Security Enhancement by using Video Surveillance System

Prof. V. K. Sambhe

Vaibhav N. Chunekar

Computer Department ,Veermata Jijabai Technological Institute
Matunga, Mumbai-400019
vksambhe@vjti.org.in, chunekar@rediffmail.com

Abstract

Video surveillance system, also called Visual monitoring and security system has been a key component in ensuring security at airports, bands, casinos and correctional institution and also used for military application and wild life preserve management. In this paper we have implemented a method for human detection by using edge detection [3] and segmentation methods[9][3] and human tracking[4][3] is done by using feature tracking method. Kanade-Lucas-Tomasi[10] proposed an effective algorithm for human tracking.Using this method a significant performance can be obtained in a surveillance system.In this paper he proposes an algorithm and stepwise result. We implemented this algorithm in C. We assume that the motion of a human is normal (approx. 7km/hr).According to our implementation and results we found that our system is more accurate and efficient as compared to the previous work.

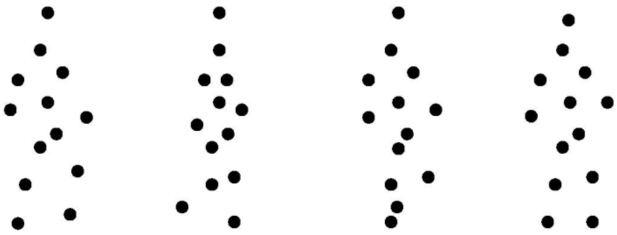
Keywords: Video Surveillance, Edge detection, Segmentation, Feature tracking.

1. INTRODUCTION:

Video Surveillance system is also referred to as visual monitoring and security system. With the increasing population there has been a proportionate growth in the number of public places. It is, thus, essential to provide a secured and safe environment at such places. A proper security mechanism must be in place so as to prevent possible terror attacks, thefts or other abuses. More recently government agencies, schools and businesses are turning towards visual monitoring and security systems as a mean to increase public security. These systems are used for ensuring security and law enforcement. Visual monitoring and security system has been a key component in ensuring security at airports, bands, casinos and correctional institution and also used for military application and wild life preserve management. The Video Surveillance System tracks individual pedestrians as they pass through the field of vision of the camera, and uses vision algorithms to classify the motion and activities of each pedestrian. With this information, the system can bring the incident to the attention of human security personnel. The system requires people to be tracked. Information about their behaviors can be obtained from their trajectories and interaction between them.

Detection of a human based only on motion may, at first, seem far-fetched. Do the motion of the limbs contain enough information to infer the presence of a human? Experiments performed by Johansson[4] in the 1970's demonstrated the answer to be 'Yes'. Johansson filmed moving humans in a pitch-black room, the only visual indicator being a white point of light attached to each limb. He showed that a viewer watching the film could easily identify human motion, despite the absence of visual cues such as shape, texture, brightness, and color. An example of these Johansson points is shown in the Figure below. Given that the human brain can effortlessly recognize this motion, it is conceivable that a computer algorithm could do the same.

In addition, single points of motion as used in the Johansson experiment can be efficiently represented on a computer. Unlike pure image processing, which must deal with large numbers of pixels at each time step, this Johansson motion can be specified by a handful of points, each represented by a 2-D position and a 2-D velocity at any given time. This gives us hope that a simple, effective algorithm is achievable.

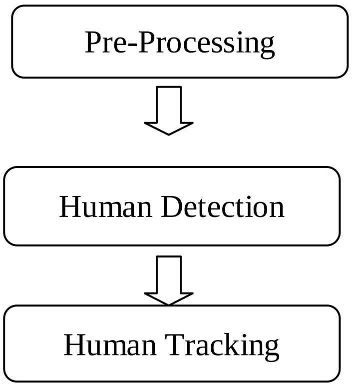


An example sequence of Johansson points above show the side view of a walking human. Taken one image at a time, the shape of the human figure is not completely apparent. However, when considered in sequence, the images clearly depict a human.

The Johansson experiment shows us that human detection from the motion of point features is a realistic goal. With this premise in mind, we can divide the system into several sub problems:

1. Distilling full-frame video into a set of individual point features.
2. Finding a model that accurately represents the human motion.
3. Apply that model to a detector that can determine if a cluster of moving points is representative of human motion.

2. SYSTEM ARCHITECTURE :



1) Pre-processing:

The pre-processing stage consists of two sub-stages:

a) Low pass filtering:

We have applied Low pass filtering on the stream of input images which involves in convolving the image by special masks specified.Low pass filter is used for removing high frequency component and it will allow only low frequency component. It will remove the noise from image by blurring it.

Ideal low pass filter has the transfer function as below:

$$H(u,v) = 1 \text{ if } D(u,v) <= D0$$
$$= 0 \text{ if } D(u,v) > D0$$

b) Edge detection:

Edge detection is the most common approach for detecting meaningful discontinuities in gray level. It is required to detect the motion of humans. Edge detection is done by making use of two techniques-sobel operators and variance method.

The sobel technique which we have implemented on filtered images does not have very clear output. This problem is solved in variance technique by taking square root of summation of difference of pixels.

c) Sobel Edge Detection:

It is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function.

