Event Triggering With Crisp Set Using Images of Surveillance Systems

Vijanth S. Asirvadam, Member IEEE, Altahir A. Altahir and Patrick Sebastian, Member IEEE,

Intelligent Signal and Imaging Research Cluster,
Universiti Teknologi PETRONAS
Tronoh, Bandar Seri Iskandar, 31750, Perak D Ridzuan, Malaysia.
Email: vijanth_sagayan@petronas.com.my

Abstract— Observing moving objects using far field’s video surveillance is core areas of research in computer vision. Similar application also can be found in surveillance systems and traffic monitoring applications. This paper presents implementing of crisp set technique on video images in order to evaluate moving objects in far field’s surveillance systems. One main objective is to create an event trigger if the scene is active. The concept is based on extracting two powerful attributes from moving objects, namely velocity and pixel frequency distribution. Combining the two measurements mentioned using crisp set rules in order to evaluate the active section in the image plane. The experimental results proved the efficiency of the technique.

Index Term— Surveillance System, Image Processing.

I. INTRODUCTION

The capability in detecting moving object or monitor monitor human activities using computers in security-sensitive areas such as airports, borders will be subject of interest great to the enforcement unit [6]. Moreover, creating such abilities will save tremendous human effort in sorting and retrieving from video sequences of images. This paper presents a robust approach of knowledge extraction from using crisp set techniques in order to trigger the system if there is significant about of movement of object in an area of focus. The main idea of crisp set mathematics is constructed on two basic components of set theory, the sets themselves and the operations on those sets.

Due to the requirements of assisting the human operators in the surveillance systems to catch the events of interest and evaluating the suitable storage rate, this paper introduces a novel trend of implementing crisp sets on image processing in order to create semi automotive surveillance system. This concept is based on extracting two main component of measurements from a moving object [1]-[3] and then combine the measurements mentioned above via crisp rules in order to evaluate the active section in the image plane. The basic operation of any form of video mining involves extraction of the pixel distribution and object tracking procedures. By extracting the moving object velocity, derived by time and the pixel distribution of object a weight is then assigned for each single second in the video stream. Finally based on the dedicated weight a suitable trigger alarmed rate is allocated to each second in the video data based on the scene.

Fig. 1 shows various surveillance image scene of a single human with different characteristics of movement.

(a)

(b)

(c)

Fig. 1. (a) Single human walking. (b) Single human walking then stopping. (c) Single human running.

The rest of this paper is organized as follows: section two discusses the object detection. The criteria used for information extraction is given in section three. Section four describes the proposed method used for evaluating the event trigger activities. The conclusion appears in section five which discuss the future direction of the work could possibly lead to.

II. OBJECT DETECTION

Detecting moving object in a video sequence is a crucial part of any video image sequence analysis applications. In this work the detecting process is required due to the need of extracting the pixels distribution and the object velocity in which the detection process is realized based on static background subtraction method created by [4-6]. The common image processing method will segment each frame into foreground regions {moving objects} and background {the static scene}. The detection process usually followed by thresholding the resulted image with a fixed threshold to
evaluate the image pixels after which the background image is updated. Meanwhile in the case of the tracking part, the set of the points over the frame sequence obtained from extracting the center of mass (or centroid) location for the moving objects from a binary image. The correspondence operation is derived by evaluating the distance between the new position of the object of interest and a reference point, the previous step followed by comparing the obtained distance with the measured distance between the previous position of the object of interest and the same reference point. The evaluation considers two close measured distances refers to the same trajectory of a particular object.

The advantages of this method are the simplicity and low computational power required which will assist in the application of real time surveillance systems, while the obvious drawback for this method it fails in handling occlusion which is not look into in this research work and should be considered. One possibility will be by including another set camera and let work cooperatively, a possibility of one of the future improvements for this research.

