Video Object Classification System with Shadow Removal using Gaussian Mixture Model

[1]Adekunle A.O., [2]Omidiora E.O., [3]Olabiyisi S.O. and [4]Ojo J.A.
[1]*Computer Science Dept. Adeyemi Federal University of Education,Ondo Nigeria*
[2, 3,4]*Computer Science and Engineering Dept, Ladoke Akintola University,Ogbomoso Nigeria*
gaacademics@gmail.com/adekunleao@aceondo.edu.ng; eoomidiora@lautech.edu.ng;
soolabiyisi@lautech.edu.ng; jaojo@lautech.edu.ng

ABSTRACT

Classification is the process of assigning a class to a group of objects. Moving objects classification can be difficult a task in the presence of dynamic factors like occlusion clutters and shadows. This paper developed a classifier for moving images (video stream) by using a modified adaptive background mixture model method. This system removes shadows and correctly classifies moving objects as human, human group and vehicles.

Background Mixture Model is common technique used in Computer Vision, Video object classification is not an exception, many background models have been designed to address different problems ranging from slow start and shadow removal; this paper presents a method which models each pixel as a mixture of Gaussians and using a Maximum A Posteriori approximation to update the model, this paper also introduces a two level shadow removal technique which suppressed shadows in colour and texture consistencies in the classified objects so that the system will not mis- classify moving shadows as objects. This work overcome the problem of slow learning in busy environment and can classify more than one object in view of the camera

Keywords-Classification, Gaussian Mixture Model, background subtraction, Maximum A Posteriori and Shadow detection

1   Introduction

It is easy for human beings to identify moving objects in a video clip; it is also not difficult for human beings to categorize such objects as a vehicle, a human being, a bike or a helicopter. However, it is rather a difficult task for a computer system to do the same (Dedeoglu, 2004), and due to this reason, computer vision has become an important field of study. In computer vision images are acquired, processed and analyzed to produce information. Such images can be taken from video sequences and multiple camera; the applications of computer vision in real life includes medical and automation industry. This work proposes a system that is able to distinguish transitory and stopped foreground objects from static ground objects in dynamic scene and classify detected objects into different groups such as human, human group and vehicle. The system will also detect shadows and suppress it; generate trajectory information even in multi-occlusion cases in video imagery.

2   Design

A reference background is initialized at the start of the system with the first few frames of video and updated to adapt to short and long term dynamic scene changes during the operational period. At

each new frame, foreground pixels are detected by subtracting the intensity values from the background and filtering the absolute value of the differences with a dynamic threshold per pixel.The Gaussian distribution for each pixel is maintained and updated by using the Maximum *A Posteriori estimate* .The individual pixels are now grouped and labelled by using the component libelling algorithm to create connected moving regions. These regions are further processed to group disconnected blobs and to eliminate relatively small sized regions.After grouping, each detected foreground object is represented by its bounding box, area, center of mass and colour histogram. Moving pixels from the static background of the scene, connected regions are classified into predetermined object categories human, human group and vehicle. The classification algorithm depends on thresholds that are set based on the distance of the camera to the objects, shadows are now removed using the HSV and Phong's illumination Model.

2.1   Current State Estimation

While Kaew Trakulpong and Bowden (2001) employed the L-recent windows update equations to determine the current state because it allows fast convergence on a stable background model, this work will estimate the current state of the model by classifying each pixel to know how it looks when the pixel is part of a different class. Maximum a Posteriori (MAP) will learn how a Mixture of Gaussian (MOG) will view such a pixel. Like Expectation and Maximization, MAP is also a two-step estimation process: the first step is used to compute the estimate of sufficient statistics of the training data for each mixture in the prior model. The second step handles the "new" sufficient statistics estimates and then combine with the "old" sufficient statistics from the prior mixture parameters. The parameters are collectively represented by the notation

$$\Theta = \{ w_i , \mu_i , \Sigma_i \} \tag{1}$$

w , μ , Σ represents the Gaussian weight, mean and covariance matrix respectively.

