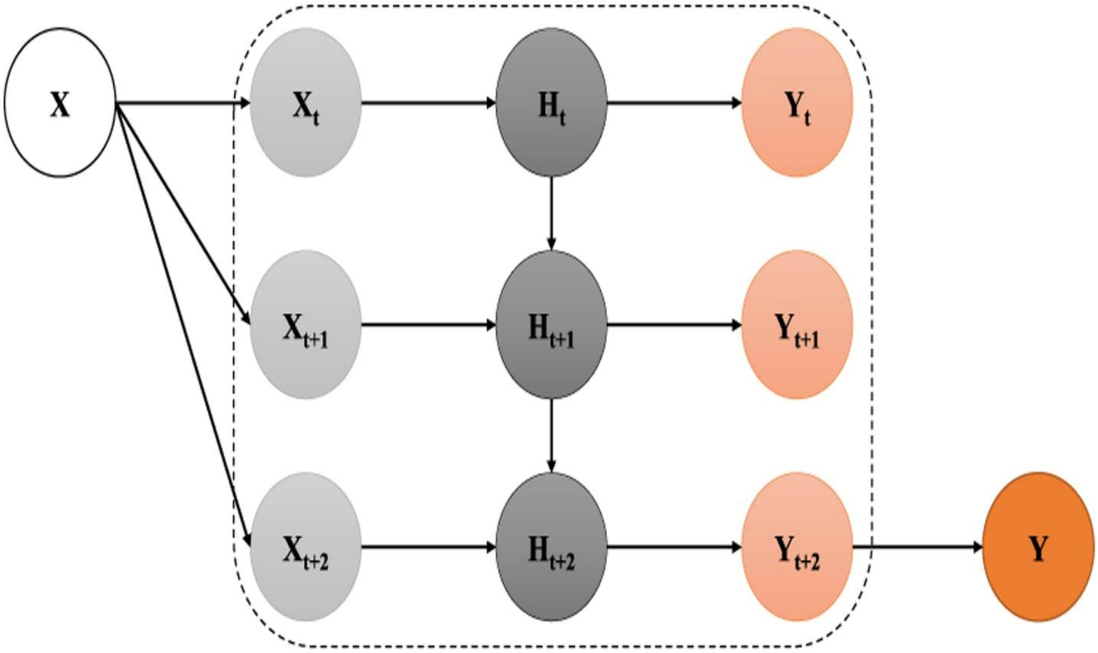


Figure 3.7: An example of an extended presentation of the recurrent neural network (RNN) with a three-length input system, three hidden units, and one output.

3.15 Sensors

The sensors and devices that we wear around us in our daily lives. Accelerometers are the most widely used sensors in monitoring activity, due to their small size and low cost. Accelerometers are often used in conjunction with other sensors such as gyroscopes, magnetometers and pressure sensors. Many other types of sensors have been used in a variety of applications. Other types of data collection and communication capabilities for sensor systems, in addition to Smartphones and smartwatches, offer IoT [102].

3.16 Data

The type of data collected refers to the second step of HAR workflow. These types of data can be broadly categorized as follows.

1. Data from weak sensors such as accelerometers, gyroscopes, magnetometers, and compass
2. Data from body sensors such as ECG, EMG, Heart Rate, or blood pressure
3. Data from environmental sensors such as temperature, pressure, CO2, moisture, and PIR.

3.17 Inertial Sensors Data

Sensors with up to nine degrees of freedom, such as accelerometers, gyroscopes, and magnetometers, are available for sale at very low prices. In addition, speed and angular velocity are the most widely used terms to describe human activities. This is supported by the previous section, because accelerometers and gyroscopes are widely used in HAR. HAR has become effective as a computer vision tool for video surveillance systems [103].

3.18 Physiological Sensors Data

Sensors of the body receive physical symptoms, which, unlike other sources of emotional information (facial expressions, gestures, and speech), provide critical benefits. Because these symbols are so unpopular, they cannot be compared to deception. They can be used to track relevant events in real time. The most widely used physical symptoms for brain electrical activity, heart rate, muscle tone, blood pressure, and skin reactions, are obtained using external data processing systems Electroencephalogram (EEG), Electrocardiogram (ECG), and Electromyography (EMG).

3.19 Environmental Sensors Data

Environmental data includes any data collection that represents environmental conditions, such as temperature, humidity, pressure, or light. The natural scale, on the other hand, exceeds the natural steps [98]. It can involve complex processes involving humans and the environment. Identifying the number of people in the environment and their positions, or actions taken from something in the environment, for example, can help in a variety of app situations related to relief services, health care, and service delivery. the difference between the titles and the middle, the topics of human mobility are very difficult to distinguish. Because of this diversity, getting used to all the subjects is very difficult.

3.20 System Description

The purpose of this type is to measure dynamic energy, store data in the cloud, and remotely display any variables stored in the cloud. Some alarms can be adjusted at different levels depending on the needs of the hospital. The system block diagram is made up of five elements (sensor data reception, visualization, HAR, Gateway, and cloud segmentation) [104]. Figure 3.8 shows this diagram:

- Sensor Data Receipt (SDR): This feature allows Bio harness 3 to receive Bluetooth data.
- HAR Divider: We use SDR data, a separation algorithm, and a machine learning process to inform the user of the type of movement currently occurring.
- Location view: You are in charge of the variable time displays found in the SDR on the graph.
- Gateway: Each dynamic routes detected from SDR to a cloud server.
- Cloud: Find and save all information posted by Gateway for later discussion online.

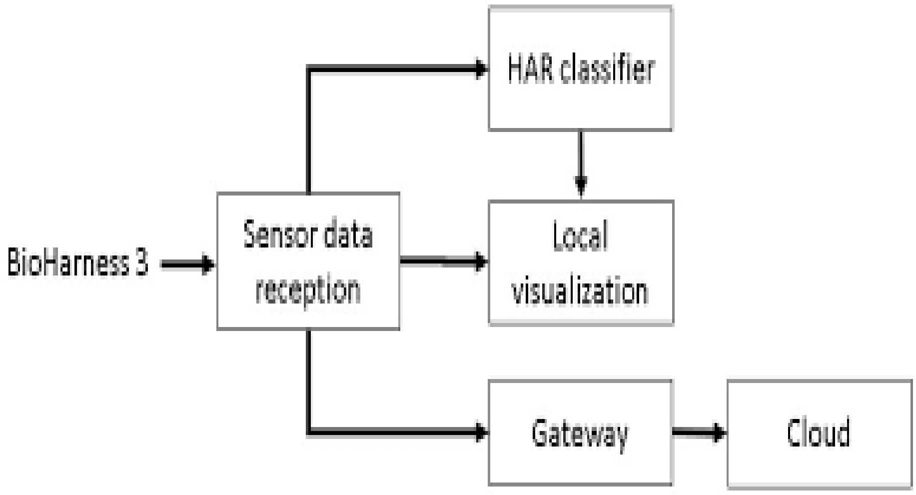


Figure 3.8: Solution architecture

3.21 Preprocessing

Work frames in a wide range of work information have different resolutions and domains, and are taken under different lighting conditions; therefore, a promotion module is required to improve the quality of the frames. Background information, lighting sound, and unnecessary details are reduced to this section for quick and easy operation [105]. Therefore, in order to solve the lighting effects, we used histogram calculations in the WS-HAR process preparation module. In addition, as shown in figure 3.9, we have removed human bodies by removing empty frames from work frames.

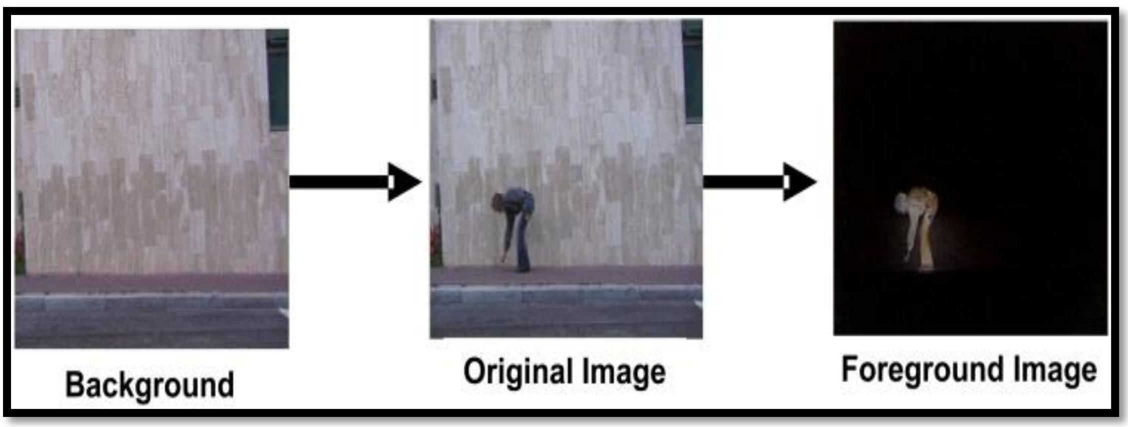


Figure 3.9: Subtracting human bodies by removing empty frame from a work frame.

3.22 1-D Convolutional Neural Network

Convolutional Variable network networks are designed to solve image separation problems, in which the model learns the internal representation of two-dimensional inputs, a process known as the learning aspect. The same procedure can be applied to single data sequences, such as acceleration and gyroscopic data for human activity recognition. The model learns to extract objects from the sequence of views and to map the internal features of different types of work. The model can learn the internal representation of time series data and, accordingly, obtain comparable performance from similar models in the built-in data version [92].

3.23 Classification of Moving Human Using KNN (K Nearest Neighbors)

To recognize human actions, the feature vector of each video must be determined. The feature vector is a person's trajectory tracking, and it is determined by the Kalman filter after background subtraction using the GMM method in the first and second steps of human action analysis [106]. Unlike many other machine learning methods such as artificial neural networks, kernel methods, wavelet networks, and so on, the k-NN method does not have a phase of parameters determining a function by bias of a mathematical optimization. Smartwatches are among the most well-known and widely used wearable devices. Because smartwatches are used by a large number of people and are constantly worn on the body, they are well suited for gathering data that can classify user activities in real-time.

3.24 Feature Fusion Algorithm

High-end features will take longer to train and differentiate. Furthermore, when the size of the output elements is too large, it is difficult to combine both fixed and dynamic features. To reduce the size of the input data d, k eigenvectors with the highest eigen values were selected to form a dimensional matrix W. Input data is converted to a new k-dimensional subspace using Eigen vector size.

