INTELLIGENT MOTION DETECTION AND TRACKING SYSTEM

Dr. Hatem Elaydi
Islamic University of Gaza
PO Box 108, Gaza, Palestine
helaydi@mail.iugaza.edu
Phone: +970-8-286-0700 ext. 2870
Fax: +970-8-286-0800

Ismail Al-Wehaidi
Islamic University of Gaza
PO Box 108, Gaza, Palestine

Waseem Shbair
Islamic University of Gaza
PO Box 108, Gaza, Palestine

Wael Al –Madhoun
Islamic University of Gaza
PO Box 108, Gaza, Palestine

Mohamed Hareb
PO Box 108, Gaza, Palestine
Islamic University of Gaza

Abstract

The rapid development in the field of digital image processing made motion detection and tracking an attractive research topic. Until recent years, real-time video applications were inapplicable due to the expense computational time. An intelligent method to analyze the motion in a stream video line using the methods of background subtraction, temporal differencing, and optical flow, methods are proposed. The new method solves the computational time problem by using a reliable technique that is called Fast Pixels Selection. A low cost tracking system is proposed. This tracking system consist of camera, PC, motor and data acquisition card. This system is designed to detect and track any moving target automatically.

Keywords:

Motion, detection, tracking, intelligent, optical flow.

Area:

Image Processing

Abstract
The rapid development in the field of digital image processing made motion detection and tracking an attractive research topic. Until recent years, real-time video applications were inapplicable due to the expense computational time. An intelligent method to analyze the motion in a stream video line using the methods of background subtraction, temporal differencing, and optical flow, methods are proposed. The new method solves the computational time problem by using a reliable technique that is called Fast Pixels Selection. A low cost tracking system is proposed. This tracking system consist of camera, PC, motor and data acquisition card. This system is designed to detect and track any moving target automatically.

1. Introduction

Organizations and institutions need to secure their facilities; thus need to use security systems that are equipped with the latest technology. Intelligent video sensors were developed to support monitoring security systems to detect unexpected movement without human intervention [1, 2]. Most of the old systems depend on the presents of guard with a duty of keeping an eye on everything. This had its draw back due to the human element.

From this point of view, the need of a system that is not based on human, avoiding its fault, does not need a lot of monitors and does not need to be looked closely at all time must be provided. This system can give an alarm when something strange is happening and can take an action when somebody attacks its structure. So this need is raised to the surface of security system.

New techniques were used in these security systems are based on target movement. Detecting of the move, location, speed and any desired information on the target from the captured frames can be taken from the camera and can be transferred to the analysis part of the system [3,4,5]. Motion detection is one of these intelligent systems which detect and track moving targets. Overhead cameras capture images of the securing area workspace; these images are processed to detect the target [6,7].

Many algorithms and techniques have been used to perform this process and improve its outcome [8,10]. Some systems have been developed to track the target after detecting it [9]. Visual tracking system is needed in many applications, including surveillance, robotic gaze control, and missile guidance [8]. This paper uses some old methods that detect a moving target and discusses their benefits and usage, and discusses the tracking process that is based on detecting algorithms. These algorithms are tested using real-time low cost system that can be used in any house or work place.

Historically, target detection has been performed on single images or static imagery [1,2]. More recently, video streams have been used for target detection [3,4]. Many methods like these are inapplicable to real time applications. However, methods such as Pfinder [4], W^4 [5], Beymer [6], and A. Lipton [1] are designed to extract targets in real-time.

A statistical model for the reliability of motion tracking using mean normalized correlation were tested [7]. Local application of optical flow to analyze and determine camera motion or reconstruct 3D scene structure depend on Alan Lipton are examined [1]. Visual behaviors for binocular tracking as A. Bernardino and J. Victor [9] and non-parametric model for background subtraction by Ahmed Elgammal and David Harwood [8] were looked at.

The purpose of visual tracking system is to follow a moving object through a sequence of images. This paper discuss detecting and tracking a moving target using a camera which is connected through a PC constraint on motor to drive it and linked with controller to control its motion, then analyzing the acquired image from video signal by applying suitable algorithms to achieve the best applicable security systems application.