A Gaussian Mixture Model of M component Gaussian density is given by

$$P(x \backslash \Theta) = \sum_{i=1}^{M} \omega_i g(x \backslash \mu_i \sum_i) \tag{2}$$

X is a D- dimensional continous valued data vector ω_i is the mixture weight,i=1,…M

$g(x \backslash \mu_i \sum_i) =$ Component Gaussian density

i=1…M

D -Variate Gaussian function

$$g(x \backslash \mu_i \sum_i) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\{ \frac{-1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \} \tag{3}$$

Where

μ_i is the mean vector

Σ_i is the covariance matrix and

$\sum_{i=1}^{M} \omega_i = 1$ (must be satisfied)

$$\Theta = \{\omega_i, \mu_i, \Sigma_i\} \quad i=1,\ldots,M$$

The Posteriori probability for component i is given by:

$$\Pr(i \setminus x_t, \Theta) = \frac{w_i\, g(x_t \setminus \mu_i, \Sigma_i)}{\sum\limits_{i=1}^{M} w_k\, g(x_t \setminus \mu_k, \Sigma_k)} \tag{4}$$

Given a prior model and training vectors from a class

$$x = \{x_1, \ldots, x_T\} \tag{5}$$

The M distributions are ordered based on the fitness value w_i / σ_i and the first B distributions are used as model of the background of the scene where B is estimated as

$$B = \arg_b \min\left[\sum_{i=1}^{b} w_i > T\right] \tag{6}$$

The threshold T is the minimum fraction of the background model, it is the minimum prior probability that the background is in the scene. Background subtraction is performed by marking a pixel foreground if it is more than 2.5 standard deviation away from any of the B distributions.

The above new sufficient statistics from the training data are used to update the priors sufficient statistics for mixture *i* to create adapted parameters for mixture *i* having the following equations will update the Gaussian

$$w_i = \{\alpha_i^w N_i / T + (1 - \alpha_i^w)\}\omega_i\}\gamma \tag{7}$$

$$\hat{\mu}_i = \{\alpha_i^m \mu_i + (1 - \alpha_i^m)\mu_i\} \tag{8}$$

$$\hat{\sigma_i^2} = \alpha_i^v \sigma_i^2 + (1 - \alpha_i^v)(\sigma_i^2 + \mu_i^2) - \mu_i^2 \tag{9}$$

Where

α_i^w is for the mixture weights

α_i^m is for the mixture means

α_i^v is for the mixture variances

γ is the scale factor, which is ensured sum to unity for all adapted mixture weight.

2.2 Shadow Removal Technique

The technique employed is a two level approach that removed shadows optimally in outdoor situations; this research work used the HSV suppression and the Phong reflection model, since each of the approaches suffers its own weaknesses. Phong reflection model, an illumination model widely used in 3D computer graphics would be employed to remove shadows from video streams, Phong reflection model helps to prove the local coherence(over a pixel neighborhood) of intensity reduction ratio used in texture verification. Phong model exploited the chromaticity, texture and intensity

reduction. According to Phong illumination model, a surface point is lit by three types of lights: ambient light i_a, diffuse light i_d, and specular light i_s. The point of luminance in the image is described by:

Our relevance is the RGB colour space that becomes

$$I^j = k^j(i_a^j + (L.N)i_a^j) \qquad (10)$$

where index j corresponds to red, green and blue.

A shadow occurs when light power from the light source to a surface is partially or completely blocked by an object. Then the point of luminance becomes:

$$I_{shadow}^j = k^j(i_a^j + \beta(L.N)i_d^j) \qquad (11)$$

where $\beta \, \varepsilon \, [0,1]$ indicating how much diffused light has been blocked.

After that HSV suppression will be applied since the phong reflection model works well in indoor environments.

3 Implementation

Several video recording were made at different times of the day to monitor designated campus sites where students' moves freely except for the human actions simulated.

SETUP

This paper worked on the setup parameters below based on set objectives:

- Uniform background, uniform illumination
- Non- uniform background, uniform illumination

3.1 Test Cases

The test cases involved the set objectives of this research work, i.e. to detect moving regions in a video frame, classify these regions as human, human group and to optimally remove shadowed regions using a combination of HSV suppression and Phong's Model.