The following are the steps involved in compiling a feature [107]:

Statistical gait features such as stride length, knee-to-ankle distance, and ankle-to-ankle distance is excluded from all frames and quoted values are calculated as statistical values.

1. Product law is used between features to determine the vector fv1 of a feature.fv1 = sl * kk * aa
2. Features based on the model reveal hip angles, knees and ankles, as well as thelevel of angles from the toes.
3. The magnitude of the resulting vector is too large to be used as a system input. Keycomponent analysis is used to reduce the vector size of the feature.

3.25 Infrastructure

Every camera in a CCTV framework requires the capacity and ability to move video information to observation and capacity frameworks. These prerequisites may expect changes to the office's framework, for example, the establishment of new camera fittings [104]. There are four key factors to consider when planning a CCTV installation:

1) How much of the system infrastructure can be reused?
2) How will the new CCTV system integrate with existing business programs?
3) How will the new CCTV framework coordinate or add other existing section and arranging frameworks?
4) To what degree can the new framework stay with existing frameworks, and which framework will affect functional cycles and responsiveness?

Using existing foundation is a vital factor in diminishing hardware and establishment costs. Old coaxial link frameworks, for instance, can be changed over into IP frameworks utilizing Ethernet over coaxial link with a transformation module. Simple video and Ethernet can both be moved to the telephone line. How much a CCTV framework meets an overall security framework will assume a significant part in its exhibition. Any new establishment should be viable with existing frameworks until another framework is received and the old framework is overhauled or delivered.

3.25.1 Reliability and Maintainability

Without show taking time, speculating the dependability and support of another CCTV framework is troublesome. The fundamental thought and essential consideration are whether the new framework will actually want to play out the necessary capacities over the long run and how the staff will actually want to work, keep up with and work on the framework. Numerous systems, like this: can be utilized to expand client trust in new establishments [108].

- Consult with security staff and administrators from different foundations utilizing similar projects or utilizing a similar seller;
- Request contending merchants to run preliminary projects nearby or on seat test hardware before establishment; and
- Installing new hardware at least to guarantee that the usefulness addresses the issues of the association.

All specialists working in apparatus may should be ensured under the conditions of the seller certificate. A few merchants may require a support agreement to guarantee that they have staff accessible to react to the CCTV Technology Handbook sooner or later. Prior to purchasing and introducing gear, support needs should be perceived.

3.25.2 Connectivity Type

CCTV cameras can utilize one of two information move techniques:

Network Cameras

Associate with IP-based Network, like the Internet, and permit far off survey and recording. Superior quality (HD) network cameras are additionally accessible, which can give a ton of picture detail. such as the Internet, and allow remote viewing and recording. High definition (HD) network cameras are also available, which can provide a lot of image detail [109].

Analog Cameras

Despite the expanding utilization of digital network cameras, thereis still an analogue camera market. This could be due to the high cost of developing and transforming a new transfer system. Analog cameras have advanced resolution options, making them ideal for a variety of surveillance applications. These cameras have some network protection benefits in light of the fact that the coaxial link to which they are associated requires actual admittance to break [86].

Day/Night Cameras

Day/night cameras give adaptability by adjusting naturally to changing lighting conditions. These cameras take shading photos during the day and afterward change to high contrast around evening time to improve the picture quality. The camera utilizes picture examination or photoelectric sensor to identify when the infrared-cut channelis taken out and change to monochrome settings.

Low-Light or Night Vision Cameras

Low light or late evening seeing cameras are utilized to catch low light conditions. Slight light cameras are intended to be utilized at a particular degree of encompassing lighting, for example, indoor eating lighting, streetlamps, or full moon; they are not intended to be utilized in complete murkiness. Close infrared (NIR) and infrared (IR) cameras with IR lights incorporated into the most widely recognized kinds of night vision cameras utilized in CCTV frameworks. They are intended to permit the administrator to see the night scenes. The width of a CCTV camera around evening time is controlled by the force of the camera material, like the focal point and sensor, and the power of the IR light utilized. Camera picture sensor is the principal choice whether the picture is sent in shading or monochrome (different tones of one tone). Monochrome cameras utilize light at NIR frequencies without human visual keenness (i.e., spectroscopic) to record pictures, and shading cameras use channels for certain light sensor picture sensors to restrict that item to certain shading frequencies. These channels empower the sensor to identify and send shading and splendid light, which will be distinguished by an unfiltered, monochrome sensor. Monochrome cameras can catch pictures around evening time or in close dull conditions with more detail than the natural eye can see. Nonetheless, pictures taken with a monochrome camera during the day might not have any correlation with the subtleties on the grounds that the picture is made in apparent and infrared light, with an assortment of center airplane. Monochrome picture sensors are made for the most part of silicon and germanium material and have two particular ghostly reactions. This permits the picture sensors to work easily in the infrared light locale [109].

Thermal Imaging Cameras

In certain working environments, a hot camera might be expected to recognize impediments like mist or smoke. Warm imaging cameras distinguish infrared radiation or warmth that the natural eye can't see. These cameras are at present delicate to temperature changes of one-10th degrees Fahrenheit. Natural aquifers can't be seen through glass or water, yet they can give the picture a fog or a smoky little energy. This outcome in more honed, more clear pictures with a superior component definition. Temperature cameras are frequently introduced on gyro-settled, skillet and-slant gadgets, just as on boats and helicopters, night reconnaissance in dreary regions. They are likewise accessible in little convenient units with worked in markers for use in security, security and crisis reaction applications [110]. Hot camera picture sensors go down over the long run, so contact the camera creator to discover the presentation highlights over the long run and financial plan for the expenses of intermittent fixes and substitutions.

Miniature or Covert Cameras

There might be exceptional applications that require the establishment of little, covered up cameras as a component of a CCTV framework. Since these cameras will in general ensure against the climate, they may require an open-air house whenever utilized outside. These cameras are battery-controlled and can join worked in transmissions inside a conservative remote arrangement. Contingent upon their requirements, associations can look over an assortment of little and secret cameras.

Optional Camera Features

CCTV cameras can join an assortment of discretionary highlights to meet explicit natural execution prerequisites. A portion of these normal highlights are portrayed underneath.

Auto Scan

Some PTZ cameras can be set up to perform robotized errands. The expression "auto output" signifies a nonstop pattern of clearing in the review region [111].

Reset

Resetting the pre-modified altering and change of the focal point that the PTZ camera relaxes or when a specific sort of occasion happens. During an information caution, for instance, the camera might be set to show a worth thing in the review region or to zero in on the passage where the alert was performed.

Encryption

The camera with the capacity to shroud security may decide to obstruct portions of the video picture to ensure protection. PTZ cameras, for instance, can be utilized to screen a parking garage almost an apartment complex utilizing window photos of this secret structure [112]. This is an arrangement highlight (programming or equipment) that can be changed over to unpredictable and costly.

Slip Ring

Sliding ring is a sort of electrical association that permits the PTZ camera to turn without misshaping the sign/control link. Smooth rings can communicate a picture utilizing blazing pillars, or make an electric flow utilizing an elusive contact in the ring. Slip rings are defenseless against messiness and temperature changes.

Motion Detection

Cameras can be fitted with implicit movement finders that can be actuated to sound a caution when movement is distinguished inside the view field (FOV). An alert can be set to begin recording, telling administrators, or both. When there is regular development in the climate, movement recognition highlights can make irritating cautions [110].

Backlight Compensation (BLC)

Some cameras incorporate BLC settings. BLC can make up for huge contrasts in splendid foundation pictures by adding picture subtleties [113]. The BLC, for instance, empowers security faculty to see the subtleties of an individual strolling before a brilliant window.

Digital Noise Reduction (DNR)

This is a typical component for cameras that are intended to take pictures in regions with low light or dim. DNR eliminates sound from a video picture (grain looks are displayed as regions known as "raster"). This further develops picture lucidity, splendor, and comprehensibility [109]. Since there is less data outside the video, less computerized sound can likewise decrease extra room necessities.

3.25.3 Types of Image Sensors

CCTV cameras typically use image sensor technology depending on the integrated charge device (CCD) or semiconductor coupled with metal oxide semiconductor (CMOS). The image element, or pixel, is the littlest piece of the picture delivered on a strong state chip. Pixels, regardless of sensor type, are numbered, numbered, and filtered to provide a wide range of resolution, sensitivity, and spectral responses.

Charge-Coupled Device Sensors

CCD picture sensor cameras previously showed up in the CCTV business during the 1980s and presently overwhelm the sunshine and low light market, or NIR cameras. CCD innovation enjoys numerous upper hands over the cylinders utilized in the first camcorders. CCD picture sensors are little, produce little warmth, and photography doesn't care for "blooming." Blurring happens when the picture sensor is encased by an exceptionally high light source in FOV, making subtleties in certain pieces of the picture be lost. The CCD camera has a life expectancy of 5 to 25 years [114].

CCDs are comprised of three-dimensional clusters pixels, each delivering an electrical sign equivalent to the measure of light it gets. This simple like electronic sign is moved to another chip, which changes over it into computerized information. This computerized information is prepared inside the camera and communicated to different pieces of the CCTV framework. Preceding transmission, a few cameras convert the pre-owned sign into a simple video.

The CCD-type face utilized on CCTV cameras resembles an airborne perspective on an all-around kept city. Each rooftop is light-touchy, and the streets are electrically controlled occasions [105]. The size of the rooftops (nerve components) in the roads (circuits) decides the affectability of the CCD to light and goal. A camera with a bigger region committed to sensors has high affectability and great goal. Low affectability and solid goal are brought about by limited quantities. Furthermore, a few makers offer CCDs with tactile focal point sensor capacities, permitting them to gather all the lighter and increment affectability.

Despite the fact that CCDs are delicate to apparent light, they are additionally touchy to approach infrared light. The unearthly sensor reaction is synthetically controlled by a portion of its segments and can be additionally handled utilizing filtration innovation. Infrared-cut channels, intended to lessen or dispose of IR brilliance, can be utilized before the sensors to work on the nature of the shading camera picture. Since monochrome cameras are intended for use close to the infrared range, they don't have such channels. During the day, monochrome cameras make a picture utilizing NIR noticeable light; hence, the picture of the day might be less clear than the picture of the evening. A few cameras utilize three CCDs to further develop goal. A crystal separates the white light going through the perspective into red, green, and blue in this setup. The vector yield is then determined and consolidated by three monochrome CCDs, every one of which has been separated to get just one tone. Similar standards apply to cameras with three CMOS sensors. A few cameras utilize three CCDs to further develop change. The crystal isolates the white light from the focal point into red, green, and blue in this change [106]. The vector yield then, at that point computes and joins three monochrome CCDs, every one of which is sifted to get just one tone. Similar standards apply to cameras with three CMOS sensors [113].