There are different ways to specify the position of a certain object in a sequence of images such as using virtual enclosed rectangle (or bounding-box) or the center of the object area. In order to construct the trajectory; the position of the object for each image frame is acquired. In this study, the position of an object in a particular frame is defined based on the center of the object region. The center of the white pixels area for binary images is the same as the center of mass if we consider the intensity at a point as the mass at that point. To calculate the position of the object, we used a set of equation mentioned in [6][8]:

\[
\bar{x} \sum_{i=1}^{n} \sum_{j=1}^{m} B(i, j) = \sum_{i=1}^{n} \sum_{j=1}^{m} j B(i, j)
\]

(1)

\[
\bar{y} \sum_{i=1}^{n} \sum_{j=1}^{m} B(i, j) = \sum_{i=1}^{n} \sum_{j=1}^{m} i B(i, j)
\]

Where \( \bar{x} \) and \( \bar{y} \) are the coordinates of the center of the region. Thus, the position of an object is:

\[
\bar{x} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} j B(i, j)}{A}
\]

(2)

\[
\bar{y} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} i B(i, j)}{A}
\]

The position calculated using first order moments is not necessarily an integer and usually lays between the integers values of the image array indices. So the calculated values are rounded to the nearest integers.

Each new incoming image frame describes an updated version of the object position in the image plane. This position defines on the variation in the object area over time, which in turn is a direct result of the object or human motion. However the variation occurred in the rows and columns of pixels in the image plane and corresponding the x-coordinates and y-coordinates in the trajectory plane. The time unit per variation leads to an event is an image frame, so each frame defines the object location at a particular time instance. Thus the x and y coordinates for the center of mass, starting from the first frame (where the object appears in the camera view for the very first time) till the last frame (where the object leaves the camera view).

The combine list of x and y coordinates will leads to pixel footing (or mapping) which form the basis of knowledge extraction of any moving objects for surveillance systems.

III. EXTRACTION OF INFORMATION

This section aims to illustrate extracting pixel frequency distribution and objects velocity in order to evaluate human activities. Throughout this work, 530 frames (30 frames per second settings) of video sample containing two agents walking, is used for the purpose of evaluating human activities [7][8]. The evaluation results are presented in the form of visual alarms in the particular region which witnessed the high level of activity. The subsections below dedicated to discuss the extraction process involves in small surveillance testcase setup.

A Pixel Frequency Distribution

Pixel frequency distribution is considered as one of the main powerful criteria for analyzing the moving objects, where it used to describe the motion activity with respect to the location, although in is computationally intensive when implemented during real time surveillance. Pixel frequency presents the details of the motion in a visual form which detects the accumulated amount activities over time [8-11]. Equation (3) illustrates the process of calculating the pixel frequency distribution from a mathematic point of view:

\[
F(i, j) = \sum_{i=1}^{K} \sum_{j=1}^{M,N} I(i, j)
\]

Equation (3) illustrates the process of calculating the pixel frequency distribution from a mathematic point of view:

Having \( F \) stands for the accumulation of pixel intensities \( I(i, j) \) which is the pixel value corresponding to the location \( i, j \) in a digital image, \( M \) and \( N \) the number of rows and columns and \( K \) is the number of frames.
Fig. 2. Pixel frequency distribution for two individuals walking in parallel and stop at the edges.

Fig. 2 shows pixel frequency distribution for the video sample used for moving objects where there are two individual walking towards the edge of the view point.

B Velocity

The definition of the term velocity in our work is a measurement describes how fast the object moves from one point in the image plane to another point in the same image plane on image pixels point of view [12][13]. Object velocity is calculated based on the crossed distance and the time needed to achieve crossing the particular distance. Regarding calculating the crossed distance we need first to form the motion trajectory, Figure 3 presents the objects trajectories:

After generating the trajectories, a suitable step between the trajectory points is needed in order to calculate the crossed distance. The necessity of utilizing time threshold comes from avoiding the ignorable small distance between the successive frames. Moreover, it also considers the details of the motion behavior, for example if we consider finding the crossed distance between the first point of the motion {the very first point in the object trajectory} and the last point, the result is straight line which in most cases differs from the actual crossed distance.

The mathematical base for measuring the crossed distance is the space between two points in two dimensions space. So crossed distance is calculated according to the Equation (4):

\[ D^k = \{d_1, d_2, ..., d_{i}, ..., d_n \} \]  \hspace{1cm} (4)

\[ d_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \]  \hspace{1cm} (5)

Having \( D \) is a crossed distances vector, \( k \) is the agent index, \( d_i \) the distance crossed in frame \( i \), \( x_i \) and \( y_i \) is the current Centroid coordinates, \( x_{i-1} \) and \( y_{i-1} \) is the previous Centroid coordinates and \( n \) the number of frames. The results of measured distances for the two agents are presented in Figure 4(a) and Figure(b):

The velocity of the agents is formed based on the calculated distance over the time needed to cross the particular distance. Equation (6) illustrates this concept:

\[ V^k = \{v_1, v_2, ..., v_i, ..., v_n \} \]

\[ v_i = \frac{d_i}{t} \]  \hspace{1cm} (6)

Where \( V^k \) is the vector containing the velocities, and \( v_i \) the velocity in frame \( i \). Figure (5) shows the velocity for the two agents participated in this video sample.
Fig. 5. The velocity the two agents.

