

VIDEO SURVEILLANCE BASED TRAFFIC MONITORING SYSTEM

¹Mr. Rahane Wasudeo P., ²H K Sawant

^{1,2}Department of Information Technology,
Bharati Vidyapeeth Deemed University
College Of Engineering, Pune-46

ABSTARACT :The ability to reliably detect and track moving objects is a challenging task. Interacting with moving bodies and understanding their activities are at the core of many problems in intelligent systems. Some examples of its applications can be: Automated surveillance of venues like airports, railway stations, highways. The software will monitor security cameras and detect suspicious behavior. Moreover, human operators can look for activities that they specify without requiring manual viewing of each sequence. Having automated surveillance increases the coverage area of surveillance, Autonomous robots can interact more effectively with the humans if they can effectively detect the presence, The system on a camera-equipped car can detect obstacles or pedestrians and warn the driver to avoid accidents, It can be used for producing computer-generated images of realistic motion which is done currently by using a motion-capture system that stores 2-D or 3-D dimension of a human body using sensors

1. INTRODUCTION: Visual analysis is concerned with detection and analysis of a target object in a sequential stream of images. While many algorithms are successful to detect objects effectively in controlled environments, they get bugged with the variations of the object's appearance or with the surrounding change. It happens as the existing algorithms employ fixed appearance surroundings for the objects. Such systems are trained using appearance of the data available before the surveillance begins, which limits the range of modeling, and ignores a large area of information (eg: shape change or lighting conditions) which becomes dynamically effective during surveillance.

traffic-video- surveillance system that incrementally learns a subspace representation, with accordance to the changes in the appearance of the target. The proposed system is decomposed into two independent sub-problems. The first problem is to detect foreground objects on a frame-by-frame basis. It is done by labeling each concerned pixel in the image frame as object foreground or object background. The second is the coupling of object observations at different points in an inter-related sequence to extract the moving object's trajectory.

The project is concerned with the task of traffic video surveillance. For a sequence of image sequence from a stationary camera, it is desired to detect and track foreground objects moving through the location as shown in Figure 1.

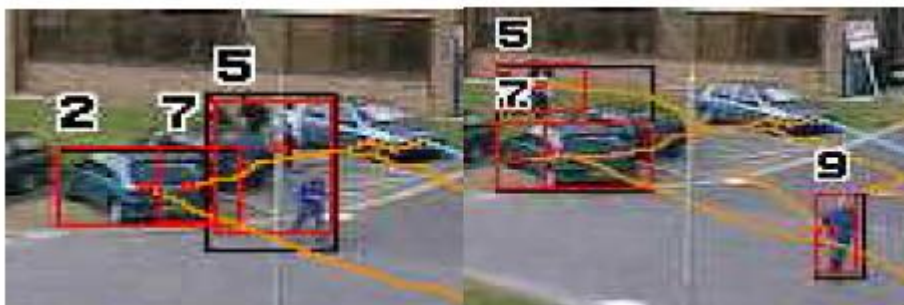


Figure 1: Object detection by video surveillance.

2. ANALYZING PROBLEM FOR RECOGNITION

It is important to adapt the appearance model online, while tracking, to reflect these changes. The appearance model we have chosen, an Eigen basis, is typically learned offline from a set

Of training images $\{I_1 \dots I_n\}$, by taking computing the eigenvectors U of the sample Covariance matrix

$$\frac{1}{n-1} \sum_{i=1}^n (I_i - \bar{I})(I_i - \bar{I})^T,$$

$$\bar{I} = \frac{1}{n} \sum_{i=1}^n I_i$$

Where

Is the sample mean of the training images. Equivalently one can obtain U by computing the singular value decomposition

$$U\Sigma V^T$$

Of the centered data matrix

$$[(I_1 - \bar{I}) \dots (I_n - \bar{I})],$$

with columns equal to the respective training images minus their mean.

Two of the vertices which are connected by any edge are considered as probabilistically dependent on one-another, and those not connected can be considered totally independent. Most important to conclude is that its efficiency at runtime is $O(MN^3)$, where M is the number of triangles in the graph and N is the number of point features we need to evaluate. Thus, for efficiency reasons, it is in our best of work to keep the number of point features in each evaluation to a minimum extend.

So, to reduce this number of point features for each model evaluation may require the supervision of intelligence to choose them. It would be not possible to exactly find the point motion into the entire sequence, and provide all resulting features to the project model in a single evaluation.