CMOS Sensors

CCD sensors are usually utilized in applications that require high-goal pictures, yet CMOS sensors are ordinarily utilized in network cameras, PC applications, and cell phones. This is on the grounds that CMOS sensors utilize less force than CCDs and are more affordable to create. CMOS sensors, in contrast to CCD sensors, manage every pixel exclusively. They take a charge from every pixel sensor and move it as computerized information. This methodology decreases the requirement for additional handling in the exchange interaction [115].

Lenses

The CCTV camera focal point is the first in a progression of photos, including focal point, camera, transmission framework, picture the board and examination programming, and observing. The focal point coordinates light or infrared sensor imaging [108]. The motivation behind the focal point is to give an unmistakable, adjusted, and precise picture of the picture. High-goal imaging programs start with focal points intended to deliver excellent picture sensor. A few pieces of the photograph series can't manage the cost of the lower focal point.

While picking a focal point, consider the distance needed to zero in on objects, see field (FOV), camera picture sensor size, and lighting conditions. The focal points are isolated by their fixed length, generally communicated in millimeters, their most stretched out openings, regularly alluded to as the number f, and the size of the picture sensor for which they were planned [116].

Back Focus

The focal point transmission should be appropriately centered around the camera picture sensor to be utilized. This camera setting is known as back center. The technique utilized by cameras to change the distance between the outside of the picture and the outside of the focal point fluctuates [117]. The picture sensor on most cameras can be moved to and from. Some have wheels that can be turned to move the picture sensor, while others have a screw that can be pivoted to move the inward way. A few focal points likewise have back center changes. Long range focal points are more inclined to zero in on the back. At the point when the picture of the fixed item stays honed all through the zoom range, the long-range focal point inverts likewise. During this zoom activity, the focal point center ring ought not be contacted.

To zoom in pleasantly to center the zoom or changed focal point, the opening is set to the most extreme opening, and the central length of the long-range focal point is set to the longest setting. Utilize faint light or spot a dark impartial channel before the focal point to decrease the measure of light and power the opening totally to accomplish greatest focal point opening. An accomplished specialist can frequently center the camera while watching something nitty gritty at the scene in a handheld screen.

Specialists, for instance, can embed a center diagram inside the scene and spotlight on the camera until the Moiré impact is decreased, or a mid-range obscure

Resolution

Camera change shows its capacity to handle data in a picture. Choice outline can be utilized to decide level camera changes. The choice diagram is comprised of a few equal lines that addition inside, making the lines more modest and closer together at the lower part of the outline. The even change is dictated by where the lines are not noticeable as the camera filters start to finish, as displayed on the diagram scale. Vertical change is restricted by the quantity of examining lines that contain the TV picture and is frequently ignored on the grounds that it surpasses the huma power [118].

3.26 Non- Sub Sampled Contourlet Transform (NSCT)

This technique will provide a novel image fusion algorithm which is based on the pulse coupled neural networks (PCNNS) and visual attention model. The idea behind that is high-pass sub band coefficients are combined with their visual saliency maps as input to motivate pulse coupled neural networks [119]. The main moto of using this technique will help us to construct a flexible and efficient transform targeting applications where redundancy is not a big issue. Non- sub sampled contourlet transform (NSCT) is a fully shift-invariant, multiscale and multi direction expansion that has a faster implementation.

3.27 Non-Dominated Sorting Genetic Algorithm (NSGA-3)

The NSGA is a widely used multi-objective algorithm for processing various parameterized steps in a machine. It is also known for its ability to minimize or maximize the performance of various machining operations. [120].

3.28 Beta Chaotic Map

A chaotic map is commonly used to generate sequences. The proposed method splits the generated sequences into three phases: Permutation, Diffusion, and Substitution. The resulting sequences are then encrypted using the encryption scheme [121].

Chapter 4

Experimental Results and Considerations

Here, d ′ 1, d ′ 2, d ′ 3, d ′ 4, and d ′ 5 are initial attributes required for 5DHCM. q, r, s, t, u, v, and w define control attributes of 5DHCM. Chaotic attractors are represented in figures 4.1 and 4.3 for q = 30, r = 10, s = 15.7, t = 5, u = 2.5, v = 4.45, and w = 38.5. Figure 4.1 demonstrates the chaotic attractor in d 1 -d 2 planes, whereas figure 4.2 shows the chaotic attractor in d 1, d 2, and d 3 plane. The time series analysis of state attributes d 1 utilizing Eq. (4.1) as depicted in figure 4.3. It is demonstrated that d 1 achieves random behavior and non-periodic in nature. It contains 5 equilibrium points like (0, 0, 0, 0, 0), (2.39, -2.39, -6.85, 8.76, -143.09), (1.56, -1.56, 10.50, -5.71, -93.33), (-2.39, 2.39, 6.85, -8.76, 143.09), and (-1.56, 1.56, -10.50, 5.71, 93.33) [25]. The eigenvalues of equilibrium points are positive integers. Thus, 5DHCM can achieve better encryption keys and can be computed as:

$$d'1 = q\,(d2 - d1) + d2d3d4$$

$$d'2 = r\,(d1 + d2) + d5 - d1d3d4$$

$$d'3 = -sd2 - td3 - ud4 + d1d2d4$$

$$d'4 = -vd4 + d2d3d1$$

$$d'5 = -w\,(d1 + d2) \qquad (4.1)$$

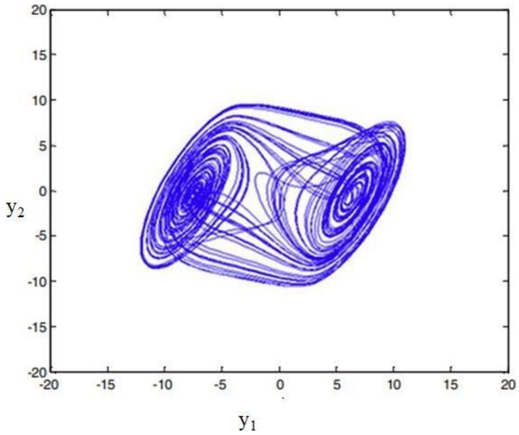


Figure 4.1: Chaotic attractor in y_1-y_2 plane

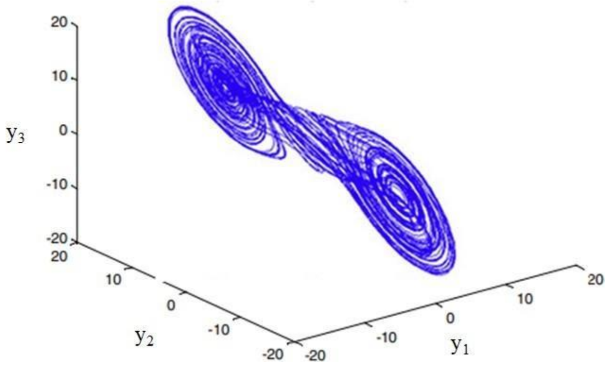


Figure 4.2: Chaotic attractor in y_1-y_2-y_3 plane

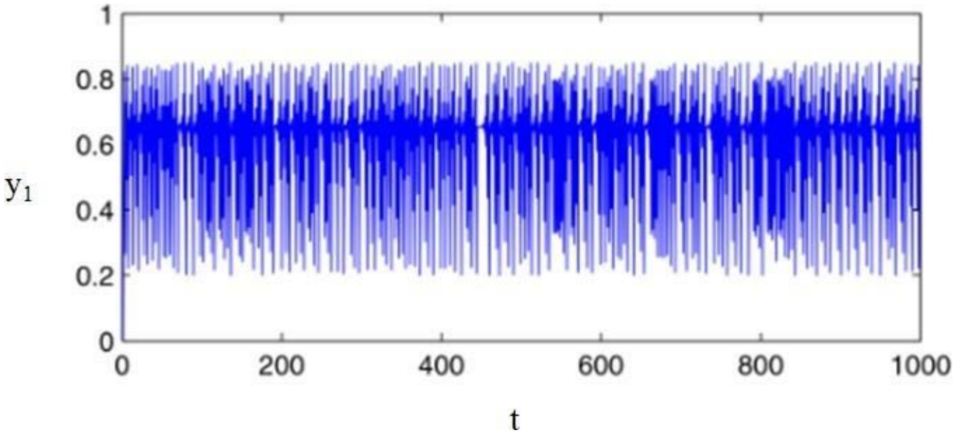


Figure 4.3: Time series plot for state variable y_1.

4.1 Secret keys

The process of secret keys generation using 5DHCM is elaborated in Algo. 5. It uses the optimized parameters that obtained from Algo. 2. The keys such as d1, d2, d3, d4, d5, d6, and d7 are generated using Algo. 5. To diffuse the image, these keys are utilized by Algo. 1 to obtain encrypted image.

4.2 Fitness Evaluation

To compute the individuals, the following fitness function is utilized Correlation and entropy are widely accepted parameters to obtain the optimal solutions because in image encryption, there is a need to reduce the relationship between the pixels. In the case of maximum entropy, each pixel carries the same amount of information. Both parameters hide the statistical behavior of the images from the attacker. The desired value of correlation and entropy should be near 0 and 8, respectively. The fitness function is described in algorithm 6.

4.3 Comparative Analysis

Simulations are done on MATLAB 2020*a* with the help of image processing toolbox. The proposed encryption model is compared with the five well-known competitive approaches by using various performance metrics. The initial parameters of PRNDO are assigned s F = 0.6, C_r = 0.2, N = 50, K_S = 170, and T = 5. Four surveillance camera images are used with 256 × 256 size.

4.3.1. Visual Analysis

Input and the respective encrypted surveillance images are shown in figure 4.4. It clearly indicates that the obtained encrypted surveillance images seem like a purely noisy images in nature. Therefore, no attacker can obtain any kind of details of the images.