3.2 Moving Object Classification Detection Results

Figure 1: Correctly classified Human Group (people on a moving bike and people moving inside the building)
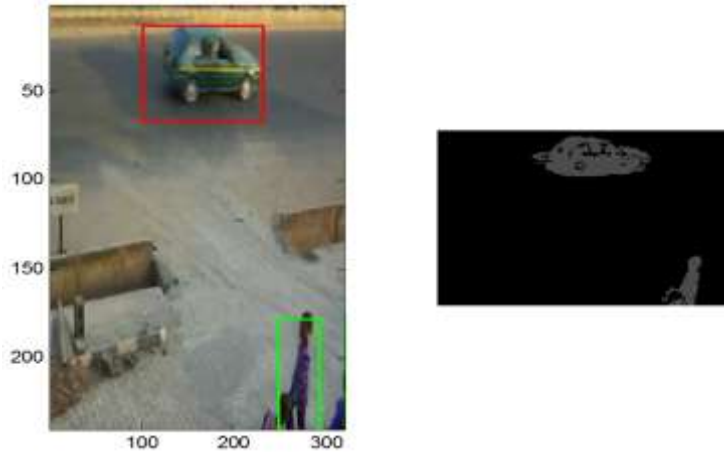
Figure 2: Correctly classified car and human group on the view of the camera
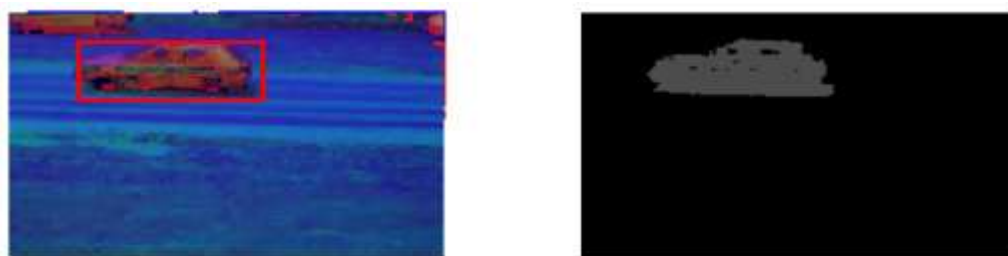

Figure 3: Correct vehicle classification with partial shadow removal(HSV suppression only)

Figure 4: Correct Human Classification and shadows completely removed

3.3   Classification Experimental Results

As seen in Figures 1,2,3 and 4.The algorithm classifies human(H)(4), human group(HG)(1 and 2) and vehicle(V)(3) correctly and generates a 3x3 confusion matrix in Table 4.3, it is observed that the algorithm has a high mis-classification(high false negative) in human group using this algorithm for some instance, this is due to the variance of the said group of people being equal to that of cars or trucks.; it can however be improved if the camera distance from the object is altered. The True positives 32, 23, and 45 obtained for Human, Human Group and Vehicles respectively.

4   Evaluation

This paper measured the sensitivity of the model to detect moving targets with low contrast against the background and how this sensitivity is affected by the target presence in the scene using the Percentage Correct Classification (PCC) and the Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC).

Table 1: The Confusion Matrix of the Moving Object Classification.

	H	HG	V	%
H	32	0	0	100
HG	0	23	2	91
V	0	0	45	100

ROC Curve

Diagonal segments are produced by ties.
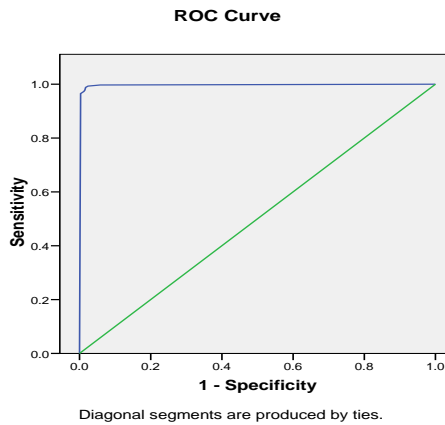
Percentage Correct Classification =98%; Area Under Curve=0.9

Figure 5: The ROC of the Moving Object Classification

4.1   Shadow Removal Technique

The performance of any shadow detection and removal technique can be tested using two metrics proposed by Prati *et al*, namely shadow detection rate (η) and shadow discrimination rate (ξ):

$$\eta = \frac{TP_S}{TP_S + FN_S} \tag{12}$$

$$\xi = \frac{TP_F}{TP_F + FN_F} \tag{13}$$

Where TP and FN stand for true positive and false negative pixels with respectively to either Shadows(S) or foreground objects (F). The shadow detection rate is concerned with labeling the maximum number of cast shadow pixels as shadows. The shadow discrimination rate is concerned with maintaining the pixels that belong to the moving object as foreground.

