Volume 4, No. 4, March-April 2013

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Caste Shadow Removal in Vehicle Detection using Mixture of Gaussian for Traffic Surveillance System

Vanraj Dangar*
Computer Engineering,
Chandubhai S Patel Institute of Technology
Changa, Gujarat, India-388421
dangar.vanraj@gmail.com

Amit Thakkar
Dept. of Information and Technology
Chandubhai S Patel Institute of Technology
Changa, Gujarat, India-388421
amitthakkar.it@ecchanga.ac.in

Abstract:
Observing moving objects in a site and removal of caste shadow is a critical task in computer vision. This paper presents an algorithm for detection and caste removal of vehicles in real-time video which is streamed by a camera with fixed position. Processing of the video is done in two steps: Vehicle Detection and Caste shadow removal. Identifying moving object is done by classification of pixels into either foreground (object) or background. Vehicle detection is achieved by the use of Background subtraction. Many existing scheme of background removal presenting different background models like Mixture of Gaussian (MoG) and Joint Random Field (JRF) will be discussed. A simple approach to remove caste shadow area from the detected foreground objects will also be discussed.

Keywords: Background subtraction; Gaussian Mixture Model; JRF model; Caste Shadow; Thresolding

I. INTRODUCTION

Identifying moving object is a critical task and used in many computer vision application such as human-computer interaction, remote sensing, and video surveillance. Detection and caste shadow removal can be viewed as lower level vision tasks to achieve higher level event understanding. Detection of moving vehicles from static camera is essential task for traffic monitoring and can provide useful piece of information of transportation system like vehicle count, speeds and vehicle classification. Many techniques have been proposed for solving tracking problems and great work has been done to meet the challenges like tracking objects in complex scenarios containing partial occlusions, rapid object motions, and fast change of illumination or complex dynamic background from moving cameras. Most of them has five basic task, which are (1) pre-processing (simple image processing tasks that change the raw input video into a format that can be processed by subsequent steps), (2) background modeling (also known as background maintenance), (3) foreground detection (also known as background subtraction) (4) Caste shadow removal and (5) data validation (also reoffered to as post-processing, used to eliminate those Moving Object Detection in Spatial Domain pixels that do not correspond to actual moving objects). Background modeling refers to the process of creating and subsequently maintaining, a model of the appearance of the background in the field of view of a camera. However, background subtraction refers to the process in which an image frame is compared to the background model in order to determine whether individual pixels are part of the background or the foreground or the shadow. So it is also referred to as foreground detection. Background subtraction is commonly used technique when the video data is captured with a fixed camera. Background subtraction is an old technique for finding moving objects in a video sequence. The idea is that subtracting the current image from a time averaged background image will leave only non-stationary objects. The background model is constructed from observed images and foreground objects are identified if they differ significantly from the background. However, accurate foreground segmentation could be difficult due to potential variability, such as moving shadows caste by foreground objects, illumination or object changes in the background and camouflage (i.e., similarity between appearance of foreground objects and the background). Besides local measurements, such as chromaticity, restrictions in temporal and spatial information from the video scene are very important to deal with the potential variability during the segmentation process. For example, if a pixel is found to be background then nearby pixels have high probability to be in background if intensity variation is not large. Similar explanation can be made for temporal dependencies. Temporal or dynamic information is a fundamental element to handle the evolution of the scene. The background model can be adaptively updated from the recent history of observed images to handle non stationary background processes (e.g., illumination changes). Different random field [1][2][3][4] has been proposed to deal with dynamic in background processes. It is assumed that background model follows morkov property which is quite accurate in case of video sequence. The background modeling is done by various stochastic filters like Kalman filter, Weiner filter and Mixture of Gaussian method [3].Some of them explained in this report.