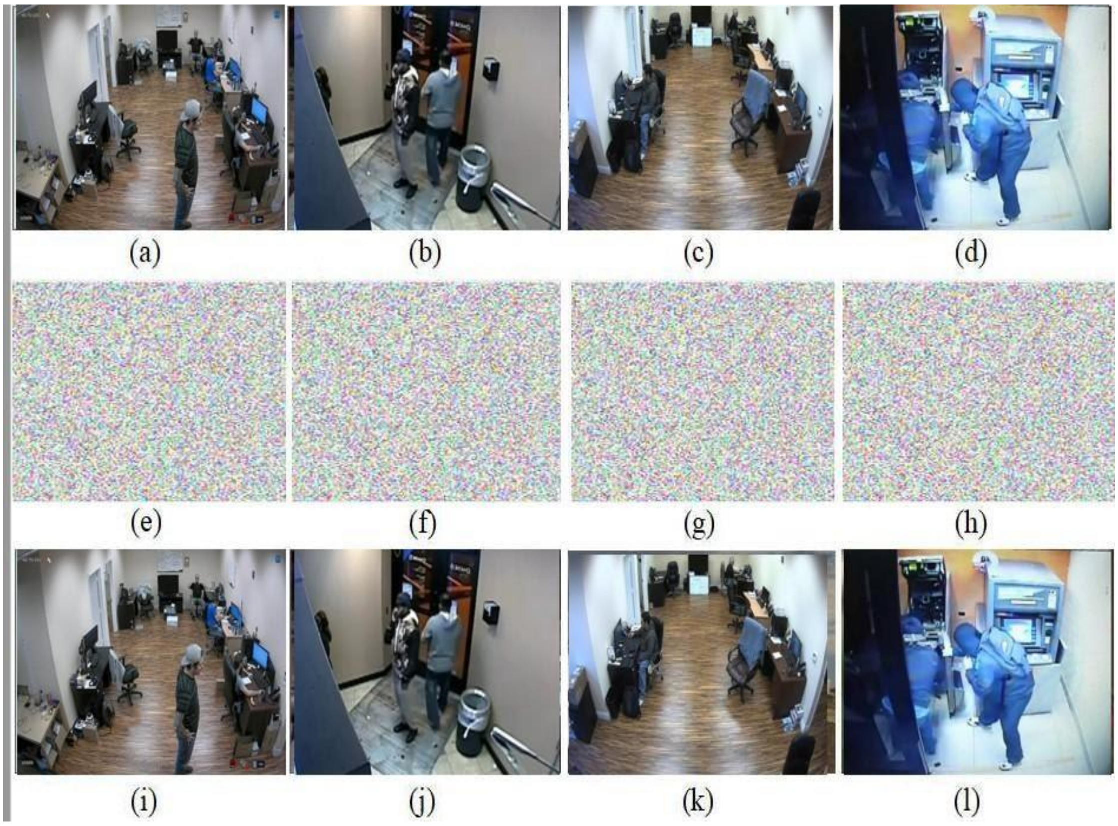


Figure 4.4: Input images: (a) to (d) input surveillance images. Encrypted images: (e) to (h) encrypted surveillance images. Decrypted images: (i) to (l) decrypted surveillance images

4.3.2. Entropy

Entropy computes the degree of randomness in images. In encrypted images it is desirable to be 8 [29]. An entropy analysis of the proposed secure surveillance system is depicted in table 4.1. It shows that the entropy value of all the techniques on surveillance images (SI_k), where k = 1, 2, 3 *or* 4 are near to 8, but the proposed secure surveillance system achieves more values than the existing techniques.

Table 4.1: Entropy analysis of the proposed secure surveillance system

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	7.9968	7.9836	7.9986	7.9886
CTCM [2]	7.9986	7.9846	7.9925	7.9860
DFRT [3]	7.9936	7.9959	7.9845	7.9899
HDNA-HC [5]	7.9880	7.9956	7.9836	7.9836
CSCM [7]	7.9972	7.9862	7.9969	7.9885
Proposed	**7.9980**	**7.9982**	**7.9986**	**7.9984**

Peak Signal to Noise Ratio (PSNR) analysis is depicted in Table 4.2 [30]. It is found that the proposed secure surveillance system outperforms the existing approaches.

Table 4.2: PSNR analysis of the proposed secure surveillance system

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	71.3826	71.7675	69.9256	69.3956
CTCM [2]	69.7869	70.4040	68.8362	70.9381
DFRT [3]	69.4288	69.1983	70.7452	71.9038

HDNA-HC [5]	70.6257	71.3539	69.9062	69.3482
CSCM [7]	69.9282	69.9264	70.9636	71.8936
Proposed	**80.9261**	**82.3689**	**80.8362**	**80.3782**

4.4.3 Mean Absolute Error

Mean Absolute Error (MAE) analysis is shown in table 4.3 [31]. It reveals that the proposed model obtains significant better values than the existing models.

Table 4.3: Mean absolute error analysis of the proposed secure surveillance system

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	79.4834	81.8382	79.3623	82.8262
CTCM [2]	83.8052	82.3617	84.3854	81.3822
DFRT [3]	81.8826	83.9083	81.8669	82.4837
HDNA-HC [5]	83.7068	81.1898	84.7656	84.0944
CSCM [7]	83.6945	85.6065	87.8084	86.8636
Proposed	**88.2847**	**90.8742**	**91.3508**	**90.6820**

4.4 Security Analysis

In this section, we will test the proposed secure surveillance system model against various security attacks. The objective is to evaluate that whether or not proposed secure surveillance system model can resist various security threats [120].

4.4.1. Histogram Analysis

Assessing the effectiveness of the proposed anti-retroviral prevention program is assessed using histogram analysis. It is found that the histograms of encrypted images should be identical. Historical analysis of the proposed model is shown in figure 4.5. DFT image frame is another global feature that contains details of the strength of the front object (e.g., title body region) if the thickness of the front object does not match the size of the background. This hypothesis is used to propose a DFT-based method of determining the geometric structure of a precursor of a local domain [29]. A typical image frame is divided into smaller blocks, and the average of all DFT values is calculated within each block. Finally, an algorithm closely relatedto K (KNN) is used to differentiate DFT features and produce the effect of job segregation. It is found that encrypted histograms acquired the proposed secure surveil- lance model are evenly distributed [121].

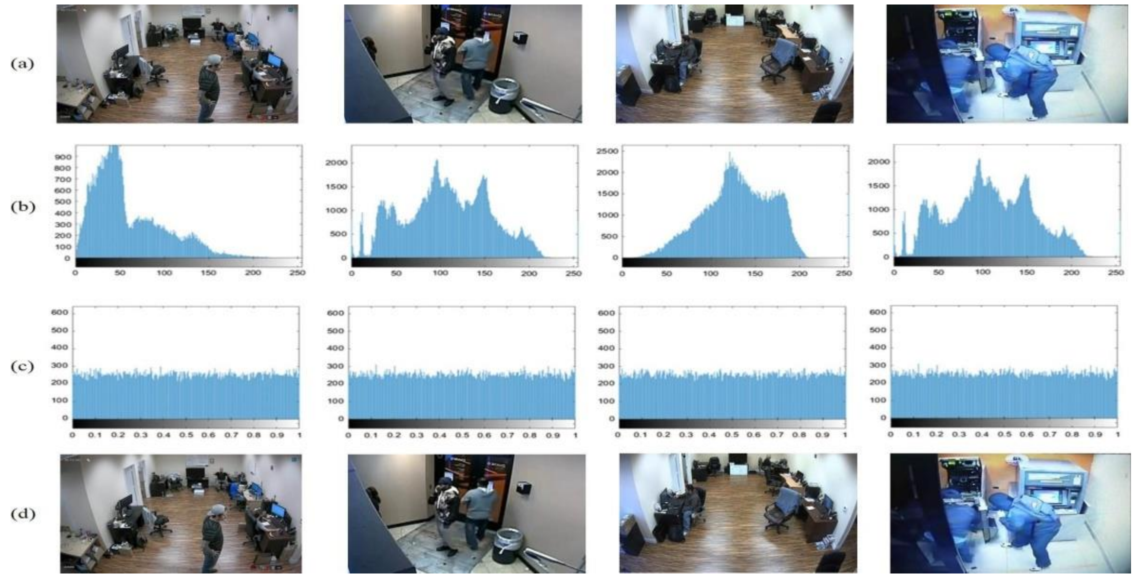


Figure 4.5: Histogram analysis: (a) surveillance images, (b) surveillance histogram, (c) surveillance images, and (d) hidden encryption histogram.

4.4.2. Correlation Analysis

Correlation can also be used to evaluate the statistical information of surveillance images. Thus, it is desirable that the correlation between the adjacent pixels vertically, diagonally, and horizontally should be minimum. The horizontal correlation analysis is depicted in table 4.4. It shows that the correlation among horizontal pixels is minimum than the existing encryption models. Vertical and diagonal correlation analysis are shown in tables 4.5 and 4.6. The minimum correlation values of the proposed secure surveillance system indicate that the proposed model can resist correlation attack.

Table 4.4: Horizontal correlation analysis

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	0.0028	0.0029	0.0038	-0.0022
CTCM [2]	0.0090	0.0060	0.0040	0.0022
DFRT [3]	-0.0050	0.0123	0.0090	0.0040
HDNA-HC [5]	0.0040	0.0070	-0.0090	0.0036
CSCM [7]	0.0072	0.0038	-0.0050	0.0055
Proposed	**-0.009**	**0.0016**	**0.0019**	**0.0021**

Table 4.5: Vertical correlation analysis

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	0.0036	0.0030	-0.0026	0.0016
CTCM [2]	-0.0066	0.0068	0.0026	0.0042
DFRT [3]	0.0026	0.0015	0.0022	0.0019
HDNA-HC [5]	0.0028	0.0016	-0.0025	0.0014
CSCM [7]	0.0016	-0.0024	0.0028	0.0024
Proposed	**0.0014**	**-0.0010**	**0.0013**	**0.0009**

Table 4.6: Diagonal correlation analysis

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	-0.0118	0.0072	0.0058	0.0028
CTCM [2]	0.0046	0.0034	0.0060	0.0032
DFRT [3]	0.0062	0.0044	0.0028	0.0026
HDNA-HC [5]	-0.0044	0.0034	0.0056	0.0038
CSCM [7]	0.0042	0.0050	0.0044	-0.0062
Proposed	**0.0019**	**0.0021**	**0.0018**	**0.0020**

4.4.3. Differential Analysis

Separated analyzes can be calculated using the Pixel Transformation Rate (NPCR) and Unified Average Change Intensity (UACI). An analysis of the NPCR and UACI of the proposed safe monitoring system is presented in tables 7 and 8. It is found that the model used gains better NPCR and UACI values than the competitive models. Therefore, the proposed model is sensitive with respect to minor modification to the input view images.

