Object Identification in Video

M. Suba [1], Dr.M. Kalaiselvi Geetha[2]
[1, 2]Department of CSE
[1, 2]Annamalai UniversityChidambaram, Tamil Nadu

Abstract-*Intelligent video surveillance system has emerged as a very important research topic in the computer vision field in the recent years. It is well suited for a broad range of applications such as to monitor activities at traffic intersections for detecting congestions and predict the traffic flow. Object classification in the field of video surveillance is a key component of smart surveillance software. The robust methodology is adopted for people and vehicle classification for automated surveillance systems is proposed. This work uses background subtraction model for detecting the moving objects and the object segmentation is done by the morphological operations. The Grey Level Co-Occurrence Matrix (GLCM) features extracted from the segmented objects and classification is done with the Artificial Neural Network (ANN). The result demonstrates the effectiveness of the proposed approach with high accuracy.*

Keywords- ANN, congestions, GLCM, morphological operations

I. INTRODUCTION

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance. Object recognition technology in the field of computer vision for finding and identifying objects in an image or video sequence. Objects can even be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems. There are two common approaches to adapting isolated object classifiers to visual scenes. The “sliding window” method considers rectangular blocks of pixels at some discretized set of image positions and scales. Each of these windows is independently classified, and heuristics are then used to avoid multiple partially overlapping detections. An alternative “greedy” approach begins by finding the single most likely instance of each object category.

II. LITERATURE SURVEY

[1] In this system, interaction between image segmentation (using different edge detection methods) and object recognition are discussed. . Expectation-Maximization algorithm and OTSU algorithm exhibited stable segmentation effect. [3] The system discusses various techniques such as the template matching technique requires large database of image templates for correct object recognition. Hence it must be used only when limited objects are to be detected. [4] This system presents a novel method for recognizing object categories when using multiple cues by separately processing the shape and color cues. Color is used to compute bottom-up and top-down attention maps. [6] In this system, we propose an object detection approach using spatial histogram features. The cascade histogram matching is trained via automatically selected discriminative features. [9] Two different types of image feature algorithms, Scale - Invariant Feature Transform (SIFT) and the more recent Speeded up Robust Features (SURF), have been used to compare the images. [10] The system describes an approach to vehicle detection and tracking fully based on the Block Matching Algorithm (BMA). Finally, the tracking algorithm establishes the correspondences between the vehicles detected in each frames of the sequence. [12] This system devoted to propose template match object detection for inertial navigation systems (INS). The proposed method is an image processing technique to improve the precision of the INS for detecting and tracking the ground objects from flying vehicles.

III. METHODOLOGY

The robust methodology is adopted for people and vehicle classification for automated surveillance system. The proposed method is evaluated using custom recorded - Chidambaram traffic dataset (showing objects such as cars, two wheeler and humans). Difference image is obtained by background subtraction algorithm in order to find the foreground object. The GLCM features are obtained from the extracted foreground objects. The extracted feature is fed to the Artificial Neural Network (ANN) classifier for classification. The overall block diagram of the proposed approach is shown in Figure 1.

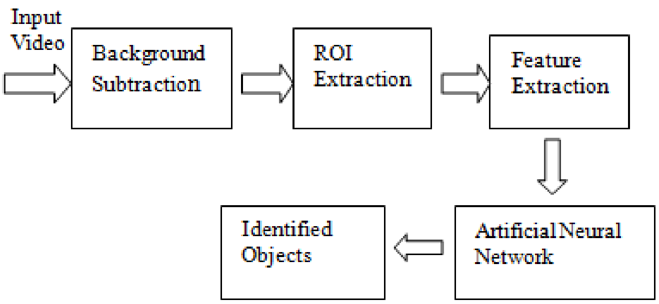

Figure 1. Overall architecture of the proposed approach

A. Preprocessing

Noise in an image is the consequence of errors in image acquisition process and results in image pixel values that do not reflect the true intensities of the real scene concerned. Noises in an image may be reduced or removed by using a filtering technique. Filtering is a technique of enhancing an image for emphasizing certain features or removing other features. Linear filtering technique is used in this proposed system.

