CHAPTER ONE

INTRODUCTION

1.1 Introduction

Computer vision has grown from the processing of single images to the processing of a video streams. Accordingly, handling video data in real-time manner has become possible due to the recent developments in computer architecture, digital signal processing and communication networks [1, 2]. These developments broadened the implementation of the computers in detecting moving objects and characterizing their activities and so, it can be described as a significant step toward designing a machine capable of understanding intelligently the objects motion.

The intention of this study is to extract the available human motion criteria from video stream taken in an enclosed environment, in order to acquire the embedded knowledge from objects motion. Based on the extracted information the system is capable of building a security scheme able to emphasize the active sections in the image plane in order to assist the surveillance systems operators to catch events of interest. In addition to that, this study intended to propose a framework for reducing the storage capacity in the 24 hours surveillance systems based on crisp set logic algorithm.

1.2 Problem Statements

Due to the current advanced in video technology, video surveillance becomes an integral part of daily life. The surveillance cameras are already prevalent in banks, stores, and parking lots for security purposes. In addition to the previous security applications, the surveillance systems has been proposed to measure traffic flow, detect accidents on highways, monitor pedestrian congestion in public spaces, and log routine maintenance tasks at nuclear facilities [3].

Mounting video cameras turn out to be economical, but finding available human resources to observe the focus view of these cameras is costly and time demanding [4]. Even with all these developments the current traditional way of surveillance systems utilized the security officers to monitor the scene projected by the video camera, where the task of those officers is to observe the current scene to catch the events of interest. This duty is a time consuming process which need additional attention and leads to inadequate the surveillance capability. Thus, the video stream data is used only "offline" as a forensic tool, thus losing its primary benefit as an active real-time medium [5].

The second serious problem in the traditional surveillance systems appears in the way it stores 24 hours continuous daily video streams of data. If the system assumes that the capturing device frame rate is 25 Frame/sec, normally the system needs 2 MB in order to store a single frame. This leads to 50 MB/sec storing rate which it equal to 3000 MB/min or 180 GB/ hour regardless the events in the current scene [Altahir A. Altahir et al, 2008c]. Figure 1.1 shows the sequence of storage requirements in the traditional video surveillance systems.

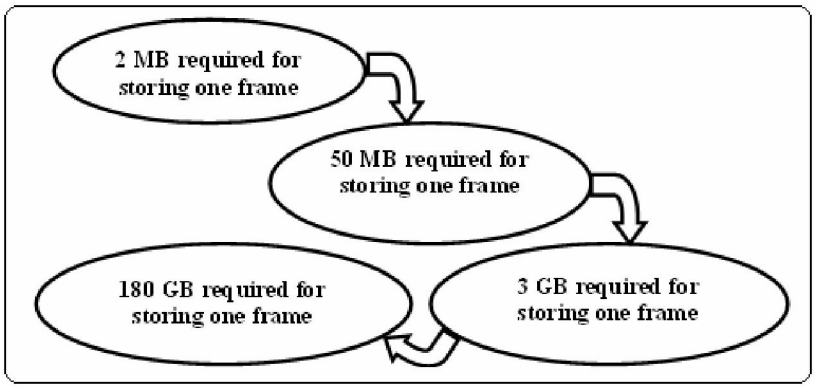

Figure 1.1: Storage Capacity Requirements.

This study is motivated by the problems stated above and the result of that is a semi automated video surveillance system capable of evaluating the moving objects activities. Moreover, the proposed system has an ability to support multi storing rates according to the current events in order to reduce the storage capacity.

1.3 Research Objectives

The target application of this research work is to adopt semi-smart surveillance systems. The following section aims to present the objectives and goals of the research study which is derived from the challenges observe from a fixed far vie the systems security surveillance video camera:

- To extract the motion trajectory, pixel frequency distribution and the time characteristics from objects motion based on full framing concepts.
- To interpret the full frame attributes from a security point of view.
- To extract the crossed distance, objects velocity and motion direction from objects motion based on inter framing concepts.
- To interpret the inter frame attributes from a security point of view.
- To implement selective criteria from the board of full frame and inter frame attributes via crisp set logic algorithm in order to emphasize the active sections in the image plane to assist the surveillance systems operators in catching the events of interest.
- To implement selective criteria from the board of full frame and inter frame attributes via crisp set logic algorithm in order to purpose a framework for reducing the storage capacity in 24 hours surveillance systems based on multi storing rate generation algorithm.
- To express the new directions of this work in order to help the future researchers toward designing an automatic, real time surveillance system.
- To present the drawbacks of the current system in details while suggesting the possible solutions.
- To express the benefits of implementing the proposed method for other applications such as traffic monitoring.