Firstly, the number of features will be huge enough, giving regressive performance. Secondly, it may be supplied with a large number of extreme point features, like those which are part of such background. In order to avoid wasting time on the background, we can use image segmentation to discriminate foreground sections of the scene from the background. Image segmentation is the action of any algorithm that separates regions of an image so that they resembles the set of human nature of perceiving. While we are interested in motion, a natural approach is the fragmentation of those parts of the image that are moving relative to the background. This process is called background subtraction.

3. VIDEO SURVEILLANCE BASED TRAFFIC MONITORING SYSTEM MODEL DETECTION

The given set of moving points, each with a position and velocity, our goal video model detection is to decide if the set of points is representative of a predefined model. In this project, we use the approach taken by Y. Song of storing the model as a probability density function, and finding a labeling for the data that maximizes the probability [4]. We will now Précis that approach.

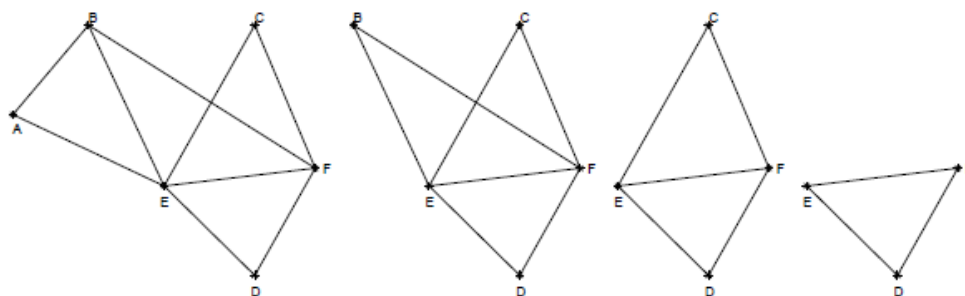


Figure 2: An example of a decomposable triangulated chart with exclusion order A,B,C,D,E, F

We first shown a set of body parts, each conforming to a point feature that could be Tracked. Let $S_{body} = \{LW, LE, LS, H, \dots, RF\}$ be the set of body parts, where LW is the Left wrist, H is the head; RF is the right foot, etc. We will also call these as the set of possible marks. Since the model for the human is learned without supervision, the actual mapping of video parts to point features is undetermined and does not correspond exactly to individual video trafficking. Each video part also has a vector of observed measurements consisting a position and velocity, which we will denote as XLW, XLE, etc. This model of signal is stored as a probability mass function P, which can be appraised for a given usual of observed data:

$$PS_{body}(XLW, XLE, XLS, XH, \dots, XRF).$$

Just, consider a vector of N observed points $X = [X_1, X_2, \dots, X_N]$ onto which the correspondence between points and labels is not known before the time. We need to find the permutation of points that maximizes the probability density function. Put different way, we need to do some labeling $L = [L_1, L_2, \dots, L_N]$ where $L_i \in S_{body}$ is the label of point X_i , that maximizes the probability density function. We will show this optimal marking as

$$\bar{L}^* = \arg \max_{\bar{L} \in \mathcal{L}} P(\bar{L} | \bar{X})$$

Where $P(L|X)$ is the provisional probability of the observation X given the marking L , and L is the set of all transformations of the labeling. the motion model is stowed as a graph with each vertex showing a point feature, and each edge representing statistical necessity of one feature over another. A obliging hypothesis is to make this graph to have the special form of a decomposable triangulated graph. Since, the graph will have one vertex that is dependent on only two other vertices. When this vertex is removed, there will be another vertex that is dependent on only two others, and consequently on. The sequence of vertex eliminations that conserves this property is known as the abolition order of the graph. Figure shows a simple example of such a graph and its elimination order. Thus, a probability density function stored in a decomposable triangulated graph can be approached as a product of independent conditional density functions. For example, if we take elimination order of the vertices of a graph are A,B,C,D,E, F. then the probability density function can be represented as:
 $P(A,B,C,D,E, F) = P(A|B,E) P(B|E, F) P(C|E, F) P(D,E, F)$.