Table 2: Table showing the discrimination rate and detection rate at each stage of the shadow removal

	Shadow Detection Rate	Shadow Discrimination Rate
HSV	0.482	0.786
HSV&PHONG	0.695	0.895

5 Conclusion

This paper employed the modified adaptive background mixture model method in detecting human and vehicular motions in video images, the Maximum A posteriori (MAP) was used to update the Gaussian to , classify the moving region as Human, Human Group or Vehicles and remove shadowed regions in the video scene optimally using the combination of HSV suppression and the Phong's illumination Model(Chromaticity and Texture constraints only).

The system performed well in detection of moving images(Human, Human Group or vehicles), the system also classifies moving objects in its view to Human ,Human Group and Vehicles, but fails in classification when a group of people are walking or running at the same speed in the same direction; this type of motion confuses the system giving this set of people the variance of a car or a truck, The system performs well in classification creating a bounding boxes on the target objects correctly even when more that one object is in its view , for example ,Figure 4.9 where a moving Taxi cab(Red Box) and a woman with her child(Green Box) were correctly classified.

REFERENCES

[1]   Cutler, R. and Davis, L.S. (2000): "Robuct real-time motion detection, analysis and Applications". In IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 8, pp. 781-796.

[2]   Dahmen, J., Schluter, R., and Ney, H. (1999): "Discriminative training of Gaussian Mixtures for image Object Recognition". In proceeding 21st Symposium of german Association for pattern Recognition, Bonn, germany, pp. 205-212.

[3]   Dedeoglu Y.(2004): Moving object detection, tracking and classification for smart video surveillance,Master's Thesis, Bilkent University, Ankara,pp2.

[4]   Duda, R.O. and Hart, P.E. (2002): "Pattern Classification and Scene Analysis". Nevada. John Wiley and Sons, inc. pp. 123-124.

[5]   Haritaoglu I.,harwood L.and Davis S. (1998): "W^4 Who? When?where? what? A real time System for detecting and tracking people" FGR98 submitted.

[6]   KaewTrakulPong, P. and Bowden, R. (2001): "An improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection". In proceeding. 2nd European Workshop on Advanced Video based Surveillance Systems.

[7]   Landabaso J-L,Pardàs M and Xu L.(2011): Shadow Removal with Morphological Reconstruction *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 3, pp.810-813

[8]   Omidiora,E.O., Fakolujo,O.A., Ayeni,R.O., Olabiyisi,S.O. and Arulogun O.T. (2008) "Quantittive Evaluation of PCA and FDA in Teaching of face Image" Journal of Computer Science and its Application 15(1), pp22-37

[9]   Phong, B.T. (1975): Illumination for computer generated pictures, Communications of ACM Volume (18) 6, Pp. 311–317.

[10]   Prati A.,Mikic I.,Trivedi M. and Cucchiara R.(2003):Detecting Moving Shadows: algorithm and Evaluation.IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(7):Pp.918-923

[11] Reynolds D,Ridder C., Munkelt O. and Kirchner H.(1995): Adaptive background estimation and foreground   detection using kalman-filtering. Proceedings of the International Conference on Recent Advances in Mechatronics,Academic Press, pp: 193-199.

[12] Stauffer C. and Grimson W. (1998):" Adaptive background mixture models for real-time Tracking". Proc. Of the IEEE Computer Society conference on Computer Vision and pattern recognition, pp 246-252.

[13] Stauffer C. and Grimson W. E. L.(1999) "Adaptive background mixture models for real-time tracking". Proc. Of the IEEE Computer Society conference on Computer Vision and pattern recognition (Cat. No PROO149). IEEE Comput. Soc. Part Vol. 2.pp 87-94

[14] Stauffer, C. and Grimson, W. E. L. (2000). Learning patterns of activity using real-time tracking.IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.22(8):pp.747–757.

[15] Wixson L. and Selinger A. (1998):" Classifying moving objects as rigid or non-rigid". In Proceeding of DARPA Image Understanding Workshop, pp. 341-358.