Table 4.7: NPCR analysis of the proposed secure surveillance system

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	99.40	99.52	99.48	99.50
CTCM [2]	99.44	99.40	99.50	99.52
HDNA-HC [5]	99.45	99.52	99.48	99.50
CSCM [7]	99.59	99.56	99.58	99.46
Proposed	**99.65**	**99.70**	**99.69**	**99.72**

Table 4.8: UACI analysis of the proposed secure surveillance system

Model	SI_1	SI_2	SI_3	SI_4
2LSS [1]	33.25	33.33	33.35	33.30
CTCM [2]	33.39	33.40	33.38	33.36
DFRT [3]	33.30	33.37	33.40	33.38
HDNA-HC [5]	33.38	33.42	33.42	33.48
CSCM [7]	33.40	33.50	33.50	33.49
Proposed	**33.55**	**33.58**	**33.64**	**33.62**

4.5 Robustness Against Various Attacks

In this section, we will evaluate the performance of the proposed secure surveillance system against various attacks such as noise and enhancement attacks.

4.5.1. Noise Attack

Robustness of the proposed secure surveillance system is evaluated against Salt & pepper and Gaussian noise attacks with various densities of noise. figure 4.6 shows noisy encrypted images using Gaussian noise with mean (μ) = 0.1 and variance (σ^2) = 0.1, μ = 0.3 and σ^2 = 0.3, μ = 0.5 and σ^2 = 0.5. The corresponding decrypted images are depicted in figures 8 (d), (e) and (f). The corresponding PSNR values among actual and decrypted images are 27.58 dB, 21.39 dB and 13.84 dB. The visual analysis shows that the proposed model can preserve details from the attacked images.

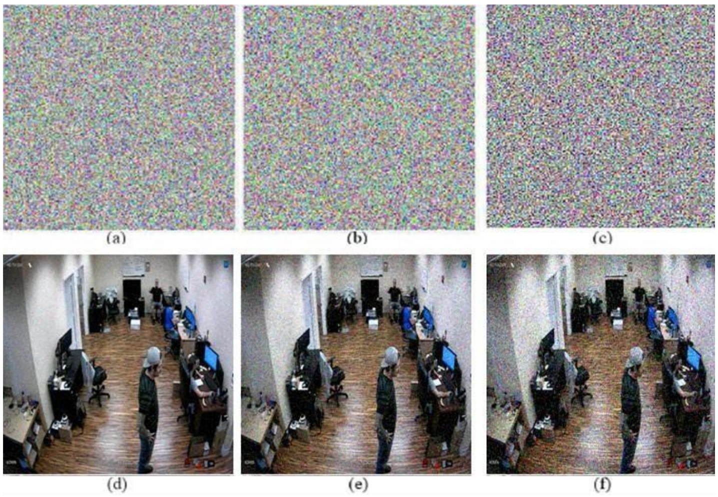


Figure 4.6: Gaussian noise attack based encrypted images: (a) with μ = 0.1 and σ^2 = 0.1, (b) with μ = 0.3 and σ^2 = 0.3, (c) with μ = 0.5 and σ^2 = 0.5. corresponding decrypted images: (d) evaluated using (a), (e) evaluated using (b), (f) evaluated using(c).

Figure 4.6 shows salt & pepper noise affected encrypted images with density = 0.1, 0.3, and 0.6, respectively. Figures 4.9 (d), (e) and (f) demonstrates decrypted images evaluated from the respective noisy images. The corresponding PSNR among SI1 and decrypted SI1 are 32.13 dB, 23.73 dB, and 17.25 dB. The decrypted images show that the proposed model can decrypt images even from high density of noise.

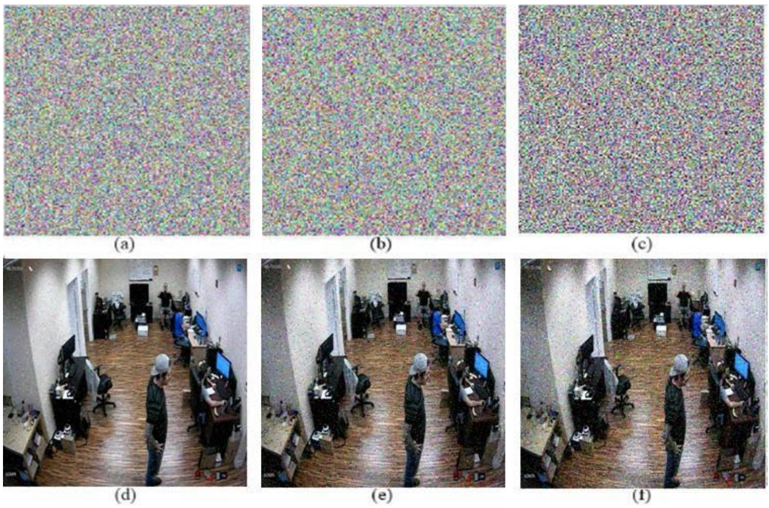


Figure 4.7: Salt & pepper noise based encrypted images (a) with density = 0.1, (b) with density =0.3, (c) with density = 0.6, corresponding encrypted images: evaluated using (a), (e) evaluated using (b), (f) evaluated using (c).

4.5.2. Enhancement Attack

Attackers may utilize image enhancement to damage the encrypted images. Attacked encrypted images obtained from various enhancement attacks such as gamma correction, histogram equalization, and adaptive histogram equalization are shown in figures 4.8 (a), (b) and (c). The corresponding decrypted images are depicted in figures 4.8 (d), (e) and (f). The corresponding PSNR values are 27.35 dB, 19.72 dB and 14.93 dB. Visual analysis also reveal that the proposed model can provide Potential information of actual images.

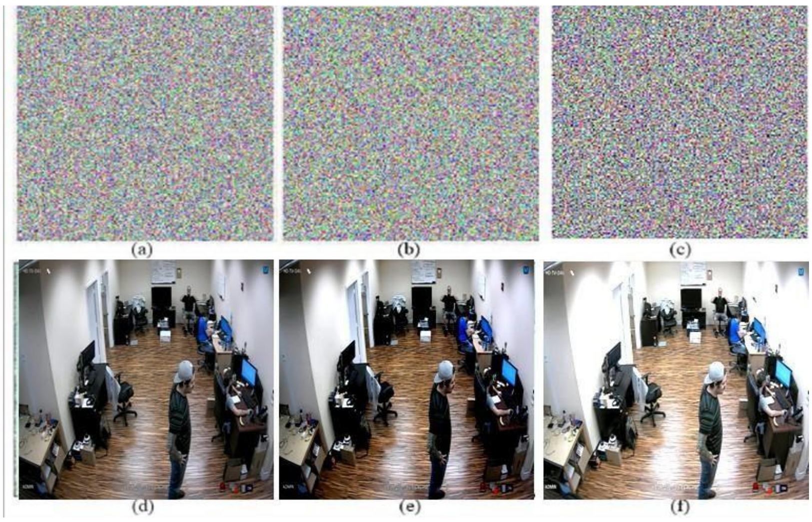


Figure 4.8: Enhancement based encrypted images: (a) gamma correction, (b) histogram equalization, (c) adaptive histogram equalization. corresponding decrypted images: (d) evaluated using (a), (e) evaluated using (b), (f) evaluated using (c).

4.6. Key Analysis

To evaluate the performance of the proposed secure surveillance system against brute-force attack , key sensitivity and key space analysis are considered [123].

4.6.1 Secret key Space

The proposed secure surveillance system computes the secret keys by 5DHCM. It requires 7 state attributes (i.e., d_1, d_2, d_3, d_4, d_5, d_6, and d_7) and 7 control attributes (i.e., q, r, s, t, u, v, and w. Assume that the precision of attributes is 10^{-16}. Thus,each attribute has 10^{16} possible values. Thus, the proposed secure surveillance system has Key Space (KS)$\approx 10^{119}$.

$KS = d_1 \times d_2 \times d_3 \times d_4 \times d_5 \times d_6 \times d_7 \times q \times r \times s \times t \times u \times v \times w$

$= 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times 10^{16} \times$

$10^{16} \times 10^{16}$

≈ 10

Since the *key space* > 10^{30} ($\approx 2^{100}$), therefore, the proposed encryption model can resist brute-force attack.

4.6.2 Secret Key Sensitivity Analysis

The secret keys are desirable to be more sensitive towards tiny changes. 7 keys are evaluated to implement the proposed encryption. These 7 keys are computed by Eq. (4.1) with the help of initial state (i.e., u1, u2, u3, u4, u5, u6, and u7) and control (i.e., s, r, t, i, q1, p, n, e, g and q2) attributes. Assume that a tiny change is done in the bits of u1. To verify the key sensitivity, we will encrypt the surveillance images using actual key and with tiny change in the same key. The encrypted surveillance images with actual and modified keys are shown in figures 4.9 (b) and (c). Figure 4.11 (d) shows the error between figures 4.9 (b) and (c). The computed error seems like a completely random image. We have also used to decrypt figure 4.9 (b) by utilizing both keys. The evaluated results are depicted in figures 4.9 (e) and (f). If the identical key is used to decryption, then the obtained image is similar to the actual image. The decrypted image obtained using a tiny change in the actual key seems like a completely noisy image [122]. Thus, proposed model is very sensitive towards the encryption key. Sensitivity analysis: (a) actual surveillance image, (b) actual key based encrypted image, (c) same key with tiny changes based encrypted image, (d) Error between (b) and (c), (e) actual key based decrypted image, and (f) tiny changed key based decrypted image.