A. ***Related Work***

Rittscher et al. use both HMM and MRF for foreground and shadow segmentation. In their work, each site (or block) is modeled by a single HMM independent of the neighboring sites (or blocks). The HMM and the MRF are employed in two different processes to impose temporal and spatial contextual constraints, respectively. Paragios and Ramesh and Wang et al. [5] use the hidden Markov random field (HMRF) model to combine different types of features and incorporate spatial constraints. Both methods detect changes between the background and the current image without

© 2010, IJARCS All Rights Reserved

utilizing the previous images. In this work, the foreground is estimated from the history of observed images. The dynamic information from previous observations enhances the confidence of object segmentation, especially at camouflage areas near foreground boundaries. In some of the approach, the variance under shadow is assumed to be smaller than the variance in the background for the same site, which sometimes is not valid for indoor environments. Recently, Cucchiara et al. integrated the knowledge of moving objects, ghosts (apparent objects), and shadows during the segmentation process to enhance object detection and background updating [6]. Points of different classes are handled individually with object-based selective update. The object-level knowledge used in their approach and the statistical model proposed in this work are complementary for efficient and effective segmentation of foreground and shadow in video sequences. For the background updating process, the Gaussian mixture method by Stauffer and Grimson [3] is slightly modified by extending shadow detection property in our work to remove the caste shadow identified in foreground segmented image.

The rest of the paper is arranged as follows: Section II explains the JRF model using CRF (Conditional Random Field). Section III proposes the Mixture of Gaussian model (MoG) for the foreground and shadow detection method with caste shadow removal using an dummy example for the algorithm. Then, our technique is concluded in Section IV. At last future extension is given in Section V.

II. JOINT RANDOM FIELD (JRF)

The JRF model extends the conditional random field (CRF) [2] by introducing auxiliary latent variable to characterize the structure. This method enhances vehicle detection in video by jointly estimating detection labels i.e. (object/background) and hidden variable (i.e. pixel intensity in shadow). The method works efficiently in different illumination, moving caste shadow/lights and dynamic background processes. The JRF is explained below.

Given an image sequence, the observed data and detection label of a point i at time instant t are denoted by l_i^t and d_i^t respectively. The observation d_i^t consists of intensity information at the site i. Label l_i^t assigns the point i to one of K classes (generally K is 3 to 5). The label $l_i^t = e_k$ if point i belongs to the K^{th} class, where e_k is a K-dimensional unit vector with its K^{th} component equal to one. Here $t \in N$, $i \in X$. Where N and X is the spatial domain of the video scene. The entire label field and observed image over the scene at time t are compactly expressed as l_i^t and d_i^t respectively.

A.JRF Model

For random variable l and observed data z over the video scene, (l, z) is a conditional random field if, when conditioned on z, the random variable obeys the markov property. i.e.

$$p(l_i | z, l_j, j \neq i) = p(l_i | z, l_j, j \in N_i)$$

where the set N_i denotes the neighbouring sites of point i. Hence l is a random field globally conditioned on the observed data. In order to introduce auxiliary hidden variables during the labeling process, the notion of JRF [1] and DCRF [2] is used.

For two random field (l, h) and observed data z, (l, h, z) becomes a joint random field if,

$$p(l_i, h_i | z, l_j, h_j, j \neq i) = p(l_i, h_i | z, l_j, h_j, j \in N_i)$$

i.e. the couple (l, h) is Marcovian when conditioned on observed data z.

For two random field (l, h) and observed data z ,(l, h, z) becomes a joint random field if,

$$p(l_i, h_i | z, l_j, h_j, j \neq i) = p(l_i, h_i | z, l_j, h_j, j \in N_i)$$

i.e. the couple (l, h) is Marcovian when conditioned on observed data z.

For given the observed image z^t at time instant t, the joint probability distribution over the label field l^t and latent field h^t is modeled by a joint random field $(l^t, h^t; z^{1:t})$ to formulate contextual dependencies. Here, $z^{1:t}=\{ z^k , k=1, 2,\ldots t\}$ is the sequence of observed data up to time t. Thus the couple (l^t, h^t) obeys the markov property when the observed data z^t is given.

For traffic monitoring, each pixel in the scene is to be classified as moving, shadow or background (roadway). For a site I at time t, label l_i^t is defined as follow.

$$l_i^t = \begin{cases} e_1 & background \\ e_2 & shadow \\ e_3 & vehicle \end{cases}$$

Here static shadows are considered to be part of the background. For each point i, the pixel intensity z_i^t has three (R, G, B) component for color images or one value for grayscale images. Since, the intensity under shadow is not given beforehand, so the auxiliary variable h_i^t is used to characterize the scene for each point i at time t. to segment the moving vehicles from the scene, the system should model the background and shadow information. Assuming each pixel is corrupted by Gaussian noise, the background is modeled as $z_i^t=b_i^t+n_i^t$, where b_i^t is the intensity mean for a pixel i within the background and n_i^t is independent zero-mean Gaussian noise with variance $(\sigma_i^t)^2$. Here b_i^t and $(\sigma_i^t)^2$ can be estimated from previous images. Similar Gaussian model for scene is defined to describe the same point when shadowed. i.e. $z_i^t=b_i^t+n_i^t$. The shadow portions always have less intensity so the value of σ_i^t is between 0 to 1. For maximum application independence, it is assumed that the intensity information of vehicle is unknown. Hence, uniform distribution is used for the pixel intensity of moving vehicle.