First derivatives in sobel technique are implemented using magnitude of gradient. For a function f(x,y) magnitude of gradient is given by

∇ $\nabla f = |Gx| + |Gy|$

The masks shown below are called sobel operators.

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

The x-coordinate is here defined as increasing in the "right"-direction, and the y-coordinate is defined as increasing in the "down"-direction. At each point in the image, the resulting gradient approximations can be combined to give the gradient magnitude, using:

$$\mathbf{G} = \sqrt{\mathbf{G}_x{}^2 + \mathbf{G}_y{}^2}$$

Using this information, we can also calculate the gradient's direction:

$$\boldsymbol{\Theta} = \boldsymbol{arctan}\left(\frac{\boldsymbol{Gx}}{\boldsymbol{Gy}}\right)$$

where Θ is 0 for a vertical edge which is darker on the left sides.

The idea behind using weight two is to achieve some smoothing by giving more important to the center points. From these masks Gx and Gy are:

Gx= (z7+2z8+z9)-(z1+2z2+z3)

Gy= (z3+2z6+z9)-(z1+2z4+z7)

d) Variance Edge Detection:

It is another technique that we have applied for the edge detection. It gives better results than sobel edge detection techniques. It is not using operators like sobel. It processes each pixel in the image.

It first takes a pixel than makes an nxn window for the processing of that pixel. The output of the variance is depending on the value of n. After processing it will replace the value of that pixel by new value.Algorithm is as follows[5]:

```
Take an image of  L x M size.
For each pixel f(x,y) in image
{
    Take an nxn window
    Take the difference of f (x,y) and other pixel in image.
    Sum the square of the difference
    Square root of sum(S)
If ( S > 255) then S = 255
        Set new value of  S to f(x,y)
}
```

1 Human Detection:

The outputs of preprocessing are fed to the human detection techniques. We have implemented two techniques for human detection. Background subtraction is a very simple approach for detecting humans which take the difference of two consecutive frames. It gives erroneous outputs in the cases of environmental changes or dynamic background.

a) Background Subtraction:

The goal in background subtraction is to separate background areas of the image from foreground regions of motion that are of interest for human tracking. It generates a foreground mask for every frame. This step is simply performed by subtracting the background image from the current frame.

$$O(x, y) = |C(x, y) - B(x, y)|$$

b) Segmentation:

We use segmentation, thresholding followed by morphological operation to solve that problem in th background subtraction.

One of the techniques of segmentation is Clustering. The K-means algorithm is an iterative technique that is used to partition an image into K clusters. The basic algorithm is:

- Pick K cluster centers, either randomly or based on some heuristic.
- Assign each pixel in the image to the cluster that minimizes the variance between the pixel and the cluster center.
- Re-compute the cluster centers by averaging all of the pixels in the Nayan Agrawal, Priyanka Agrawal, Manisha Dandge, Vibhuti Thakkar, Akash Wakodkarcluster
- Repeat steps 2 and 3 until convergence is attained (e.g. no pixels change clusters).

Histogram based segmentation finds foreground and background peaks and using those peaks it calculates threshold values.

- First Compute the histogram H of the image..
- Histogram Smoothing.
- Calculate the peak values in the histogram.
- Calculate two threshold points from computed peak values.
- Threshold the image using threshold points.

After finding threshold points from the image (High and Low) , we will perform threshold operation on the image. If the pixel value is between High and Low then it will set as a white pixel otherwise set it to 0 which the value for black. The algorithm is shown below[4].

High and Low Threshold Value

```
For each pixel f (x, y) in the image
If ( f(x,y) > Low and f(x,y) < High )
        Set the value to 255 (white pixel)
Else
Set the value to 0 (Black pixel)
```

2 Morphological Operation:

Morphological operators often take a binary image and a structuring element as input and combine them using a set operator (intersection, union, inclusion, complement). They process objects in the input image based on characteristics of its shape, which are encoded in the structuring element.

a) Dilation:

Dilation is one of the two basic operators in the area of mathematical morphology, the other being erosion

The dilation operator takes two pieces of data as inputs. The first is the image which is to be dilated. The second is a (usually small) set of coordinate points known as a structuring element (also known as a kernel). It is this structuring element that determines the precise effect of the dilation on the input image.

Dilation is used for bridging gaps. A and B are two sets in Z2. The dilation A by B is defined as[1].

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \emptyset\}$$

This equation is based on obtaining the reflection of B about its origin and shifting this reflection by z. Set B is commonly referred as structuring element in dilation.

b) Erosion :

Erosion is the dual of dilation, i.e. eroding foreground pixels is equivalent to dilating the background pixels[1].

$$A \ominus B = \{z \mid (B)_z \subseteq A\}$$

Thus erosion of A by B is the set of all points z such that B translated by z is contained in A. Erosion is used for eliminating irrelevant detail in an image.

1. Human Tracking:

Human tracking involves detecting the human motions and their activity. One technique tracks the motion by finding the objects in the image sequence. But the windows which are created on the objects are not identical when two persons overlap.