Placing aside for the instant the problematic of the sample mean, suppose we have a $d \times n$ Data matrix $A = \{I_1, \dots, I_n\}$ where each column I_i is an remark, for which we have already computed the singular value breakdown $A = U_V >$. When a $d \times m$ matrix B of new observations is available, the goal is to professionally compute the SVD of the concatenation of A and B

$$\begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} U & \tilde{B} \end{bmatrix} \begin{bmatrix} \Sigma & U^T B \\ 0 & \tilde{B}^T B \end{bmatrix} \begin{bmatrix} V^T & 0 \\ 0 & I \end{bmatrix}$$

Let

$$R = \begin{bmatrix} \Sigma & U^T B \\ 0 & \tilde{B}^T B \end{bmatrix}$$

Which is a square matrix of size $k + m$, where k is taken as the number of singular values in Σ . The time needed to compute the SVD of R ,

R does not depend on n . Now the SVD of $[A B]$ can be expressed as:

$$\begin{bmatrix} A & B \end{bmatrix} = \left(\begin{bmatrix} U & \tilde{B} \end{bmatrix} \tilde{U} \right) \tilde{\Sigma} \left(\tilde{V}^T \begin{bmatrix} V^T & 0 \\ 0 & I \end{bmatrix} \right)$$

To deal with this complexity, we present variables to model the connectivity among sticks, and the (unnoticed) positions of switch endpoints and linkages in each frame. Each stick consumes binary endpoints, both of which are prearranged with just one vertex. Every vertex can resemble to one or other stick. We will let k_i stipulate the coordinates of endpoint, i absolute to the local coordinate arrangement of the stick.

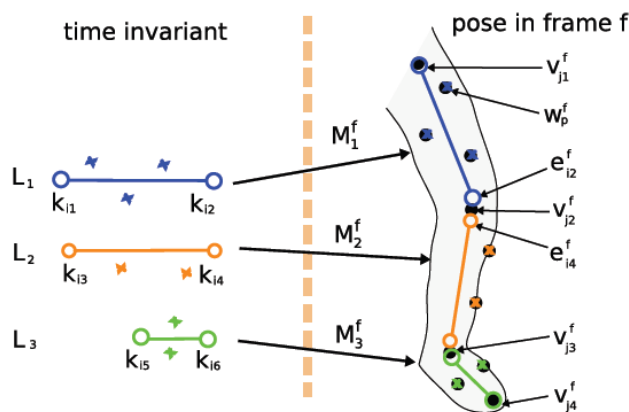


Figure3: The propagative procedure for the experiential feature places, and the credited sites of the stick endpoints.

For each stick, the comparative sites of its feature points and endpoints remain signified in a time-invariant native synchronize system (left). For each frame in the arrangement (right), gesture variables attempt to appropriate the experiential feature positions by plotting local coordinates to domain coordinates, though upholding physical unity by plotting stick endpoints to incidental vertex (joint) positions.

By method of can be implicit, joining sticks adds additional limitations that reduce the probable prospect of the explanations (top left), the endpoints assumed vertices (top right), and the endpoints specified M_k (bottom left). In distinction the vertex entropy period (bottom right) turns as a per-vertex drawback, which decreases as we combine vertices, preferring additional highly linked models.

4. IMPLEMENTATION OF VIDEO SURVEILLANCE BASED TRAFFIC MONITORING SYSTEM

the practical details of how each algorithm is implemented and how they come together to form a complete system.