Traditionally, tracking systems require dedicated hardware and software to handle the computational demands and input/output rates imposed by real time video sources. Our tracking system compatible with computer PII 400 MHz platform, with high performance camera 30 frames/sec, yielded by high level language software exactly MATLAB and Visual C++.

The paper is organized as follows: Section 2 gives a brief background on image processing. Section 3 talks about target detection techniques. Section 4 covers tracking with motion approaches. Section 5 talks about motion control. Section 6 describes the system configuration and section 7 displays the results.

2. Computer Vision

Many processes as image compression, restoration, enhancement and analysis are used in dealing with the digital image [7]; the analysis of the image is the most used process in this paper.

Digitizing an analog video signal is to convert the signal to be compatible with computer media to store and restore the signal, this is done by sampling and quantization. The value of the analog signal at each instant is converted pixels, the smallest element of an image as shown in Figure(1). After the process of digitizing is completed, we have 2D of data constructed of pixels, which contain the color information in each position of the image.

I(x, y) = the brightness of image at the point (x, y). Where x= row and y= column.

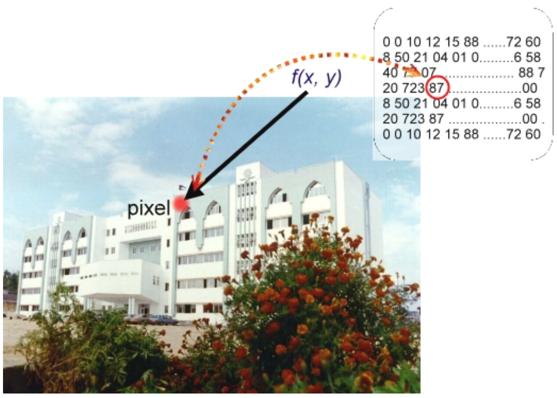

Figure (1): Digital image representation by pixels

Each colored image is stored in three dimensional array MxNx3, number 3 comes from creating three layers of the basics colors, Red, Green and Blue. These layers are composed to represent the true color image. Each colored pixel composed of three values (red, green, blue), these values are having the ranged between 0-255.

3. Motion Detection

This section describes four methods for extracting moving targets from real time video stream. The first approach is *Background Subtraction* which is the simplest but not without disadvantages that the second approach *Temporal Differencing* handle it. Then, a direct improvement has been added to temporal differencing to present a new algorithm called *Fast Pixel Selection*. Finally, an intelligent approach has been presented here that is based on *Optic Flow* method.

3.1 Background subtraction

The main steps of detecting the target system are shown in Figure (2) are:

1. Frames differencing.
2. Detecting the moving target.

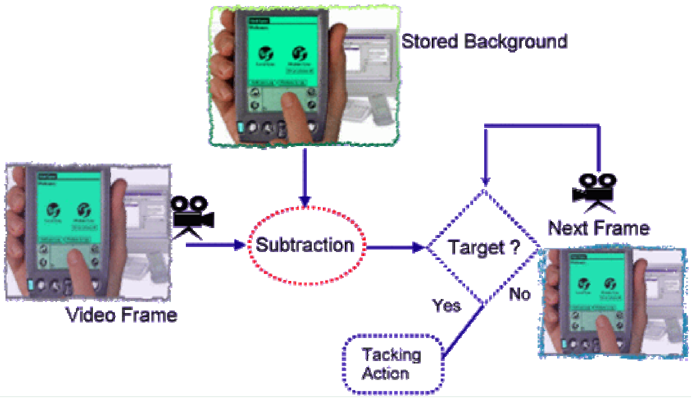

Figure (2): Background subtraction model

Frames differencing: The idea is to take the absolute difference between a reference stored background and a new frame which is grabbed from the camera [1]. If no target is entered, the result pixel is expected to be zero value (difference result), else, if a target exists, the resulting pixels will give a value different than zero as shown in Figure (3).

$$D = | I_{K-1} - I_K |$$

K: The frame number, I: frames, D: difference result.

The Background Frame	**The Grabbed Frame**	**The Output Of Frame Differencing**
		

Figure (3): Frame Differencing

3.2.2 Detecting the moving target

After taking the difference, the number of nonzero value pixels will specify the size of the target; if these pixels met the requirement of the system an action will be taken to alarm, or lights, etc.