B. Background Subtraction Algorithm

Background Subtraction algorithm attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image. The background image BT is updated by the use of an Infinite Impulse Response (IIR) filter as follows:

$$B_{t+1} = \alpha I_t + (1-\alpha)B_t \qquad (1)$$

The foreground pixel map creation is followed by morphological closing and the elimination of small-sized regions.

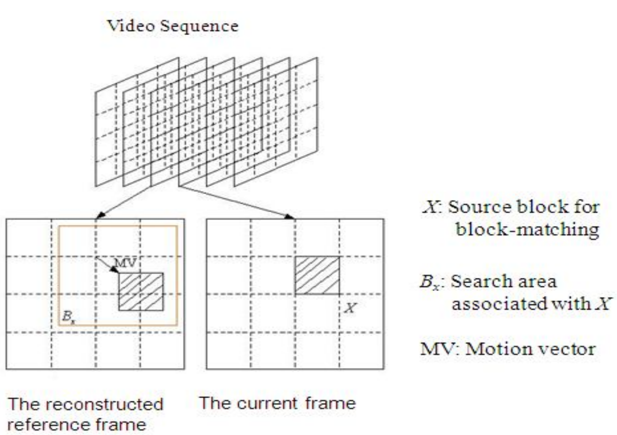

Figure 2. Moving object from background subtraction

C. Morphological operations

Morphology is a vast extent of image processing operations that modifies the images based on shapes. It is considered to be one of the data processing methods useful in image processing. It has many applications like texture analysis, noise elimination, boundary extraction etc. Morphological image processing follows the goal of eliminating all these defects and maintaining structure of image. Morphological operations are confident only on the associated ordering of pixel values, rather than their numerical values, so they are focused more on binary images, but it can also be applied to grayscale images such that their light transfer functions are unknown and thus their absolute pixel values are not taken into consideration. The base of the morphological operation is dilation, erosion, opening, closing expressed in logical AND, OR notation and described by set analysis. Dilation adds pixels while erosion removes the pixels at boundaries of the objects. This removal or adding of pixels depends on the structuring element used for processing the image.

D. Grey Level Co-Occurrence Matrix (GLCM)

The GLCM was introduced by Haralick. It is a second order statistical method which is reported to be able to characterize textures as an overall or average spatial relationship between grey tones in an image. Its development was inspired by the conjectured from that second order probabilities were sufficient for human discrimination of texture. In general, GLCM could be computed as follows. First, an original texture image D is re-quantized into an image G with reduced number of grey level, Ng. A typical value of Ng is 16 or 32. Then, GLCM is computed from G by scanning the intensity of each pixel and its neighbor, defined by displacement d and angle ø. A displacement, d could take a value of 1, 2, 3…n whereas an angle, ø is limited 0□, 45□, 90□ and 135□.

The GLCM P (i,j|d,ø) is a second order joint probability density function P of grey level pairs in the image for each element in co-occurrence matrix by dividing each element with Ng. Finally, scalar secondary features are extracted from this co-occurrence matrix. 8 of the most commonly used GLCM secondary features, are employed as inputs to the neural network classifier. To visualize the mechanism of GLCM, it is best described by the example shown in Figure 3.