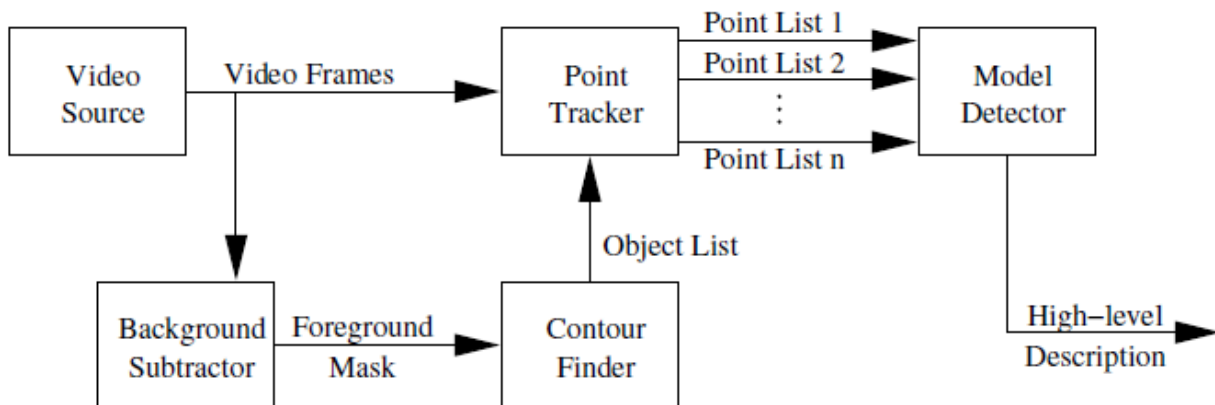


Figure4 : The procedure of complete system

Our application of background removal follows the humble procedure. Throughout the period between model generations, figures for the background are collected from every frame and are incrementally processed. When 240 frames have been reached, a fresh background ideal is calculated for usage preceding the next 240 frames. Besides, subsequently consecutive frames incline to be very alike, only every fourth frame is recycled for statistics gathering. Another interpretation which increases rapidity is to only achieve background deduction on a sub-sampled variety of every image. Therefore, a 640 pixel by 480 pixel image can remain re-scaled to a firmness of 160 by 120, a subdivision of the linear magnitudes.

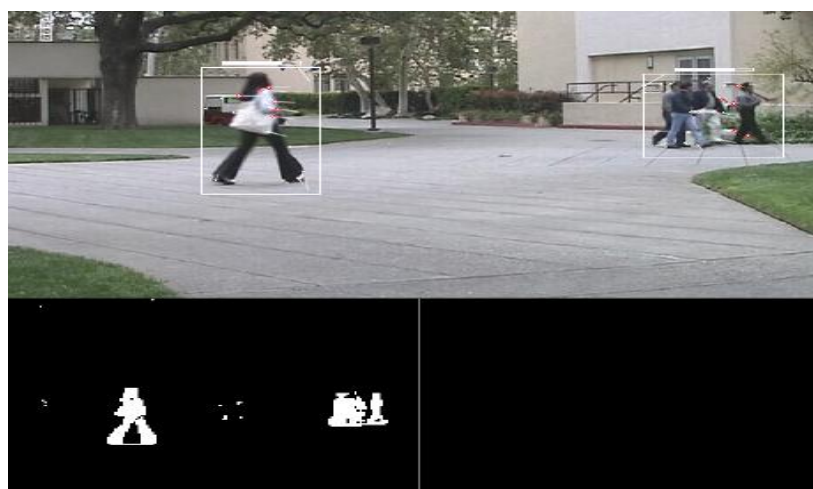


Figure5 :background deduction

Since this image is one sixteenth of the total, the processing time for background subtraction is also one sixteenth of the original time.

background subtraction in terms of three basic ways :

- Background Update, which is run one time for every 4 frames, gathers statistics for the background.
- Model Generation, which is run once every 240 frames, computes the background image from the statistics which have been gathered during the Background Update time.
- Background Subtraction, which generates a foreground mask for every single frame. This way is simply performed by subtracting the background image from the current frame, taking the absolute value of the difference, and thresholding it with the value of three standard deviations of the average image noise.

As discussed in the approach, we made a set of Gaussians for all pixel of the background. Each Gaussian has a mean, variance, and weight. To simplify, the variance is taken to be fixed, similar to the variance of the image noise. After each Model Generation phase, these Gaussians are reboot to their uninitialized state so that they may produce from the Background Update phases. In our algorithm, the weight of a Gaussian is simply similar to the number of frames from where the pixel has taken a value within three standard deviations of that Gaussian's mean. Further, we keep track of the sum of these pixel values, so that the Gaussian's mean is simply sum divided by the number of frames. Also, each pixel has exactly five Gaussians associated with it to simplify data structures. The process of updating each pixel during the Background Update phase are:

- Compute the mean of each of the five Gaussians by dividing each sum by each of the frame count (weight).
- If the current pixel value is upto three standard deviations of any of the five means, deviate that Gaussian's weight by 1 and add the current value to its sum.