If the number of the nonzero pixels did not approach the minimum number of the required pixels to decide, another frame is taken and goes on to the difference process again.

3.2.4 Disadvantages

This is a fast algorithm and compatible with real time video analysis, but, many problems appears in using the still background, these can be summarized as follows:

1. Sensitivity to lights, where any lighting condition will cause the system to recognize it as a moving target.
2. Sensitivity to clutter, and image texture, the same thing will be done by the system in error decision.

3.3 Temporal Differencing

The temporal differencing is the next algorithm which will solve the still background problems. The basic principle of this method depends on extracting (two-frames or three-frames), and taking the first frame as background and the subtraction is applied between it and between the next two frames. So we have an updating background as shown in Figure (4).

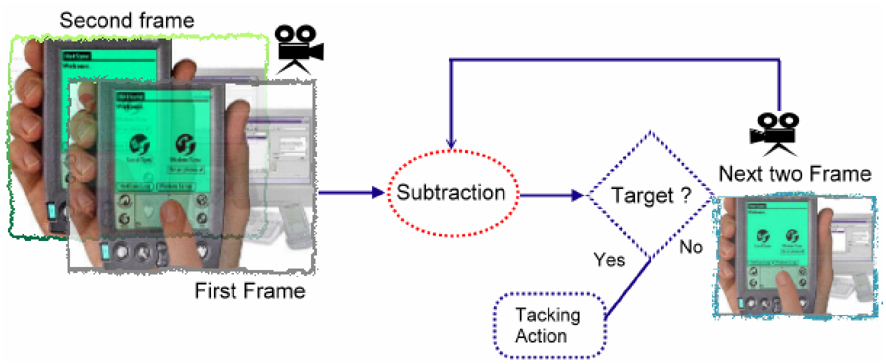

Figure (4): Temporal Differencing Model

3.3.1 Motion Analysis

Motion analysis using temporal difference is simply done by taking a consecutive video frames and determines the absolute difference. A threshold value is then chosen to determine the change.

$$M_n(i,j) = \begin{cases} I_n(i,j) & , D_n(i,j) \geq T \\ 0 & , D_n(i,j) < T \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.1)}$$

Where T is the threshold, where the difference between the pixels of the consecutive frames will be little more than zero because the capturing quality of the camera.

3.3.2 Lighting Error

If the light changed through the motion detection process an error decision will occur.

Here the benefits of temporal method appears, the third frame will be compared too if it contains a target then the decision will be taken, else it will be classified as an error in lighting or something.

3.3.3 Disadvantages

Time consuming of temporal method cause that the method to be inapplicable with security system. Where the time of differencing two frames and taken the decision and possibility of using a third frame takes a lot of time. So the search goes on to find a better algorithm to enhance the working in still camera motion detection.

3.4 Fast Pixel Selection

If the frames were of big dimensions in a normal speed of measuring the difference between the consecutive frames to detect and there is a target in the zone, it will take a lot of time that may permit the target to pass the secure area without taking its frame because the system was busy in computing the difference in the previous frames as shown in Figure (5). So the need for more a faster computing algorithm is required in order to recover the missing frames during that time.

The approach here depends on: the number of the white pixels, which are the output of the difference process where they decide if there is a target or not. So instead on taking the difference of all the pixels in the consecutive frames, it would be faster to take a selection of the pixels and reduce the decision number of pixels.