From the above discussion, the local intensity likelihood of a point at time t becomes

$$p(z_i^t | l_i^t) = \begin{cases} N(z_i^t; b_i^t, (\sigma_i^t)^2) & l_i^t = e_1 \\ N(z_i^t; \rho_i^t b_i^t, (\sigma_i^t)^2) & l_i^t = e_2 \\ uniform - distribution & l_i^t = e_3 \end{cases}$$

However, the observation model tends to confuse cast shadow and moving vehicle at boundary areas or in uniform regions, especially when the vehicle is darker than the background and the road surface is untextured. Such detection error can be effectively reduced if the intensity of shadowed points is known, i.e. $z_i^t=h_i^t+n_i^t$ if $l_i^t=e_2$, when h_i^t is the mean intensity under caste shadow (or light) for site i. since the intensity under shadow points is not given beforehand, in this work is used as the auxiliary latent variable to characterize the visual scene for each point i at time t .The probability $p(z_i^t | l_i^t)$ is given by the local intensity likelihood. For pixel intensity under cast shadow (or light), the posterior

$$p(h_i^t | l_i^t, z_i^t) = \begin{cases} N(h_i^t; z_i^t, (\sigma_i^t)^2) & l_i^t = e_2 \\ uniform - distribution & otherwise \end{cases}$$

in the original image.

© 2010, IJARCS All Rights Reserved

In case of Gaussian mixture model, the probability distribution of pixel intensity is modelled by a mixture of Guassian

$$p(z) = \sum_k w_{i,k}^t N(z_i^t | \mu_{i,k}^t, (\sigma_{i,k}^t)^2),$$

Where $N(z | \mu, \sigma^2)$ a Guassian distribution with argument is z, mean μ, variance σ^2 and $\{w_{i,k}^t\}$ denotes the corresponding weights for the Gaussian mixture. Given the current video frame z^t, each pixel value is checked to match the existing Gaussian distribution. For a matched Gaussian, its weights increases and the corresponding mean and variance are updated using the pixel value. For unmatched distribution, the mean and variance remain the same, while weight should be renormalized. If none of the distributions match the pixel value, the distribution of the lowest weight is replaced with a Gaussian with the pixel value as its mean, initially low weight and high variance. For each point *i*, the Gaussian distribution that has the highest ratio of weight over variance $w_{i,k}^t / ((\sigma_{i,k}^t)^2)$ is chosen as the background model at time instant *t* [1].

III. FOREGROUND AND SHADOW SEGMENTATION

A. Mixture Of Gaussian (MoG)

In [7], the background was modeled using single Gaussian assumption. So Kalman filter approach was used. However, it wasn't that accurate assumption. Mixture of Gaussians removes the limitation of single Gaussian. It gives the PDF (Probability Density Function) for the pixel intensity which can be dependent on other pixel also. The background of the scene contains many non-static objects such as tree branches and bushes whose movement depends on the wind in the scene. This kind of background motion causes the pixel intensity values to vary significantly with time. So a single Gaussian assumption for the pdf of the pixel intensity will not hold. Instead, a generalization based on a mixture of Gaussian has been used in to model such variations. The pixel intensity was modelled by a mixture of many (K-number which is small between 3 and 5) Gaussian distributions. For traffic surveillance system, 3 Gaussian distribution need to be used to model the pixel value, corresponding to road, shadow and vehicle distributions. Although, in this case, the pixel intensity is modelled with three distributions, still unimodel distribution assumption is used for the scene background, i.e. the road distribution. Unlike kalman filter which tracks the evolution of a single Gaussian, the MoG method tracks multiple Gaussian distributions simultaneously. MoG has tremendous popularity since it was first proposed for background modeling. The generalized mixture of Gaussians (MoG) has been used to model complex, non static background. Stauffer and Grimson [3] allow the background model to be a mixture of several Gaussians. Every pixel value compared against the existing set of models at that location to find a match. The parameters for the matched model are updated based on a learning factor. If there is no match, the least-likely model is discarded and replaced by a new Gaussian with statistics initialized by the current pixel value. The models that justify some predefined fraction of the recent data are deemed "background" and the rest "foreground". There are additional steps to cluster foreground pixels into semantic objects and track the objects over time [3].