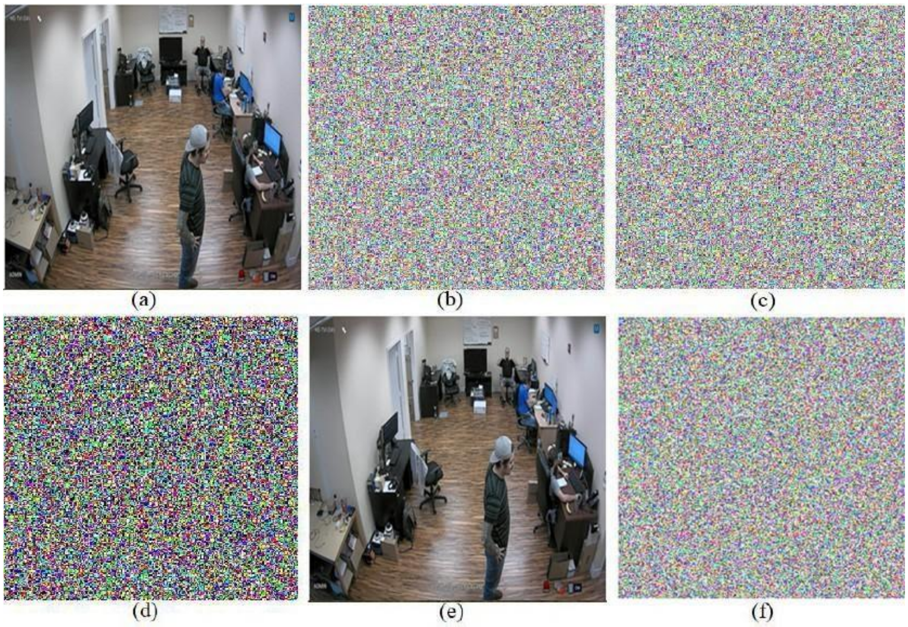


Figure 4.9: Sensitivity analysis: (**a**) actual surveillance image, (**b**) actual key-based encrypted image, (**c**) same key with tiny changes-based encrypted image, (**d**) error between (**b**) and (**c**), (**e**) actual key-based decrypted image, and (**f**) tiny changed key-based decrypted image

Table 4.9 : Shows the error between a modified an actual key based encrypted images. It demonstrates that the proposed secure surveillance system is sensitive towards initial key values.

Table 4.9: Error between actual and modified keys based encrypted images

	SI_1	SI_2	SI_3	SI_4
Error	99.9965	99.9958	99.9969	99.9971

4.6.3 Execution Time

Encryption and decryption execution time investigation are displayed in tables 4.10 and 4.11. These tables reveal the proposed model can encode and decode pictures at a higher speed, as it devours the least time than the current models.

Table 4.10: Encryption time analysis

Image	Dimensions	2LSS [1]	CTCM [2]	DFRT [3]	DNA-HC [5]	CSCM [7]	Proposed approach
SI_1	256 × 256	11.3	15.4	12.4	12.6	11.1	10.3
SI_2	512 × 512	58.8	60.8	56.8	55.6	58.6	39.4
SI_3	1024 × 1024	648.6	601.6	626.8	611.8	598.8	592.6
SI_4	2048 × 2048	2368.8	2658.8	2358.4	2261.2	1996.2	1872.1

Table 4.11: Decryption time analysis

Image	Dimensions	2LSS [1]	CTCM [2]	DFRT [3]	DNA-HC [5]	CSCM [7]	Proposed approach
SI_1	256 × 256	0.062	0.051	0.042	0.039	0.041	0.034
SI_2	512 × 512	0.049	0.051	0.043	0.052	0.046	0.037
SI_3	1024 × 1024	0.061	0.051	0.046	0.051	0.045	0.037
SI_4	2048 × 2048	0.064	0.051	0.043	0.052	0.039	0.035

Chapter 5

Conclusions and Future Scope

5.1 Conclusions

Human resource recognition has emerged as quite possibly the most well-known popular research topics in the field of mechanical learning. Typically, the proposed methods are intended for unique. Angles are flexi- ble features of scale and rotation, which is why this method is used. Because human activity occurs in a time zone, the slide kernel method is used to store these angle values. A common domain is used to test stored kernel objects. We have used HWT to compress data without losing time-tested data [123]. To reduce the cost of computers, the number of coefficients in Haar was reduced from 128 to 16 by converting it to a common base. Important information is retained and eliminated abstraction is used using this method. Following the reduction in size, a way to capture the novel feature is suggested. The opposite HWT method is used to extract the features. The KNN calculation is utilized to identify human exercises [124]. As a study, different classifiers and element vectors were contrasted with deciding if the proposed technique was a decent strategy for human-made acknowledgment. As indicated by test results, the proposed technique accomplished a high precision of 86.1 percent. All in all, an audit of different re-designing measures for HAR programs has been explored. The pro- posed programmed framework for RGB depth camera-based programming is by all accounts reasonable for video reconnaissance and grown-up care programs. Not with- standing, there are a few mistakes in the proposed calculation [125]-[127]. To start with, some unacceptable bone identification prompts the estimation of some unacceptable points, which consequently deludes the partition. Since the framework is prepared in positions in two diverse ways and in various situations, there might be disarray when attempting to distinguish work because of the closeness of the points and places of various undertakings [128]. Ultimately, as far as cost and usability, a basic RGB camera instead of an RGB depth camera can be neglected. Later on, assessing the exhibition of the proposed HAR framework utilizing a straightforward RGB cam- era could build its application re- gions [129],[130].

5.2 Future Scope:

This section discusses the future scope of the proposed technique.

- The balanced-whale, fruit-fly, Salp-Swarm, Spotted-hyena, etc. has been recently designed which can be utilized to tune the hyper-parameters in more efficient and effective manner.
- As TFCM has been implemented on low computing devices such as mobile, tablet, etc. So, Integrated NSCT, NSGA-3 and beta chaotic map (INNB) has been utilized in high computing device. Therefore, one may also relax hyper-parameters tuning and consider compressive sensing to reduce the execution time.
- This proposed algorithm has overcome the issue of parameter tuning, Insecure surveillance system etc., with existing variants of chaotic maps. Therefore, in the near future, we can propose a novel chaotic map with high dimension to enhance the performance of INNB.
- INNB can also be tested against different security attacks and cryptanalysis.
- To hide the existence of different secret images in surveillance, steganography canbe used in INNB.
- In the near future, the researcher can use higher dimension of chaotic map to secure the images during the communication. More the dimension of the chaotic map higher the security of the images.
- Different optimization technique can be used in a procedure which will execute iteratively by comparing various solutions or values until it reaches an optimum solution or satisfactory solution.
- The researchers while designing an algorithm different analysing factor should be considered, so that the proposed algorithm will provide good solution with better performance.

References:

1. J. Aggarwal and M. Ryoo, "Human activity analysis", *ACM Computing Surveys*, vol. 43, no. 3, pp. 1-43, 2011.
2. W. Lao, J. Han and P. De With, "Automatic video-based human motion analyzer for consumer surveillance system", *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 591-598, 2009.
3. X. Xu, J. Tang, X. Zhang, X. Liu, H. Zhang and Y. Qiu, "Exploring Techniques for Vision Based Human Activity Recognition: Methods, Systems, and Evaluation", *Sensors*, vol. 13, no. 2, pp. 1635-1650, 2013.
4. J. Johnson, "Design of experiments and progressively sequenced regression are combined to achieve minimum data sample size", *International Journal of Hydro mechatronics*, vol. 1, no. 3, p. 308, 2018.
5. A. Jalal and IjazUddin, "Security architecture for third generation (3G) using GMHS cellular network," *Proc. - 3rd Int. Conf. Emerg. Technol. ICET 2007*, pp. 74–79, 2007.
6. A. Jalal, S. Kamal and D. Kim, "A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments", *Sensors*, vol. 14, no. 7, pp. 11735-11759, 2014.
7. A. Jalal, S. Lee, J. T. Kim, and T. S. Kim, "Human activity recognition via the features of labeled depth body parts," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7251 LNCS, no. June, pp. 246–249, 2012.
8. Ç. Erdaş, I. Atasoy, K. Açıcı and H. Oğul, "Integrating Features for Accelerometer-based Activity Recognition", *Procedia Computer Science*, vol. 98, pp. 522-527, 2016.
9. A. Jalal and S. Kim, "Algorithmic implementation and efficiency maintenance of real-time environment using low-bitrate wireless communication," *Proc. - Fourth IEEE Work. Softw. Technol. Futur. Embed. Ubiquitous Syst., SEUS andthe Second Int. Work. Collab. Comput. Integr.*,pp. 81–86, 2006.
10. J. Aggarwal and M. Ryoo, "Human activity analysis", *ACM Computing Surveys*, vol. 43, no. 3, pp. 1-43, 2011.
11. A. Jalal, S. Kim, and B. J. Yun, "Assembled algorithm in the real-time H.263 codec for advanced performance," *Proc. 7th Int. Work. Enterp. Netw. Comput. Healthc. Ind. Heal.*, pp. 295–298, 2005.

12. K. Nai, Z. Li, G. Li, and S. Wang, "Robust Object Tracking via Local Sparse Appearance Model," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4958–4970, 2018.

13. D. Figo, P. Diniz, D. Ferreira and J. Cardoso, "Preprocessing techniques for context recognition from accelerometer data", *Personal and Ubiquitous Computing*, vol. 14, no. 7, pp. 645-662, 2010.

14. S. Kamal, A. Jalal and D. Kim, "Depth Images-based Human Detection, Tracking and Activity Recognition Using Spatiotemporal Features and Modified HMM", *Journal of Electrical Engineering and Technology*, vol. 11, no. 6, pp. 1857-1862, 2016.

15. D. Tao, Y. Wen and R. Hong, "Multicolumn Bidirectional Long Short-Term Memory for Mobile Devices-Based Human Activity Recognition", *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1124-1134, 2016.

16. Y. H. Li, Y. G. Pang, Z. X. Li, and Y. L. Liu, "An intelligent tracking technology based on Kalman and Mean shift algorithm," *ICCMS 2010 - 2010 Int. Conf. Comput. Model. Simul.*, vol. 1, pp. 107–109, 2010.

17. D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan, "Real-time vision-based camera tracking for augmented reality applications," *ACM Symp. Virtual Real. Softw. Technol. Proceedings, VRST*, pp. 87–94, 1997.

18. C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, pp. 246–252, 1999.

19. B. Engineering and S. Korea, "SELF-ORGANIZING MAPS Computer Science & Engineering , POSTECH , South Korea School of Computing and Maths , Charles Sturt University , Australia," no. July, 2017.

20. B. Morris and M. Trivedi, "A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1114-1127, 2008.

21. D. Gowsikhaa, S. Abirami, and R. Baskaran, "Construction of Image Ontology using low-level features for Image Retrieval," *Int. Conf. Comput. Commun. Informatics, ICCCI*, 2012.

22. A. Jalal, J. T. Kim, and T. Kim, “Development of a Life Logging System via Depth Imaging- based Human Activity Recognition for Smart Homes,” *Int. Symp. Sustain. Heal. Build.*, no. September, pp. 91–104, 2012.