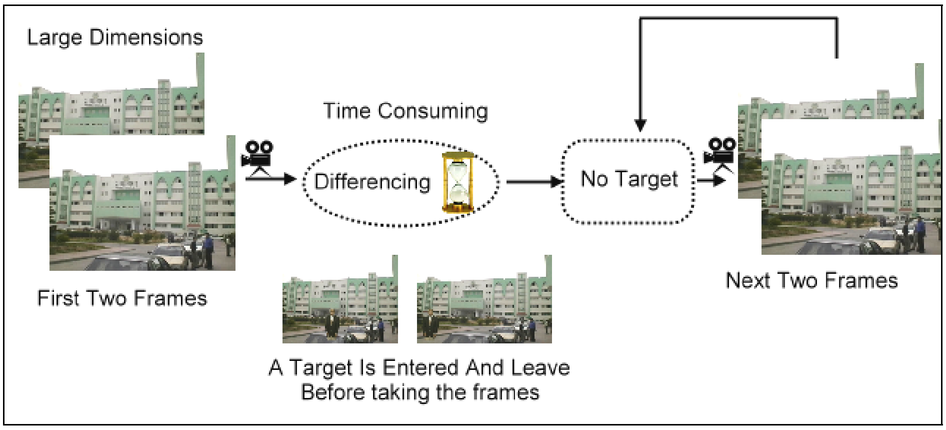

Figure (5): Time consuming problem overview

Since no target takes the size of one pixel or even ten pixels the pixel differencing can be spanned, such that the pixels in each row can be taken after ignoring a number of them and the same thing can be done with the columns.

The pixel selection can be made by taking not the adjacent rows or columns of the pixels in the frames this will speed the process of differencing to four times the normal differencing.

3.5 Optical Flow

This method based on the differential methods and the region-based matching methods [17]. The differential methods effectively track intensity gradients. The advantage of this is that flow vectors can be determined based on the information (measured by intensity gradient).

With reliable flow vectors for every pixel in an object it becomes possible to track individual pixels from frame to frame. This capability can be employed to cluster of pixels into “body parts” for model-based motion analysis. It also means that an object’s rigidity can be determined by calculating residual flow that is the motion of “body parts” relative to the target gross motion [17].

Detection of bodies in a video stream is performed by real-time human motion analysis. This is basically a process of background subtraction using a dynamic updating background model. First, each frame is smoothed with a (3X3) Gaussian filter to remove video noise (as lights effect or tree vibration). Then the differencing method is applied. After this, non-moving pixels are updated using the Gaussian filter to reflect changes in the scene (such as lighting).

The idea is to use edge gradients as a measure of information and, at every pixel position, grow the support region until there is enough edge gradient information to justify matching. Furthermore, flow needs to be computed only for pixels contained within the moving object. Consequently, this particular implementation is only valid for video streams derived from stationary cameras, or streams which have been stabilized [17].

The Region Matching Algorithm

First, a region $W(x)*I_n(x)$ is defined by multiplying I_n with a 2-D windowing function W(x) of size (Wi;Wj) centered on pixel x, which is the position of the object. This region is then convolved with the next image in the stream I_{n+1} to produce a correlation surface D(x; d).

$$D(x;d) = \sum_{i=1}^{i=w_i} \sum_{j=1}^{j=w_j} \frac{|W(i,j)I_n(i,j) - I_{n+1}((i,j)+d)|}{\|W(x)\|} \quad \text{.......................} \ (3.2)$$

Where ||W(x)|| is normalization constant given by

$$\|W(i;j)\| = \sum_{i=1}^{i=w_i} \sum_{j=1}^{j=w_j} W(i,j) \quad \text{.......................} \ (3.3)$$

To measure horizontal and vertical information, a –Sobel- filter is used. Prior to the matching process, two images $S_H(x)$ and $S_V(x)$ are computed from I_n by filtering using the standard Sobel operators.

4. Tracking Through Motion

After going over different motion detection algorithms, now we will discuss the tracking algorithms which depend on previous methods discussed in Section 3. Three approaches to track moving objects are discussed as follows:

4.1 Tracking using Background Subtraction

This algorithm does efficient work with still camera because the time between updating backgrounds is still large. So, in a tracking system with moving camera this algorithm will be useless. Tracking the target can be done just by drawing boundary boxes around the target as long as the target appears in the frames as shown in Figure (6).

Figure (6): surround the detected target

4.2 Tracking by Temporal Difference

Detecting moving objects in a video stream by temporal difference algorithm, works well with tracking system. This method works well with moving camera tracking system.

The approach is based on dividing the grabbed frame into four equal parts vertically where the division is used to determine the target location in these quarters. And after detecting the target location the camera may move to center the target.

4.3 Tracking by Optic Flow

Optic flow does more efficient work for tracking system, because this algorithm manipulates many critical points in the tracking process. This point will be discussed later as tracking through occlusion.

The important feature of this tracker is that, unlike traditional template tracking which has an aim to "drift", this method only matches pixels which have been identified as moving by the motion detection stage thus making it more robust.