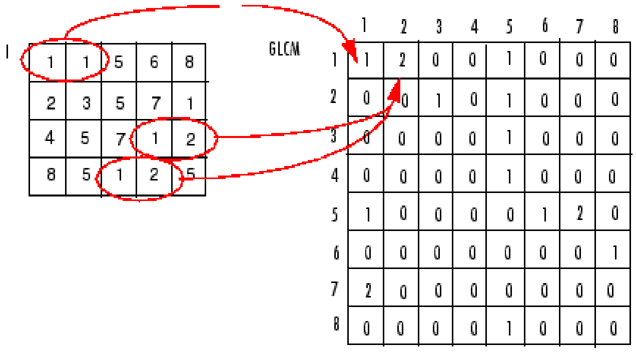

Figure 3. GLCM Co-occurrence matrix method

A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the image. The matrix element P(i,j| Δx, Δy) is the relative frequency with which two pixels, separated by a pixel distance (Δx, Δy), occur within a given neighborhood, one with intensity I and the other with intensity j. One may also say that the matrix element P(i,j | d, θ) contains the second order statistical probability values for changes between gray levels I and j at a particular displacement distance d and at a particular angle (θ).Given an M x N neighborhood of an input image containing G gray levels from 0 to G — 1, let f(m,n) be the intensity at sample m, line n of the neighborhood, then

$$P(i,j|\ \Delta x, \Delta y) = WQ(i,j|\ \Delta x, \Delta y) \qquad (2)$$

where

$$W = \frac{1}{(M - \Delta x)(N - \Delta y)}$$

$$Q(i, j / \Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A \qquad (3)$$

$$A = \begin{cases} 1 & \text{if } f(m,n) = i \text{ and } f(m+\Delta x, n+\Delta y) = j \\ 0 & \text{elsewhere} \end{cases} \qquad (4)$$

A small (5 x 5) sub-image with 4 gray levels and its corresponding gray level co-occurrence matrix (GLCM) P (i,j| Δx = 1, Δy = 0) is illustrated below.

0	1	1	2	3	
0	0	2	3	3	i= 0
0	1	2	2	3	1
1	2	3	2	2	2
2	2	3	3	2	3

j=0	1	2	3
1/20	2/20	1/20	0
0	1/20	3/20	0
0	0	3/20	5/20
0	0	2/20	2/20

Figure 4. 5×5 sub image and its grey level co – occurrence matrix

Using a large number of intensity levels G implies storing a lot of temporary data, i.e. a G x G matrix for each combination of (Δx, Δy) or (d,θ). One sometimes has the paradoxical situation that the matrices from which the texture features are extracted are more voluminous than the original images from which they are derived. It is also clear that because of their large dimensionality, the GLCM's are very sensitive to the size of the texture samples on which they are estimated. Thus, the number of gray levels is often reduced.

Even visually, quantization into 16 gray levels is often sufficient for discrimination or segmentation of textures. Using few levels is equivalent to viewing the image on a coarse scale, whereas more levels give an image with more detail. However, the performance of a given GLCM-based feature, as well as the ranking of the features, may depend on the number of gray levels used.

Because a G x G matrix (or histogram array) must be accumulated for each sub-image/window and for each separation parameter set (d,θ), it is usually computationally necessary to restrict the (d, θ)-values to be tested to a limited number of values. Figure 5 below illustrates the geometrical relationships of GLCM measurements made for four distances d (d = max{| Δx|,| Δy|}) and angles of 9 = 0,π/4,π/2 and 3π/4 radians under the assumption of angular symmetry.

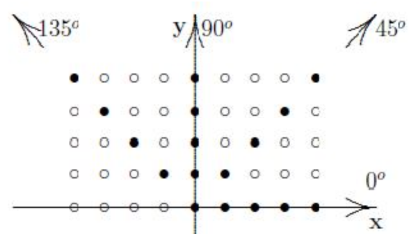

Figure 5 Geometry for measurement of gray level co-occurrence matrix for 4 distances (d) and 4 angles (θ)

In order to obtain a statistically reliable estimate of the joint probability distribution, the matrix must contain a reasonably large average occupancy level. This can be achieved either by restricting the number of gray value quantization levels or by using a relatively large window. The former approach results in a loss of texture description accuracy in the analysis of low amplitude textures, while the latter causes uncertainty and error if the texture changes over the large window. A typical compromise is to use 16 gray levels and a window of about 30 to 50 pixels on each side.