23. K. Wang, X. Wang, L. Lin, M. Wang, and W. Zuo, “3D Human activity recognition with reconfigurable convolutional neural networks,” *MM 2014 - Proc. ACM Conf. Multimed.*, pp. 97–106, 2014.

24. M. Sreedevi et al., “Real Time Movement Detection for Human Recognition,” *Proceedings of the World Congress on Engineering and Computer Science*, Vol I WCECS, San Francisco, USA October pp. 24-26, 2012.

25. J. Wu, J. Wang and L. Liu, "Feature extraction via KPCA for classification of gait patterns", *Human Movement Science*, vol. 26, no. 3, pp. 393-411, 2007.

26. L. Jianguo, Anni Cai, Lili Li, “A Detection-Aided Multi-target Tracking Algorithm,” *International Conference on Machine Vision and Human-machine Interface IEEE computer society*, 2010.

27. P. Felzenszwalb, R. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645, 2010.

28. F. Sikder and D. Sarkar, "Log-Sum Distance Measures and Its Application to Human-Activity Monitoring and Recognition Using Data from Motion Sensors", *IEEE Sensors Journal*, vol. 17, no. 14, pp. 4520-4533, 2017.

29. L. Ovsenik et al., “Video Surveillance Systems with Optical Correlators,” MIPRO, Opatija, Croatia, pp. 23- 27, 2011.

30. L. Sigal, M. Isard, H. Haussecker and M. Black, "Loose-limbed People: Estimating 3D Human Pose and Motion Using Non-Parametric Belief Propagation", *International Journal of Computer Vision*, vol. 98, no. 1, pp. 15-48, 2011.

31. D. Singh and C. Krishna Mohan, "Graph formulation of video activities for abnormal activity recognition", *Pattern Recognition*, vol. 65, pp. 265-272, 2017.

32. J. Gall and V. Lempitsky, “Class-specific hough forests for object detection,” *IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, pp. 1022–1029, 2009.

33. J. Arunnehru, G. Chamundeeswari and S. Bharathi, "Human Action Recognition using 3D Convolutional Neural Networks with 3D Motion Cuboids in Surveillance Videos", *Procedia Computer Science*, vol. 133, pp. 471-477, 2018.

34. Y. Lin and Y. Jeon, "Random Forests and Adaptive Nearest Neighbors", *Journal of the American Statistical Association*, vol. 101, no. 474, pp. 578-590, 2006.

35. H. Kuehne, J. Gall, and T. Serre, "An end-to-end generative framework for video segmentation and recognition," *IEEE Winter Conf. Appl. Comput. Vision, WACV*, no. September, 2016.

36. T. Zhang, S. Liu, N. Ahuja, M. Yang and B. Ghanem, "Robust Visual Tracking Via Consistent Low-Rank Sparse Learning", *International Journal of Computer Vision*, vol. 111, no. 2, pp. 171-190, 2014.

37. F. Xiao, P. Hua, L. Jin, and Z. Bin, "Human gait recognition based on skeletons," *ICEIT 2010 - 2010 Int. Conf. Educ. Inf. Technol. Proc.*, vol. 2, no. Iceit, pp. 83–87, 2010.

38. A. Jalal and M. A. Zeb, "Security Enhancement for E-Learning Portal," *Int. J. Comput. Sci. Netw. Secur.*, vol. 8, no. 3, pp. 41–45, 2008.

39. A. Vincentelli and B. Vigna, "Autonomous vehicles: A playground for sensors", *7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)*, 2017.

40. T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications", 37*th IEEE Applied Imagery Pattern Recognition Workshop*, 2008.

41. K. Huang, Y. Zhang, and T. Tan, "A discriminative model of motion and cross ratio for view-invariant action recognition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2187–2197, 2012.

42. K. Moustakas, D. Tzovaras and G. Stavropoulos, "Gait Recognition Using Geometric Features and Soft Biometrics", *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 367-370, 2010.

43. L. Piyathilaka and S. Kodagoda, "Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features", IEEE *8th Conference on Industrial Electronics and Applications (ICIEA)*, 2013.

44. Z. Musa and J. Watada, "Multi-camera tracking system in a large area case", *IEEE International Conference on Fuzzy Systems*, 2009.

45. A. Jalal and S. Kamal, "Real-time life logging via a depth silhouette-based human activity recognition system for smart home services", *11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2014.

46. M. Hofmann, D. Wolf, and G. Rigoll. “Hyper graphs for Joint Multi-View Reconstruction and MultiObject Tracking”, *In Proc. CVPR*, 2013.

47. D. Ross, J. Lim, R. Lin and M. Yang, "Incremental Learning for Robust Visual Tracking", *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125-141, 2007.

48. G. Abdollahian, C. Taskiran, Z. Pizlo and E. Delp, "Camera Motion-Based Analysis of User Generated Video", *IEEE Transactions on Multimedia*, vol. 12, no. 1, pp. 28-41, 2010.

49. J. Shotton *et al.*, “Real-Time human pose recognition in parts from single depth images,” *Commun. ACM*, vol. 56, no. 1, pp. 116–124, 2013.

50. Z. Kalal, J. Matas and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints”, IEEE *Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.

51. P. Guo and Z. Miao, "Action Detection in Crowded Videos Using Masks", *20th International Conference on Pattern Recognition*, 2010.

52. J. Shajeena and K. Ramar, "A novel way of tracking moving objects in video scenes", *International Conference on Emerging Trends in Electrical and Computer Technology*, 2011.

53. J. Henriques, R. Caseiro and J. Batista, "Globally optimal solution to multi-object tracking with merged measurements", *International Conference on Computer Vision*, 2011.

54. M. Kolekar and D. Dash, "Hidden Markov Model based human activity recognition using shape and optical flow-based features", IEEE Region 10 Conference (TENCON), 2016.

55. N. Karunanayake, M. Gnanasekera, and N. D. Kodikara, “A Robust Algorithm for Retinal Blood Vessel Extraction,” *Int. J. Innov. Res. Comput. Commun. Eng. (An ISO Certif. Organ.*, vol. 3297, no. 9, pp. 8116–8121, 2007.

56. Y. Hbali, S. Hbali, L. Ballihi, and M. Sadgal, “Skeleton-based human activity recognition for elderly monitoring systems,” *IET Comput. Vis.*, vol. 12, no. 1, pp. 16–26, 2018.

57. A. Jalal, M. Quaid and A. Hasan, "Wearable Sensor-Based Human Behavior Understanding and Recognition in Daily Life for Smart Environments*", International Conference on Frontiers of Information Technology (FIT),* 2018.

58. C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “P finder: real-time tracking of the human body,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, 1997.

59. A. Vedaldi, V. Gulshan, M. Varma and A. Zisserman, "Multiple kernels for object detection", *IEEE 12th International Conference on Computer Vision*, 2009.

60. T. S. Cho, C. L. Zitnick, N. Joshi, S. B. Kang, R. Szeliski, and W. T. Freeman, “Image restoration by matching gradient distributions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 683–694, 2012.

61. S. Singh, S. Velastin and H. Ragheb, "MuHAVi: A Multicamera Human Action Video Dataset for the Evaluation of Action Recognition Methods", *7th IEEE International Conference on Advanced Video and Signal Based Surveillance,* 2010.

62. J. Berclaz, F. Fleuret, E. Turetken and P. Fua, "Multiple Object Tracking Using K-Shortest Paths Optimization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1806-1819, 2011.

63. K. Zhang, L. Zhang and M. Yang, "Fast Compressive Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 2002-2015, 2014.

64. W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, “Human detection using partial least squares analysis,” no. Iccv, pp. 24–31, 2010.

65. F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal and R. Bajcsy, "Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition", *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 24-38, 2014.

66. G. Wu and W. Kang, "Robust Fingertip Detection in a Complex Environment", *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 978-987, 2016.

67. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005*, vol. I, pp. 886–893, 2005.

68. R. Okada, "Discriminative generalized hough transform for object dectection", *IEEE 12th International Conference on Computer Vision*, 2009.

69. J. Yamato, Ohya, J & Ishii, K, “Recognizing human action in time- sequential images using hidden markov model”, Proceedings of the *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, vol. 92, pp. 379-385, 1992.

70. A. Jalal, M. Uddin and T. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home", *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 863-871, 2012.

71. A. Jalal and S. Kim, “Global Security Using Human Face Understanding under Vision Ubiquitous Architecture System,” *Eng. Technol.*, no. November, pp. 7–11, 2006.

72. M. Al-Nawashi, O. Al-Hazaimeh and M. Saraee, "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments", *Neural Computing and Applications*, vol. 28, no. 1, pp. 565-572, 2016.

73. R. Marie, A. Potelle and E. Mouaddib, "Dynamic background subtraction using moments", *18th IEEE International Conference on Image Processing*, 2011.

74. K. Goya, X. Zhang, K. Kitayama and I. Nagayama, "A Method for Automatic Detection of Crimes for Public Security by Using Motion Analysis", *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2009.

75. H. Possegger, T. Mauthner, P. M. Roth, and H. Bischof, “Occlusion geodesics for online multi-object tracking,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1306–1313, 2014.

76. Yuan Li, Haizhou Ai, T. Yamashita, Shihong Lao and M. Kawade, "Tracking in Low Frame Rate Video: A Cascade Particle Filter with Discriminative Observers of Different Life Spans", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1728-1740, 2008.

77. A. Jalal and M. A. Zeb, “Security and QoS Optimization for Distributed Real Time Environment,” pp. 369–374, 2008.

78. D. Sun, S. Roth, J. Lewis and M. Black, "Learning Optical Flow", *Lecture Notes in Computer Science*, pp. 83-97, 2008.

79. K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M. H. Yang, “Fast visual tracking via dense spatio-temporal context learning,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8693 LNCS, no. PART 5, pp. 127–141, 2014.

80. B. Benfold and I. Reid, “Guiding visual surveillance by tracking human attention,” *In BMVC,* September 2009.

81. Y. Li, Shi, D., Ding, B., & Liu, D., “Unsupervised feature learning for human activity recognition using smartphone sensors,” *In Mining Intelligence and Knowledge Exploration*, 2014.

82. J. Yuan, Z. Liu, and Y. Wu, "Discriminative video pattern search for efficient action detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 128, 2011.

83. Y. Chen and C. Shen, “Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition,” *IEEE Access,* vol. 5, pp. 3095–3110, 2017.