(a) Maximum values - Pixel Frequency distribution per image frame.
(b) Maximum values - Pixel Frequency distribution per second.
(c) Maximum values - Objects Velocities per second.

Table II

<table>
<thead>
<tr>
<th>Pixel Frequency Distribution</th>
<th>Velocity</th>
<th>Active region</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Not Active</td>
</tr>
<tr>
<td>$P_i \in M_P$</td>
<td>$\bar{V}_i \in N_V$</td>
<td>Active</td>
</tr>
<tr>
<td>$P_i \notin M_P$</td>
<td>$\bar{V}_i \notin N_V$</td>
<td>Active</td>
</tr>
</tbody>
</table>

Fig. 6. Rescaling the Pixel Frequency Distribution and Objects Velocities.
(a) Maximum values - Pixel Frequency distribution per image frame.
(b) Maximum values - Pixel Frequency distribution per second.
(c) Maximum values - Objects Velocities per second.
The crisp set algorithm works on the pixel frequency distribution and object velocity of a moving object. One has to note for given image plane the pixel distribution will tell where active activity is compared to the still background. Whereas the velocity of an image plane is taken for this case study is based on a fixed time interval (one second). Before passing the inputs to the algorithm, a normalization step has to take into consideration before the crisp set is used to combine the two main attributes. The normalization steps or rescaling procedure will be described next.

C. Rescaling the pixel distribution

The process of evaluating human or any moving object activities requires rescaling the pixel frequency distribution by choosing the maximum value recorded in each second, moreover it requires also matching up the velocity vectors by means of comparing the current recorded velocities and considering the high velocity during each second of recording.

IV. RESULTS ON ACTIVITIES CLASSIFICATION

The algorithm evaluates the rescaled values of the pixel frequency distribution with a term known as normal behavior is set as $N_p$, and evaluates and records similarly for velocity attributes with a normal behavior set as $N_v$. In order to determine the region of interest (catch region of interest), Table II shows and summarizes the processes of classifying the active and non active regions. Again the term $N_p$ and $N_v$ are the normal behavior set for the pixel frequency distribution and velocity respectively.

Constructing the normal behavior sets is a supervised operation relies on the security requirements and always set at priori. The two sets used in this case study are capable of handling any other similar scenarios, while considering the initial parameters like the position of the camera and scene topology.

To trigger the visual alarm, the image plane is segmented into sub zones. The concept of dividing the image plane into group of engrossed sections facilitates the progress of concentrating the effort in a highly sensitive area in the camera view. We implement a based pixel segmenting algorithm to perform this task. Each new coming frame is divided into four zones. After that, we checked the existence of the human in the dedicated zone. Based on the learned activity status, a visual alarm will trigger to assist the security officers to concentrate in the active regions.

Figure shows samples of the visual alarm triggering event. Fig. 7(a) presents the entrance of the objects to the scene. The visual alarm appears as a red square and indicates the active region. As explained before the region’s activity status is determined based on the velocity and pixel distribution level exhibited in the certain region.

Fig. 7(b) shows that region two is witnessing high activity more than the rest of the image regions, this fact appeared
clearly through the visual alarming system, where it detects that the second agents (red bounding box) presents more activity at that time compared to the first agent (blue bounding box). Fig. 7(c) shows that region three is now witnessing high activity more than the rest of the image regions, this fact appeared clearly through the visual alarming system, where still detects that the second agents (red bounding box) presents more activity at that time compared to the first agent (blue bounding box). Fig. 7(d) illustrates the most active region currently is region one and that’s because it contains the activities of the two agents participated in this scenario.

V. CONCLUSION

This paper looks into a form of smart triggering event approach by analyzing the image properties taken from video surveillance camera. The main motivation of the work is due to the necessity to create a semi-automated surveillance system in order to assist the enforcement personals to catch the event of interest from the continuously running real-time scene by observing humans or moving objects. Future work will look into the on how to reduce storage in video surveillance backup system or server using similar principle discuss in this papers. On possible enhancement will be rather than using simple crisp set implementation a fuzzy approach can be extended.

REFERENCES