For a given distance d we usually have four angular gray level occurrence matrices. A number of scalar texture measures T(d, θ) that may be extracted from these matrices (see below). If one wants to avoid dependency of direction one may calculate an average (isotropic) matrix out of four matrices, θ = 00, 45°, 90°, 135°. In [79]they have suggested to use the angular mean, MT(d), and range, Rt(d), of each of the proposed textural measures, T, as a set of features used as input to a classifier:

$$R_T(d)=\max_{\theta}[T(d,\theta)]-\min_{\theta}[T(d,\theta)] \quad (5)$$

where the summation is over the angular measurements, and Nθ represents the number of such measurements (here 4). Similarly, an angular independent texture variance may be defined as

$$V_T^2(d)=\frac{1}{N_\theta}\sum_{\theta}\left[T(d,\theta)-M_T(d)\right]^2 \quad (6)$$

Within the large number of texture features available, some of the features are strongly correlated with each other. A feature selection procedure may be applied in order to select a subset or a linear combination of the features available, either using a set of training image regions to establish the set of features giving the smallest classification error, or using some functional feature space distance metric such that a large feature space distance implies a small classification error.

Texture Features from GLCM

A number of texture features may be extracted from the GLCM. The following notations are used:

-G is the number of gray levels used.

-M is the mean value of P.

-μx, μy, σx and σy are the means and standard deviations of Px and Py.

- Px(i) is the I then try in the marginal-probability matrix obtained by summing the rows of P(i,j).

Contrast

$$CONTRAST=\sum_{N=0}^{G-1}n^2\left\{\sum_{i=0}^{G}\sum_{j=0}^{G}P(i,j)\right\},\quad |i+j|=n \quad (7)$$

This measure of contrast or local intensity variation will favor contributions from *P(i,j)* away from diagonal, i.e. $i\neq j$.

Entropy

$$ENTROPY=\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}P(i,j)\times\log\left(P(i,j)\right) \quad (8)$$

Inhomogeneous scenes have low first order entropy, while a homogeneous scene has a high entropy.

Correlation

Correlation is a measure of gray level linear dependence between the pixels at the specified positions relative to each other.

$$CORRELATION=\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{(i\times j)\times P(i,j-\{\mu_x\times\mu_y\})}{\sigma_x\times\sigma_y} \quad (9)$$

Sum of Squares, Variance

This feature puts relatively high weights on the elements that differ from the average value of *P(i, j).*

$$VARIANCE=\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i,\mu)^2P(i,j) \quad (10)$$

Sum Average

$$AVER=\sum_{i=0}^{2G-2}iP_{x+y}(i) \quad (11)$$

Sum Entropy

$$SENT=\sum_{i=0}^{2G-2}P_{x+y}(i)\log\left(P_{x+y}(i)\right) \quad (12)$$

Difference Entropy

$$DENT=\sum_{i=0}^{G-2}P_{x+y}(i)\log\left(P_{x+y}(i)\right) \quad (13)$$

E. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is an important tool for classification. The recent vast research activities in neural classification have established that ANN is a promising alternative to various conventional classification methods. The advantage of neural networks lies in the following theoretical aspects. First, neural networks are data driven, self-adaptive methods in that they can adjust themselves to the data without any explicit specification of the functional or distributional form of the underlying model. Second, they are universal functional approximations in thatneural networks can approximate any function with arbitrary accuracy. Since any classification procedure seeks a functional relationship between the group membership and the attributes of the object, accurate identification of this underlying function is doubtlessly important. Third, neural networks are nonlinear models, which makes them flexible in modeling real world complex relationships. Finally, neural networks are able to estimate the posterior probabilities, which provide the basis for establishing classification rules and performing statistical analysis.

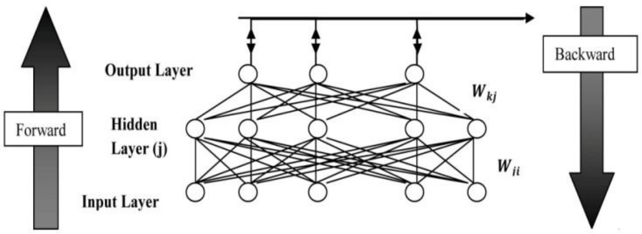

Figure 6. Structure of Artificial Neural Network

An ANN receives a set of input (X1, X2, Xn,) and the set of inputs is multiplied by a set of weights (w1, w2, wn). These weighted values are then summed and the output is passed through an activation (transfer) function. Figure 6 illustrates the structure of artificial neural networks.