5. CONCLUSION

This background subtraction algorithm is simply changed to color images by applying the procedure separately to each of the red, blue, and green channels. If any of the three channels is found to be foreground for a given pixel, the whole pixel is marked as foreground. After background subtraction is complete, the dilate operation is applied two times to the foreground phase. On the other hand to measure accuracy, we developed a procedure to explain the ground truth of a video sequence. Ground truth refers to the actual presence of video surveillance based traffic monitoring system. Once this ground truth is known for a sequence, the performance of our system in detecting video surveillance based traffic monitoring system can be evaluated. In our interface, ground truth is explained by running the detector on a pre-recorded video sequel and manually labeling all frames. The background subtracted will find coordinating regions of all frame that show motion of video surveillance based traffic monitoring system. Each of these regions is manually marked as either video or non- video. In order to update the process, the operator simply corrects any error the detector makes rather than marking every single object. If the system fails to detect video surveillance based traffic, the operator clicks the mouse on that region to indicate the mistake. Likewise, if the system falsely detects a video where there is none, the operator clicks at that location too. Frames that contain unclear objects can be specially marked so that they are not included in the accuracy statistics. For example, if a video is only half-visible at the superiority of a frame, the decision between video and non- video would be meaningless. Once this procedure is over for a sequel, the software will output a file that found the ground truth for each frame. This file can then be used by the software to evaluate its own exactness.

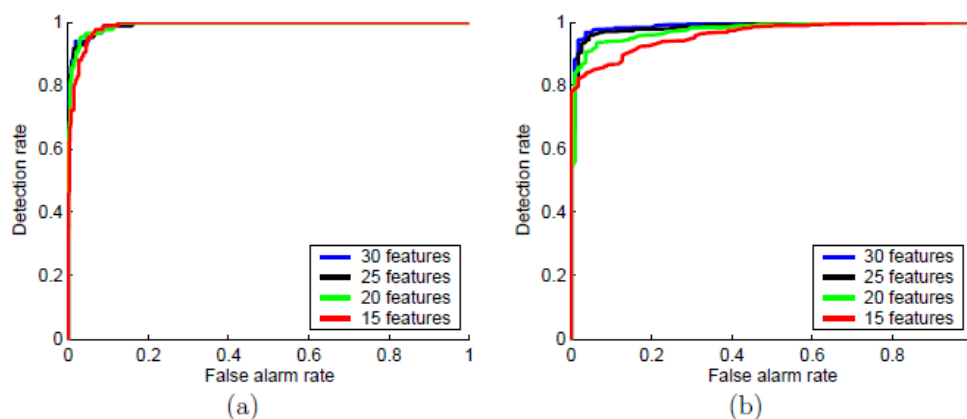


Figure 6 : Receiver Operating Characteristics (ROC) curve

A useful tool for the picturization of these results is the Receiver Operating Characteristics (ROC) curve. Since the system generates a numerical probability for every video object in the scene, there is no clear cut boundary between what should be interpreted as video surveillance based traffic monitoring system and what should be interpreted as non- video surveillance based traffic monitoring system. In order to make such a distinction, a cut-off edge must be chosen for the possible values. A ROC curve shows the rate of precise detection versus the rate of incorrect detection for any choice of blink.

References

- [1] Botelho, L.M. "A Control Structure for Agent Interaction". Proceeding of the Intelligent Vehicles Conference. Submitted. 2000.
- [2] Botelho, L.M.; Lopes, R.; Sequeira, M.M.; Almeida, P.; and Martins, S. "Inter-agent communication in a FIPA compliant intelligent distributed dynamic-information system". *Proc. 5th International Conference on Information Systems Analysis and Synthesis (ISAS99)*. 1999.
- [3] Bretier, P.; and Sadek, D. "A rational agent as the kernel of a cooperative spoken dialogue system: implementing a logical theory of interaction". *Proc.3th ATAL Workshop*. 1996
- [4] Foundation for Intelligent Physical Agents. FIPA 97 Specification, Version 2.0 Part 2 "Agent Communication Language".1998
- [5] Georgeff, M.P.; and Rao, A.S. "The semantics of intention maintenance for rational agents". *IJCAI'95*, p704-710. 1995
- [6] Haykin S., "Neural Networks: a comprehensive foundation". Prentice Hall. 1999
- [7] Helbing, D.; and Huberman, B. A. "Coherent moving states in highway traffic". *Nature* **396**, 738- 740. 1998
- [8] Moravec H.P., "Towards automatic visual obstacle avoidance". *IJCAI'77*, pp. 584. 1977.
- [9] Piscaglia, P.; and Mory, B. ACTS AC304 project MODEST. "Xeno-Object Description Scheme". P134, MPEG-7 Evaluation Ad Hoc meeting.1999
- [10] Searle, J.R. 1969. Speech Acts. Cambridge University Press, 1969
- [11] Ziliani, F.; and Cavallaro, A. "Image Analysis for Video Surveillance based on Spatial Regularization of a Statistical Change Detection". *Proc. 10th Int Conf. on Image Analysis and Processing, Venice*, pp. 1108-1111.1999