Here, the regions are dynamic and are grown until there is enough information to ensure a reliable match. The flow computation per pixel is a two-pass process. Region matching is performed to compute the flow vector (track). Motion of single pixels can be modeled as the collected motion of the object $V_B(x)$ plus some residual pixel motion V_R (x). Thus

$$V(x) = V_B (x) + V_R (x) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.3)$$

Where $V(x) = (V_i(x), V_j(x))$, the velocity of the pixel in 2-D image "flow field".

It is assumed that the residual motion of pixels within Z_n is not very great, then the correlation surface "distance function" D(x; d) to compute the flow vector for a given pixel x need only to be evaluated over a small range around $d = V_R\, \delta t$.

Using the 2-D windowing function W(x) as determined before the flow V(x) for pixel x can be computed. One limitation of this method is that background texture can corrupt the flow vector calculation. To improve this, the elements of W(x) can be weighted to give preference to "moving" pixels (expressed as a binary image M_n).

5. Motion Control

Motion control is the process that control movement objects by computer. Automatic camera working as security system can track a moving object to capture its image. A movement control mechanism has to be realized to adjust the positions and angels of camera.

When speed becomes a critical point, it is important to have a high performance real-time computer in the system. This kind of computer system will acquire information from a working process; manipulate this information, then take action by returning a set of decisions.

Figure (7) illustrates the structure of control functions and I/O interfaces to be used. For different control instructions, the system will select suitable control modes to correct the motion of the servo motors in the working process.

The speed of motor to respond to an action depends on the delay due to physical movement and the mechanical components to settle down and computer response time. Another factor that will speed up the process is the size of the image.

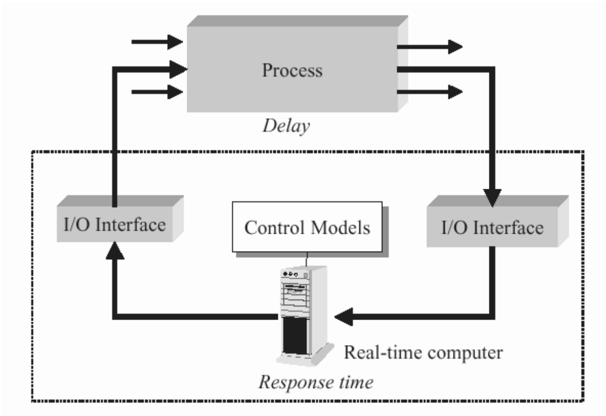

Figure (7): System Structure Overview

6. System Configurations

A low cost system is developed and is based on a camera which is connected to a PC in order to detect unwanted or unnatural targets in certain places with sensitive nature or high security nature. When there is a target in camera Filed Of View (FOV) the detecting process begins. This system as any other security system depends on hardware and software component.

Color Camera which is the main device needed to capture frames of a zone. Some specifications such as auto-focus, 30 fames/second and high quality (800 X 600 pixel resolutions) are required. **T. V. Card** which is the intermediate component between the camera and PC. **Motion Control** has the capability to control a camera manually to the right and left, and automatically for tracking targets. It can send alarm signal when the system detects a target or a system hacking. **Monitor** is needed to display the results.

This collection of compound devices must have some tools to get them connected and interfaced together thus the need for programming languages as software, to make these components alive.

The software used here is MATLAB 6.0, Data Acquisition Toolbox 2.0 and Image Processing Toolbox.

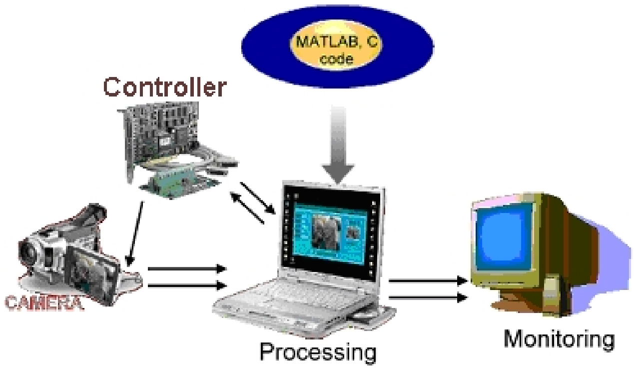

Figure (8): System Description

7. Results

This system has been implemented and tested on PII 400MHz under Microsoft Windows 98 & XP with a 32bit ATI T.V. Card digitizer. The system can detect and track any moving target at 30 frame/second over 160 x 120 pixel image. The configuration was built by connecting the computer with 12MHz Intel 80C51 Microcontroller through a data acquisition card.