1.4 Research Methodology

The object detection part of this work is based on an existing method [6] and the consistent labeling technique is defined by Matlab software to track the foreground objects over the frames sequence. The video samples used in this work consist of blocks of 20 second length of video data, captured in far view, enclosed environment using a fixed camera. The set of the extractable motion attributes is classified into two categories full frame based attributes and inter frame based attributes. Each one of these categories consists of a group of attributes, these attributes is defined based on particular concepts. The output of the extraction phase is fed as input to crisp set image algorithm in order to emphasize the active sections in the current view and to generate the suitable storage rate for the current frame sequence. The rest of this section is dedicated to demonstrate the proposed system and its building blocks:

1.4.1 System Description

The video data manipulated at three levels: preprocessing level, full frame level and inter frames level. These three levels are followed by implementing the crisp set image algorithm in order to understand the underlying motion. The output of the crisp set image algorithm is implemented to support the operators in the process of generating the suitable reaction according to the current events. In addition, a suitable storage rate based on current events is determined.

The figure appears below describes the proposed system levels and illustrates the dependency between these levels, where the proposed system started by preprocessing level in order to prepare the input video signal for operations of later stage. The second level is extracting the full frames based attributes, this level followed by extracting inter frames based attributes. Finally, the crisp set logic algorithm is implemented and fed by a selective criteria from full frame and inter frames sets of attributes.

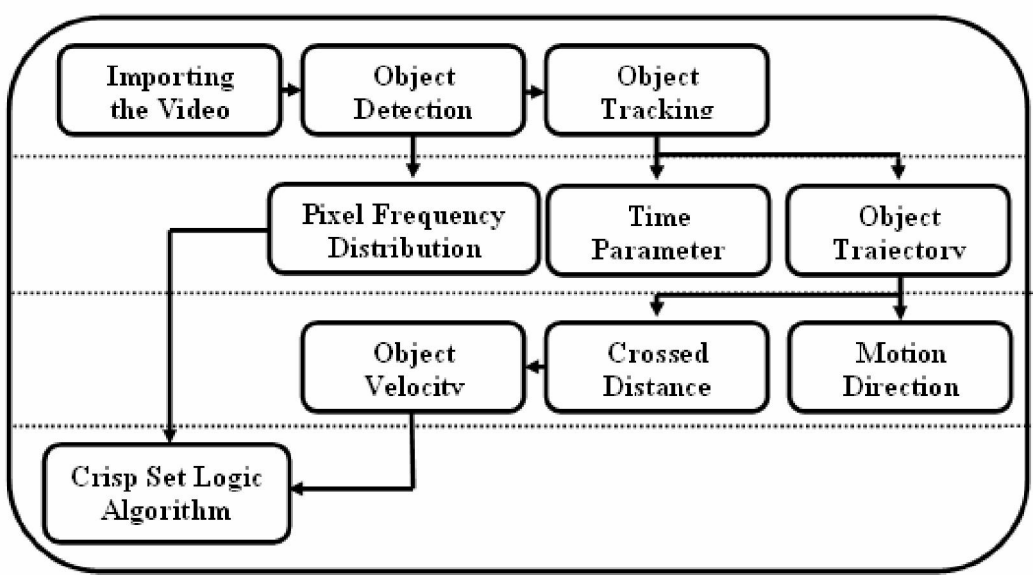

Figure 1.2: System Levels Dependency

In the same context, the next section describes these levels and the methods used to extract the components of each one of the levels.