The simple tracking algorithm is as follows [6].

- Take the images from the background subtraction.
- For each image in the video
- Find out the two points for each objects
- To find out points use manual search and sorting
- Draw the windows around objects

This technique has some problems. When two objects are coming into view area then they are assigned with one window to each. For example person A has window of x color and person B has a window of y color. When they both are cross each other then at the time of overlapping two both the objects come in only one window. So we have one window for two objects. This is not the expected output.

The second technique involves finding feature points from the image. The method for identifying points of interest in a series of images and tracking their motion is called feature point selection method.

Take an image I (x , y)

$$Gx = \text{Convolve image I with} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$Gy = \text{Convolve image I with} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

Gxy = Gx * Gy
Gxs = Gx * Gx
Gxy = Gy * Gy

$$A = \text{Convolve Gxs with} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$B = \text{Convolve Gys with} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$C = \text{Convolve Gxy with} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

T = A + B
D = AB – C2
Q = 4D / T2
If Q (x , y) > Qmin
W = D / T
else
W = 0 (set it to black)

If the value is greater than Qmin then set the value to D/T other wise make it black. So the pixels which have the maximum value on the output image will be best tracking point. Here we will find the maximum value from the region of the object.

Then we apply Knade–Lucas–Tomasi filter to track those feature points[10].

- Choose a small window; say 7 pixels on a side, around a pixel of interest in frame 1. This pixel of interest will be called pixel A.

- For each pixel near A in frame 2, call it pixel B, and perform the following:
- Subtract the value of each pixel in the 7 by 7 region around pixel A from each pixel in the 7 by 7 region around pixel B. Square the result of the difference, and sum these 49 values to produce'dissimilarity' for this choice of pixel B.
- The pixel B in frame 2 with the smallest dissimilarity is considered to be the new location of pixel A in frame 1.

3. Results:

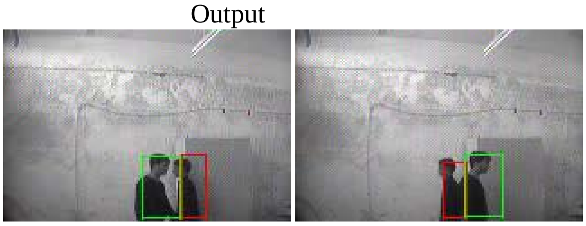
Sample output of the system are shown below.We can see human detection and tracking in the output.

Input

Output

Input

Output



Input Output

4. Conclusions and Future Work:

We have successfully developed a system for detecting human and their motions. Thus, any moving object that has roughly the shape of a human and moves with the speed expected of a human will be detected as a human. The outputs of tracking algorithm are used for the real time video surveillance system. The system presented here serves as a successful proof of concept for a robust human motion detector in the field.

There are a variety of enhancements that could be made to this system to achieve greater detection accuracy and increased robustness.

1 Objects could be tracked between frames rather than simply performing human motion detection on single frames.

2 The current segmentation algorithm can be confused by too much fast lighting change moving shadows.

3 From the current system, client-server model can be created ,in which client have access to view the tracked videos and images.

Acknowledgements:

We take this opportunity thanks to Nayan Agrawal, Priyanka Agrawal, Manisha Dandge, Vibhuti Thakkar, Akash Wakodkarur, B.Tech student for their valuable contribution.

References:

1 R.C.Gonzalez and R.E.Woods, "Digital Image Processing Second Edition"

2 R. Border, B. Jackson and N. Papanikolopoulos, "Vision-Based Human Tracking and Activity Recognition," AIRVL, Dept. of Computer Science and Engineering, University of Minnesota.

3 D. Moore, "A Real-World System for Human Motion Detection and Tracking," California Institute of Technology. dcm@acm.org June 5, 2003

4 R. T. Collins, A. J. Lipton and T. Kanade, "A System for Video Surveillance and Monitoring," Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. The Sarnoff Corporation, Princeton, NJ, Robotics Institute, Carnegie Mellon University Technical Report, (CMU-RI-TR-00-12), May 2000.

5 Y. Sheikh, Y. Zhai, K. Shafique and M. Shah, “Visual Monitoring of Railroad Grade Crossing,” University of Central Florida, Orlando FL-32816, USA.

6 C. Jaynes, S. Webb, R. M. Steele and Q. Xiong, “An Open Development Environment for Evaluation of Video Surveillance Systems,” Metaverse Lab, Dept. of Computer Science University of Kentucky, Lexington, KY 40513.

7 I. O. Sebe, J. Hu, S. You and U. Neumann, “3D Video Surveillance with Augmented Virtual Environments,” Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089-0781. IWVS’03, November 7, 2003, Berkeley, California, USA.

8 Raju Rangaswami, Zoran Dimitrijevi´c, Kyle Kakligian, Edward Chang, Yuan-Fang Wang *f*raju@cs, zoran@cs, smallart@cs, echang@ece, yfwang@csg.ucsb.edu University of California, Santa Barbara.

9 Efficient Moving Object Segmentation Algorithm Using Background Registration Technique Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen, *Fellow, IEEE. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 12, NO. 7, JULY 2002*