7.1 Detection

The small change in image produce the difference between the frames with nonzero pixels as shown in Figure (9). A threshold value was applied based on the system requirement which eliminates the noise effects.

If a trees are blowing in the wind they will be correctly rejected and classified as background clutter.

7.2 Tracking

The tracking system allows a continuous tracking through a moving camera. Because of using 160 x 120 pixel frames the system where the computational time is about 4 seconds. Using fast pixel selection, the frame is reduced to 5 pixels in each column and row spanning. This

First Frame Second Frame The Result

Figure (9): Capturing cause error in difference process.

reduces the computational time by a magnitude of 20, thus reducing the computational time to takes less than a second. This high speed is needed to determine the direction of the target to move the camera toward the target direction before going out of the camera's FOV.

The rotating order that is given to the motor, the response time, rotating time, and stability time are taken into consideration and were measured to stop the capturing process while the camera is moving. It gives an error decision if the capturing is done and the camera is rotated.

8. Conclusion and Future Work

The rapid growth in computing and digital image processing made motion detection and tracking an attractive research topic. Until recently, real-time video applications were inapplicable due to computational time. A new low cost intelligent system was presented based on fast pixel selection and optic flow. The results shown the system was capable of detecting and tracking any moving target. Future work can be concentrate on classifying the target.

References

[1] A. Lipton, H. Fujiyoshi, R. Patil, the Robotic Institute, http:// www.cs.cmu.com, 1998.

[2] R. Kasturi and R. C. Jain, Computer Vision Principles, Computer Society Press, 1991.

[3] D. Koller, K. Danilidis, Model-Based Object Tracking in Monocular Image Sequences of Road Track Scenes, International Journal of Computer Vision, 1993.

[4] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland , *Pfinder: Real-Time Tracking of the Human Body*, IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997.

[5] I. Haritaoglu, D. Harwood, L.S. Davis, *W^4 Who? When? Where? What?* A Real Time System for Detecting and Tracking People, 1998.

[6] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik, *A real-time computer vision system for measuring track parameters*, Proceedings of IEEE, 1997.

[7] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Addison Wesley, 1992.

[8] A. Elgammal and D. Harwood, On-parametric Model for Background Subtraction, Computer vision Laboratory, University of Maryland, College Park, 1997.

[9] A. Bernardino and J. Santos-Victor, Visual Behaviors for Binocular Tracking, Institute Superior, 1096 Lisboa Codex, http://www.isr.ist.ut1.pt.

[10] W. Bell, P. Felzenszwalb, D. Huttenlocher, *detection and long term tracking of moving objects in aerial video*, http://www.cs.cornell.edu, March 26th, 1999.

[11] F. Dellaert and R. Collins, Fast Image-Based Tracking by Selective Pixel Integration, http://vasc.ri.cmu.edu, 1999.

[12] C. Nelson and R. Polana, Motion Detection and Recognition Research, http://www.cs.rochester.edu, 1998.

[13] Anti Katridge, *Vision for Matlab*, http://www.sys.uea.ac.uk/, 1997.

[14] Brad Johanson, Rachel Kolodny, Daniel Russel, *A Hand Pose and Position Tracker*, http:// www. graphics.stanford.edu, 1999.

[15] Marcus Breig and Markus Kohler, Development of a Motion Detection and Prediction for Human Motion, http://www.cs.uni-dortmund.de, 2000.

[16] Lijun Ding, Thisath Kularatna, Ardeshir Goshtasby, *Volumetric image registration by template matching,* http:// www.cs.wright.edu, Wright state University, Dayton.

[17] A. J. Lipton, Local Application of Optic Flow to Analyze versus Non-Rigid Motion, the robotic institute, http://www.cs.cmu.edu/~vsam.

[18] J. Oakley and W. Powell, *Statistical Model for Reliability of Motion Tracking*, University of Manchester, Manchester M13 9PL, 1998.