1.4.2 Preprocessing level

The raw digital video signal at hand in this level is the intensity values at each picture element, pixel. The video signal is represented in terms of a multi-dimensional array (i.e., three-dimensional color at each pixel by two-dimensional image plane per video frame by one-dimensional time series). The main process in this level is segmenting the frames into foreground objects and background. This achieved based on a background subtraction method proposed by [6]. Differentiating between moving object and the background is represented in a binary image form, where the white regions correspond to the objects and black regions match up the background of the current scene. This level is considered as a preparation level before the high levels of the computer vision operations.

1.4.3 Full Frame Level

The term full frame refers to a continuous period of time represented by a sequence of successive frames. Full frames level intends to extract the motion attributes for the

objects of interest from the new arriving frames [Altahir A. Altahir et al, 2007, Altahir A. Altahir et al, 2008a, Altahir A. Altahir et al, 2008d]. The obvious example for this group of attributes is the object position, where a new position is defined based on the arrival of each new coming frame. Beside the object location, there are two more measurements the second one is temporal behavior of the existing objects and finally the pixel frequency distribution for the particular frame. More details about this topic will hold in chapter three. The next section provides an overview of these three measurements.

(a) Human Trajectory

Visualizing the object trajectory is counted as the first step in interpreting the global human motion [3]. The reason for that is because the trajectory generates accurate acquaintance about the object location at any instants during the simulation process. Moreover the trajectory is a functional tool to derivate more complex and powerful measurements such as time calculations, crossed distance and object velocity.

In this work the trajectory is generated via tracking the center of mass for the white regions over the time. The point correspondence is implemented based on evaluating the distance between the new position and a reference point and compare it with the measured distance between the previous position and the same reference point.

(b) Time Parameter

The temporal behavior of the object of interest is considered in this work. This concept is implemented via calculating the time spent by the object of interest in particular zones in image plane [Altahir A. Altahir et al, 2007, Altahir A. Altahir et al, 2008b, Altahir A. Altahir et al, 2008d, Altahir A. Altahir et al, 2008e]. Dividing zones realized through segmenting each frame into four equal zones, and then with the intention of calculating the time per zone, the system tests the existents of the object in each single zone. The idea behind that is the entire image plane is not important enough for the analyzing processes (e.g. there are walls, high places or any other area out of the range of the system consideration). Minimizing the area of interest provides more deterministic results about time spend by the objects in this area [Altahir A. Altahir et al, 2007, Altahir A.

Altahir et al, 2008a, Altahir A. Altahir et al, 2008d, Altahir A. Altahir et al, 2008e]. This concept is achieved via segmenting each zone mention earlier in new four sub zones. Determining the number of zones and zones boundaries is a supervised operation achieved based on the topology of the camera view and the security requirements determined by the user of the system.

(c) Pixel Frequency Distribution

The third full frame attributes in our work is pixel frequency distribution. This attribute is used to describe the activities in online manner for the entire frame region; also it may used to describe the activities in user defined zones [Altahir A. Altahir et al, 2008d].
The concept of the pixel frequency distribution is formed through the accumulating of the pixels intensity during the simulation time. Highly active area produce low peaks in the pixel frequency distribution surface and high peaks refer to the low activity area [Altahir A. Altahir et al, 2007, Altahir A. Altahir et al, 2008a].

1.4.4 Inter Frame Level

This section discuss the inter frame based concepts and analysis. As mentioned previously the system extracted the available measurements for the extracting object motion. Thus the system obtain some of these measurements required by determining a certain step through the frame sequence, so that the system are able to calculate them correctly [Altahir A. Altahir et al, 2008a, Altahir A. Altahir et al, 2008d].
The main reasons for determining a predefined step is that, the step provides a possibility to convert the system from depending on the arrival of each new frame as time unit into relying on the second as a time unit. Three different measurements can be defined based on inter frame based analysis and more details about these measurements will hold in chapter four. The next section provides a brief description for these three measurements.

(a) Crossed Distance

This measurement stands on the concept of inter frame based analysis in order to describe the crossed distance per a pre defined time step through the frame set, in this research 25

frames is used as fixed time step [Altahir A. Altahir et al, 2007, Altahir A. Altahir et al, 2008a]. Calculating the crossed distance is an essential stage in order to compute the object velocity. The mathematical base for computing the crossed distance is achieved via estimating the differences between the current center of mass coordinates and the previous one.