84. N. Prabhakar, V. Vaithiyanathan et al., “Object Tracking Using Frame Differencing and Template Matching,” *Research Journal of Applied Sciences, Engineering & Technology* 4, 2012.

85. N. Nguyen and N. Q. Ly, “Abnormal activity detection based on dense spatial-temporal features and improved one- class learning,” in *Proc. Int. Symp. Inf.Commun. Technol.*, Nha Trang City, Viet Nam, Dec. 2017.

86. N. K. Pareek, V. Patidar, and K. K. Sud, “Image encryption using chaotic logistic map,” *Image and vision computing*, vol. 24, no. 9, pp. 926–934, 2006.

87. S. Behnia, A. Akhshani, H. Mahmodi, and A. Akhavan, “A novel algorithm for image encryption based on mixture of chaotic maps,” *Chaos, Solitons & Fractals,*vol. 35, no. 2, pp. 408–419, 2008.

88. T. Gao and Z. Chen, “A new image encryption algorithm based on hyper-chaos,” *Physics Letters A*, vol. 372, no. 4, pp. 394–400, 2008.

89. G. Ye, “Image scrambling encryption algorithm of pixel bit based on chaos map,” *Pattern Recognition Letters*, vol. 31, no. 5, pp. 347–354, 2010.

90. ZL. Zhu, W. Zhang, K.-w. Wong, and H. Yu, “A chaos-based symmetric image encryption scheme using a bit-level permutation,” *Information Sciences*, vol. 181,no. 6, pp. 1171–1186, 2011.

91. C. Li, “Cracking a hierarchical chaotic image encryption algorithm based on per-mutation,” *Signal Processing*, vol. 118, pp. 203–210, 2016.

92. Y. Q. Zhang and X. Y. Wang, “Analysis and improvement of a chaos-based symmetric image encryption scheme using a bit-level permutation,” *Nonlinear Dyn.*, vol. 77, no. 3, pp. 687–698, 2014.

93. X. Y. Wang, T. Wang, D. H. Xu, and F. Chen, “A selective image encryption based on couple spatial chaotic systems,” *Int. J. Mod. Phys. B*, vol. 28, no. 6, 2014.

94. D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, “Incremental learning for robust visual tracking,” *Int. J. Computer Vision*, vol. 77, nos. 1–3, 124, 2008.

95. W. Zeng, W, Wang, C & Yang, “Silhouette-based gait recognition via deterministic learning,” *Pattern recognition*, vol. 47, no. 11,2014.

96. D. Geronimo, “A global approach to vision based pedestrian detection for advanced driver assistance systems,” *PhD thesis, Computer Vision Centre, Universitat Autonoma Barcelona*, 2009.

97. O. Barnich, “A universal background subtraction algorithm for video sequences,” *IEEE Transactions on Image Processing*, 20(6), pp. 1709-1724, June, 2011.

98. S. wei Li, J. qiang Wang, and Y. zhao Zhang, “Simulation study on pedestrian strategy choice based on direction fuzzy visual field,” *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 2, pp. 827–843, 2020.

99. F. Ince, “Human identification using video- based analysis of the angle between skeletal joints,” *J. Institute Contr. Robot. Syst.* 24, 2018.

100. A. Farooq, A. Jalal, and S. Kamal, “Dense RGB-D map-based human tracking and activity recognition using skin joints features and self-organizing map,” *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 5, pp. 1856–1869, 2015.

101. M. Mahmood, A. Jalal, and M. A. Sidduqi, “Robust spatio-Temporal features for human interaction recognition via artificial neural network,” *Proc. - 2018 Int. Conf. Front. Inf. Technol. FIT 2018*, no. December, pp. 218–223, 2019.

102. H. Yoshimoto, N. Date, and S. Yonemoto, “Vision- based real- time motion capture system using multiple cameras,” in *Proc. IEEE Int. Conf. Multisensor FusionInteger. Intell. Syst.*, Tokyo, Japan, Aug. 2003.

103. M. Ye, Y. Shen, C. Du, Z. Pan, and R. Yang, “Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1517–1532, 2016.

104. Q. Zhang and K. N. Ngan, "Segmentation and tracking multiple objects under occlusion from multiview video," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3308–3313, 2011.

105. J. Huang, J. Yang, and Y. Qiao, "Person re- identification across multi- camera system based on local descriptors," in *Proc. Int. Conf. Distribut. Smart Cameras*, Hong Kong, China, Oct. 2012.

106. D. Z Uddin, N. D. Thang, and T.- S. Kim, "Human Activity Recognition via 3- D joint angle features and Hidden Markov models," in *Proc. Int. Conf. Image Process.*, Hong Kong, China, Sept. 2010.

107. T. Ben-Zvi and J. V. Nickerson, "Intruder detection: An optimal decision analysis strategy," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 2, pp. 249–253, 2012.

108. D. Thang et al., "Estimation of 3- D human body posture via co- registration of3- D human model and sequential stereo information," *Appl. Intell.* 35, no2, 2011.

109. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

110. A. Jalal, N. Sarif, J. T. Kim, and T. S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home," *Indoor Built Environ.*, vol. 22, no. 1, pp. 271–279, 2013.

111. L. Johnson, "Reynold's stress statistics in the near nozzle region of coaxial swirling jets," *Int. J. Hydro mechatronics* 1, 2018.

112. N. Ravi et al., "Activity recognition from accelerometer data," in *Proc. Conf. Innovative Applicat. Artif. Intell.,"* Pittsburgh, PA, USA, July 2005.

113. V. Lumelsky, "Whole- body robot sensing and human- robot interaction," in *Proc. Int. Symp. Micro- Nanomechatronics Human Sci.*, Nagoya, Japan, Nov. 2012.

114. A. Jalal, S. Kamal, and D. Kim, "Individual detection-tracking-recognition using depth activity images,"*12th Int. Conf. Ubiquitous Robot. Ambient Intell. URAI 2015*, no. Urai, pp. 450–455, 2015.

115. M. T. Uddin and M. A. Uddiny, "Human activity recognition from wearable sensors using extremely randomized trees," *2nd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2015*.

116. A. Jalal, S. Kamal, and D. Kim, “A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 4, p. 54, 2017.

117. M. Ding and G. Fan, “Articulated and generalized Gaussian kernel correlation for human pose estimation,” *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 776–789, 2016.

118. V. R. Satpute, K. D. Kulat, and A. G. Keskar, “Variance Method for Human Detection and Tracking in Video Surveillance Applications,” *9th Int. Conf. Ind. Inf. Syst.*, pp. 1–6, 4793,2014.

119. A. Jalal and Y. Kim, “Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data,” *11th IEEE Int. Conf. Adv. Video Signal-Based Surveillance, AVSS 2014*, pp. 119–124, 2014.

120. B. K. Y. Chuen, T. Connie, O. T. Song, and M. Goh, “A preliminary study of gait-based age estimation techniques,” *Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA ASC 2015*, no. December, pp. 800–806, 2016.

121. V. O. Andersson and R. M. Araujo, “Person identification using anthropometric and gait data from kinect sensor,” *Proc. Natl. Conf. Artif. Intell.*, vol. 1, pp. 425–431, 2015.

122. C. BenAbdelkader, R. G. Cutler, and L. S. Davls, “Gait recognition using image self-similarity,” *EURASIP J. Appl. Signal Processing*, vol. 2004, no. 4, pp. 572–585, 2004.

123. B. Babenko, S. Belongie, and M. H. Yang, “Visual tracking with online multiple instance learning,” *IEEE Conf. Comput. Vis. Pattern Recognition, CVPR* , pp. 983–990, 2009.

124. A. Jalal, Y. H. Kim, Y. J. Kim, S. Kamal, and D. Kim, “Robust human activity recognition from depth video using spatiotemporal multi-fused features,” *Pattern Recognit.*, vol. 61, no. June, pp. 295–308, 2017.

125. A. Jalal, Y. Kim, and D. Kim, “Ridge body parts features for human pose estimation and recognition from RGB-D video data,” *5th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT* , no. September 2016, 2014.

126. B. Leibe, A. Leonardis, and B. Schiele, “Robust object detection with interleaved categorization and segmentation,” *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 259–289, 2008.

127. X. Luo, H. Tan, Q. Guan, T. Liu, H. H. Zhuo, and B. Shen, "Abnormal activity detection using pyroelectric infrared sensors," *Sensors (Switzerland)*, vol. 16, no. 6, pp. 1–17, 2016.

128. A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," *15th Learn. Technol. Conf. L T 2018*, pp. 29–34, 2018.

129. A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition,"*15th Learn. Technol. Conf. L T*, pp. 29–34, 2018.

130. F. U. M. Ullah, A. Ullah, K. Muhammad, I. U. Haq, and S. W. Baik, "Violence detection using spatiotemporal features with 3D convolutional neural network," *Sensors (Switzerland)*, vol. 19, no. 11, pp. 1–15, 2019.

Dissemination of Work

1. Gopal Ghosh, Sahil Verma, Kavita, Monica sood, “Internet of Things Based Video Surveillance Systems for Security Applications”, Journal of Computational and Theoretical Nanoscience, Vol. 17, No. 6, 2020.
2. Gopal Ghosh, Kavita, Divya, Sahil Verma, M N Talib, Mudassar Hussain Shah, “A Systematic Review on Image Encryption Techniques”, Turkish Journal of Computer and Mathematics Education, Vol.12, No.10, 2021.
3. Gopal Ghosh, Kavita, Divya Anand, Sahil Verma, Danda B. Rawat, Jana Shafi, Zbigniew Marszałek and Marcin Wozniak, “A Comparative Review on Non-chaotic and Chaotic Image Encryption Techniques”, Symmetry, Vol.13 No. 8, 2021.
4. Gopal Ghosh, Kavita, Sahil Verma, Jhanjhi, M N Talib, “Secure Surveillance System Using Chaotic Image Encryption Technique”, IOP Conference Series: Materials Science and Engineering, Vol. 993, 2020.
5. Gopal Ghosh, Divya Anand, Kavita, Sahil Verma, NZ Jhanjhi, MN Talib, “A Review on Chaotic Scheme based Image Encryption Techniques”, 2nd International Conference on Technology Innovation and Data Sciences-2021, Vol. 248,2021.
6. Gopal Ghosh, Divya Anand, Kavita, Sahil Verma, NZ Jhanjhi, MN Talib, “A Comparative Review on Non-chaotic and Chaotic Image Encryption Techniques”, 2nd International Conference on Technology Innovation and Data Sciences-2021, Vol. 248,2021.