B.Algorithm

Input: Video file (in .asf/ .avi / .mp4 format) captured by Traffic Surveillance Camera

Output: Blobs of detected moving vehicles without Caste shadow at given instance of time with total no of detected vehicles.

Method: Background model (Frames), Caste shadow (z)

Subroutines: $p(z), N(z | \mu, \sigma^2)$

Parameters: μ: mean, σ^2: variance, α: learning rate, $w_{i,k}^t$: weight of pixel i at time instance t for Gaussian component k, τ: threshold

1. Model Adaptive Background:
   (a) Convert video file into frames for time instance t
   (b) From first 50/100 frames model background by calculating probability distribution function for k=3 using following equations

$$p(z) = \sum_k w_{i,k}^t N(z_i^t | \mu_{i,k}^t, ((\sigma_{i,k}^t)^2)$$

$$N(z | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

2. Model adaption using K-mean algorithm:
   For every pixel in each new frame,
   If ($p(z) \geq \tau$) // generally $\tau = 2.5$ as per [4], [8]
   Update the statistic for Gaussian component k…

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho z_{i,t+1}$$

$$\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + \rho(z_{i,t+1} - \mu_{i,t+1})^2$$

$$\rho = \alpha N(z_i^t | \mu_{i,k}^t, ((\sigma_{i,k}^t)^2)$$

$$w_{i,t+1} = (1-\alpha) w_{i,t} + \alpha M_{i,t+1}$$

$$\text{where } M_{i,t+1} = \begin{cases} 1 & \text{for matched Gaussian} \\ 0 & \text{for all other} \end{cases}$$

   //
   else
   Keep same mean and variance but renormalized the weights

3. Background Subtraction: Get a new frame and subtract modeled background image
   F=I – B
   for each new blob increment the count
4. Removal of caste shadow:
   Caste shadow (B, I)
   if ($(R \geq R_p \geq \alpha R)$&$(G \geq G_p \geq \beta G)$ & $(B \geq B \geq \gamma B)$
   return true;
   else return false;

Each pixel is represented by three Gaussian distribution. It has been observed that three distributions gives better performance. Different Weights Wk are assigned to each distribution $z \sim N(z_i^t | \mu_{i,k}^t, ((\sigma_{i,k}^t)^2)$ where k = 1, 2, 3 for each new frame and most suitable background value is selected as an estimation. After generating background image, it is subtracted from original frame data and difference image is obtained. By doing thresholding of difference image, moving object is segmented and foreground is obtained. For each new frame background need to be update for real time system. For that K-mean algorithm has been used to reconstruct the background if there is any significant change has been done.

This foreground also has shadows. Shadow expands the region of vehicle in foreground image and which results in false contouring of vehicle. It will also merge foreground blob of two or more vehicle in case of large shadow. This is why it should be removed. After shadow removal, median filtering is done to remove salt noise. The foreground image also has region other then road scene.

© 2010, IJARCS All Rights Reserved

This region contains additional moving elements like trees and pedestrian. A binary mask is created to cover such unnecessary part of image. The obtained foreground has regions which are scattered in small groups. This kind of foreground cannot be further processed without combining these small groups; otherwise it will lead to false detection of vehicles. They are combined using morphological process like closing using appropriate structuring elements to create a single blob which represents a single vehicle.

Shadow = if (($R \geq R_p \geq \alpha R$) & ($G \geq G_p \geq \beta G$) & ($B \geq B \geq \gamma B$))

return true;

else return false;

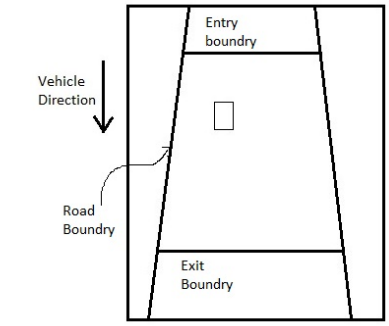
Here, (Rp, Gp, BP) is RGB color value of pixel.