(b) Object Velocity

The most significant attributes in this work is the velocity of the object of interest. The importance of calculating the objects velocity in the current scene comes from the consideration of the velocity as fundamental tool for assessing the objects activity [Altahir A. Altahir et al, 2008a, Altahir A. Altahir et al, 2008d]. Objects with high velocities are considered more active and more interesting from a security point of view and thus having more suspicious behavior. The mathematical base for computing the object velocity in this work is based on the simple physics rule which it says the velocity is equal to the crossed distance divided by the time required to cross this distance. The detailed aspects of calculating the object velocity and the relative results are presented in chapter four.

(c) Motion Direction

The meaning of calculating the motion direction attribute is to provide basic information about the estimated direction of the objects motion. For example, if a person starts walking towards a forbidden zone or security-sensitive area, the motion direction gives an early warning to the security operators to deal with the current case. In addition to that, this attributes has a promising future in case of implementing this study in traffic monitoring, where each car in the highway must follow a well known directions.

The mathematical base for calculating the motion direction is relied on calculating the gradient of the object location function based on the partial derivative of the center of mass coordinates with respect to the time. It can be thought of as a collection of vectors pointing in the direction of increasing in the function values, then representing the

collection of the calculated values of the object position function as vectors pointing towards the estimated direction of the object of interest.

1.4.5 Crisp Set Images Implementation

The previous sections discussed extracting the motion features based on two levels of processing, full frame level and inter frames level. This section demonstrates the decision-making process based on the knowledge learned from the video data. The decision is necessary due to the requirements of the alerting system and reducing the storage capacity for the 24 hours surveillance systems.

These requirements lead the research work to implement a crisp set logic algorithm designed for decision-making process. The requirements are covered by proposing a set of crisp set image logic rules. The proposed rules derived from evaluating the system powerful attributes from the board of extracted measurements. Two features are chosen to fulfill the requirements; the pixel distribution from the board of full frame attributes and the velocity from inter frame based attributes. The ambition from generating decisions is to help human operator.

It's important to mention that the final decision must be taken by the human operator who controls the validity of the proposed decision [7]. More details about this topic will hold in chapter five.

1.5 Assumptions Taken Into Consideration

This section presents assumptions taken into consideration in order to overcome certain problems appeared in the course of executing this research experimental work. These assumptions are capturing the video samples in enclosed environment, implementing constrained tracking system and projection from three dimensional real worlds coordinates into two dimensional image plane coordinates. The justification of utilizing these assumptions is presented below:

- In order to detect objects in dynamic environment with illumination changes, motion of small objects and the shadow of the moving objects, there is a need to use background subtraction method capable of adapting itself in such a dynamic environment. Hence, the background subtraction method must consider all these factors mention above and generate a suitable background in order to grantee correct detecting process.
- The second assumption is the constrained tracking system. The prefect tracking system must avoid occlusion problem which it occurred when one or more object cover another object from the view of the camera.
- The study also includes projection from three dimensional real worlds coordinates to two dimensional image plane coordinates, which it results in approximate values for the set of extracted attributes. The validation of the obtained results is considered out of the range of this study.

Due to the reasons mentioned above this work, this study choose to focus on the implementation of a basic background subtraction method proposed by [6], accompanied by a constrained tracking system. So one of the suggestions in the future work mentioned in chapter six, is to replace these methods by powerful methods capable of overcoming the problem discussed in this section.

1.6 Organization of the Thesis

The main aspects of this thesis are structured into six chapters. These chapters are organized as follows:

- Chapter one is referred to as the introduction chapter, the problem statements, objectives and the research methodology are discussed in this Chapter.
- The essential preparation work for this study is given in chapter two.
- Chapter three describes the concepts, implementation and results of full frame based analysis for different case studies.

- Chapter four describes the concepts; implementation and results of inter frame based analysis for different case studies.
- Chapter five discusses the mining operations based on crisp set logic algorithm.
- Chapter six concludes this thesis by discussing the current drawbacks and the future directions of this work.