IV. PERFORMANCE MEASURES

In order to evaluate the performance of the classifier, and to allow comparisons, several ratios have been taken into account.

True Positive (TP) : Test samples correctly identified.
False Positive (FP) : Test samples incorrectly identified.
True Negative (TN) : Test samples correctly rejected.
False Negative (FN) : Test samples incorrectly rejected.

From these quantities, Precision, Recall, Specificity and F-Measure are chosen as performance measures and are calculated using the following equation.

Precision is a description of a level of measurement that yields consistent results when repeated.

$$\text{Precision (P)} = 100 \times \frac{TP}{TP+FN} \quad (14)$$

Recall (also called the true positive rate or the sensitivity in some fields) measures the proportion of positives that are correctly identified.

Recall(R) =(TP/TP)+FN (15)

F-Measure is harmonics mean of precision and recall.

F-Measure (F) = 2PR/(P+R) (16)

Specificity (also called the true negative rate) measures the proportion of negatives that are correctly identified. The higher the sensitivity and specificity values, the better the procedure.

$$\text{Specificity (SP)} = 100\times \frac{TN}{TN+FP} \quad (17)$$

In order to examine the classification techniques more closely the sensitivity, specificity and accuracy value have been calculated using the confusion matrix shown in Table 1.

Table 1. Confusion matrix

Actual	Predicted	
	Positive	Negative
Positive	TP	TN
Negative	FP	FN

V. EXPERIMENTAL RESULTS

The experiment is performed on Chidambaram traffic dataset using ANN classifier. The results show that the classifier is able to identify the Car, Two Wheeler and Human. This proposed experimental results focus only on moving object detection from traffic surveillance video. Sample frames for Chidambaram traffic dataset are shown in Figure 7. The Sequences were recorded over static background with the fixed frame sizes at a rate of 25 fps. In this work, vehicles which are available in traffic video (car, bike and humans) are used for experimental purpose.

Figure 7 Sample frames from Chidambaram traffic dataset

Table 2. Chidambaram Street Dataset

DATA SET	TYPE	TIME (SEC)	FRAME SIZE	FRAME S
Chidambaram Street Dataset	Car (Four Wheeler)	30	640×360	750
	Bike (Two Wheeler)	40	640×360	1000
	Human	60	640×360	1500

The recognition results obtained by the proposed method on traffic dataset with ANN is summarized in a confusion matrix in Table 3, where correct responses define the main diagonal, the majority of objects are correctly classified, An average recognition rate of ANN is 83.66%. Table 4 shows the performance measure obtained with ANN classifiers.

Table 3. Confusion matrix for ANN classifier

Objects	Car	Two Wheeler	Human
Car	78	17	5
Two Wheeler	12	82	6
Human	5	4	91

Table 4 Performance measure obtained with ANN classifiers

Objects	Precision (%)	Recall (%)	Specificity (%)	F-Measure (%)
Car	83.31	99.23	98.53	90.58
Two Wheeler	82.42	98.45	94.62	89.72
uman	80.54	97.32	93.12	88.14
Average	82.09	98.35	95.42	89.48

Figure 8 Two wheeler and four wheeler (car) objects are detected using ANN

VI. CONCLUSION

The novel methodology is proposed for people and vehicle classification for automated surveillance systems. This work uses background subtraction model for detecting the moving objects and the object segmentation is done by the morphological operations. The GLCM features extracted from the segmented objects. The extracted features are fed to Artificial Neural Network (ANN) classifier. This approach evaluates the performance of GLCM features in Chidambaram traffic video sequences and the performance measures such as and Precision, Recall, Specificity and F-Measure were calculated. The system gives a good classification of accuracy of 83.66%.