As shadow cast over the surface, the intensity of pixel (Rp, Gp, Bp) decreases. If it is assumed that intensity of pixel for background region in RGB gray level is (R, G, B) and it decreases by factor α, β, γ accordingly then following threshold can be taken into account.

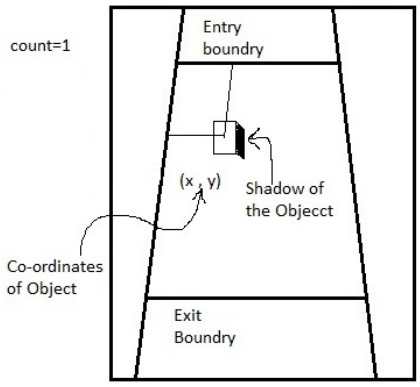
factor α, β, γ are less than 1 and varies as per lighting condition. These factors can be estimated by overall illumination condition of scene. This kind of approach is not quite efficient for umbra type of caste shadow which is almost dark region.

C. Example

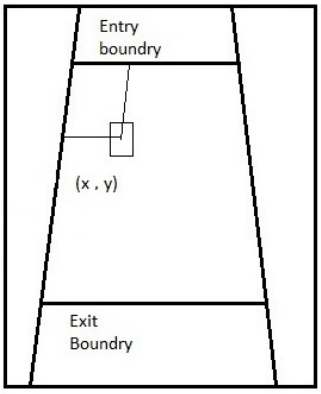
The proposed approach will work on gray scale video sequences as well as color image sequences. Here one dummy example for detecting vehicle and caste shadow removal is shown in figure: 1. Fig. 1 (a) shows the initial conditions for the input image. Fig.1 (b) shows the detected foreground with shadow region within boundary area. Fig.1 (c) shows the output image after removal of caste shadow.



(a)



(b)



(c)

Figure 1. (a) Original Video frame image. (b) Output after Background Subtraction using MoG. (c) Output after Shadow Removal.

IV. CONCLUSIONS

The work presented in this report is motivated in large part by a practical need identified in the security and transportation planning industry to collect volume information from transportation scene in an accurate and cost effective way. The introduction of computer vision based approach for the conducting this type of vehicle volume data collection represents a significant improvement over the current manual data collection methods. The manual methods are expensive, due to the labor involved, and plagued by inaccuracies caused by fatigue. The data collected using this approach shall be helpful in monitoring the road traffic in a better way. The information about traffic patterns shall also be used in future planning of road networks. The report essentially proposes an approach incorporing Adaptive Gaussian Mixture Model and blob based detection and then caste shadow removal. The usage of various modules is optimized to minimize the processing time.

V. FUTURE EXTENSIONS

The work can be extended for tracking the path of moving vehicles, average speed measurement and also classification based on the size of the vehicle so that as much as possible traffic data can be efficiently collected.

VI. REFERENCES

[1] Joint Random Field Model for all-Weather moving vehicle Detection By Yang Wang, IEEE Transaction on IMAGE PROCESSING, Vol. 19, No. 9, September 2010

[2] A Dynamic Conditional Random Field Model for Foreground and Shadow Segmentation By Yang Wang, Kia-Fock Loe and Jian-Kang Wu, IEEE Transaction on PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol. 28, No. 2, FEB 2006

[3] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," IEEE Transaction on PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol. 22, No. 8,pp. 747-757, Aug. 2000

[4] Adaptive background mixture models for real-time tracking by Chris Stauffer and W.E.L Grimson, 1999

[5] N. Paragios and V. Ramesh, "A MRF-Based Approach for Real-Time Subway Monitoring," Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 1034-1040, 2001.

[6] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams,"IEEE Trans. Pattern Analysis and Machine Intelligence,vol. 25, pp. 1337-1342, 2003

[7] N.Martel-Brisson and A. Zaccarin, "Moving cast shadow detection from a Gaussian mixture model shadow model",

© 2010, IJARCS All Rights Reserved

IEEE Transaction. Intelligent Transportation System, vol. 9, no. 1,pp. 148-160, March 2008.

[8] Simple Vehicle Detection with Shadow Removal at Intersection by Chao Yuan, Chenhui Yang, Zhiming Xu, Xiamen University, Second International Conference on Multi Media and Information Technology, 2010

© 2010, IJARCS All Rights Reserved