REFERENCES

[1] Y.Ramadevi, T.Sridevi, B.Poornima, B.Kalyani, “SEGMENTATION AND OBJECT RECOGNITION USING EDGE DETECTION TECHNIQUES”, International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, December 2010.

[2] Kumar Rajpurohit, Rutu Shah, Sandeep Baranwal, Shashank Mutgi, "Analysis of Image and Video Using Color, Texture and Shape Features for Object Identification", IOSR Journal of Computer Engineering , Volume 16, Issue 6, Ver. VI (Nov – Dec. 2014), PP 29-33.

[3] Khushboo Khurana, Reetu Awasthi, "Techniques for Object Recognition in Images and Multi-Object Detection",International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 4, April 2013.

[4] Fahad Shahbaz Khan, Joost van de Weijer, Maria Vanrell,"Modulating Shape Features by Color Attention for Object Recognition", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), April 2013.

[5] K.Saranya , M.Kalaiselvi Geetha , J.Arunnehru, "Motion Detection and Tracking of Multiple Objects for Intelligent Surveillance", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727.

[6] Hongming Zhang, Wen Gao, Xilin Chen , Debin Zhao,"Object detection using spatial histogram features",Image and Vision Computing 24 (2006) 327–341

[7] Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost van de Weijer, "Color Attributes for Object Detection", Computer Vision Center, CS Dept. Universitat Autonoma de Barcelona,2007

[8] S. H. Shaikh et al., Moving Object Detection Using Background Subtraction, SpringerBriefs in Computer Science, DOI 10.1007/978-3-319-07386-6_2, 2014

[9] Osameh Biglari, Reza Ahsan, Majid Rahi, "Human Detection Using SURF and SIFT Feature Extraction Methods in Different Color Spaces", Journal of mathematics and computer Science 11 (2014) 111 – 122

[10] Luigi Di Stefano, Enrico Viarani,"Vehicle Detection and Tracking Using the Block Matching Algorithm", 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, September 2000.

[11] T. Gao, B. Packer, D. Koller, “A Segmentation-aware Object Detection Model with Occlusion Handling”, In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), June 2011.

[12] W. Hu, A.M.Gharuib, A.Hafez, “Template Match Object Detection for Inertial Navigation Systems,” Scientific research (SCIRP), pp.78-83, May 2011.

[13] E.Shectman, M.Irani, “Matching Local Self-Similarities across Images and Videos,” In IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[14] F. Khan, R. Muhammad, et.al., “Color Attributes for Object Detection,” IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp.3306 – 3313, 2012.

[15] T. Gevers, A. Smeulders, “Color-based object recognition,” Pattern Recognition, 1999.

[16] P. Viola and M. Jones, “Robust real-time object detection,” International. Journal of Computer Vision, 57(2), pp.137–154, 2004.

[17] R. Fergus, P. Perona, A. Zisserman, “Weakly supervised scale-invariant learning of models for visual recognition,” International Journal of Computer Vision, 2006.

[18] G. de Croon, “Active Object Detection,” In 2nd International conference on computer vision theory and applications (VISAPP 2007), Barcelona, Institute for Systems and Technologies of Information, Control and Communication (INSTICC), pp. 97–103, 2007.

[19] A. Berg, T.Berg , J. Malik, “Shape Matching and Object Recognition using Low Distortion Correspondences,” In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 26 – 33, 2005.

[20] H. Moballegh, N. Schmude, and R. Rojas, “Gradient Vector Griding: An Approach to Shape-based Object Detection in RoboCup Scenarios,” from: www.ais.uni-bonn.de/robocup.de/systems/RS11_Moballegh.pdf

[21] Kai She, George Bebis, Haisong Gu, and Ronald Miller, "Vehicle Tracking Using On-Line Fusion of Color and Shape Features", International Journal of Advanced Research, 2006

[22] Shalu Gupta, Yumnam Jayanta,"Region and Shape Features Based Image Retrieval", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 11, November 2015.