Combination of In Vogue Algorithms for Human Detection and Tracking

Bhuvanarjun K.M.
PG Scholar, Signal Processing,
Siddaganga Institute of Technology,
Tumakuru, Karnataka (India).

T.C. Mahalingesh
Associate Professor, Dept. of ECE,
Siddaganga Institute of Technology,
Tumakuru, Karnataka (India).

ABSTRACT
Human detection and tracking is one of the tasks in computer vision which needs a lot of understanding and appreciable effort. It has a lot of use in visual surveillance, human machine interaction, robotics and many more. A lot of algorithms have been proposed by various researchers but the problem of detection and tracking has not yet been solved efficiently. There are a lot of problems for which there exists no generic solution using a single algorithm. Hence a combination of contrastive algorithms yields a comparatively good result. This paper mainly focuses on developing an algorithm using various image processing techniques without increasing the complexity and achieving comparatively better accuracy. In this paper a new method has been proposed using combination of algorithms. The algorithms used here are Histogram of Oriented Gradients (HOG), Covariance based method and Kalman Filter. The combined algorithms yield a reasonably good result.

General Terms
Computer Vision, Covariance method, Histogram of Oriented Gradients, Kalman Filter.

Keywords
Object Detection, Object Tracking, Video Processing, Visual Surveillance.

1. INTRODUCTION
Varieties of approaches have been proposed to detect and track objects/pedestrians. The general excellence of detection and tracking algorithms depend often on video content such as occlusion and illumination. Motion picture contains an abundance of valuable information, which helps in reliable estimation of motion of the object. Errors may occur in any detection and tracking system due to algorithms or measurements. Measurement errors occur mainly when the algorithm fails to identify the new position and algorithmic error occurs due to incorrect feature extraction by the algorithm. There exist a lot of areas where detection and tracking is essential such as automotive safety, human computer interaction, surveillance, intelligent control systems, entertainment, and care for the elderly and disabled etc.

The developed system should be fast, reliable and robust for moving object detection and tracking. Varieties of problems are associated with tracking such as changing appearance patterns of the object and the scene, abrupt object motion, non- rigid object structures, camera motion and object to object occlusion and object to scene occlusion. The main aim of this paper is to construct an algorithm using various image processing techniques to reliably detect and track people in a video sequence without much complexity. In order to do so the region of interest should be extracted first and the object should be detected and it should be tracked in all frames until it disappears from the frame. It is necessary to make a note of computation time required as well. The system is developed using HOG, Covariance and Kalman Filter for efficacious detection and tracking of pedestrians. This paper tries to give best result even in case of occlusion and low illumination.

The rest of the paper is organized in the following manner: Section 2 presents the previous work carried out in the field of object detection and tracking. Section 3 deals with the proposed methodology. Section 4 explains the HOG, Covariance and Kaman filter. Experimental results are discussed in Section 5.

2. LITERATURE SURVEY
Several algorithms have been proposed by various researchers for the purpose of detection and tracking of pedestrians.

A region of interest based pedestrian detection system has been proposed which reduces the overall size of the image thus decreasing the detection time and increasing the accuracy of the algorithm for detection of pedestrian[1]. Dalal et al. [2] have proposed a reliable pedestrian detection system known as Histogram of Oriented Gradients (HOG). It is evident from various implementations that HOG outperforms all other detection algorithms and are computationally efficient [3, 4, 5]. The blend of region of interest based HOGs make the algorithm computationally simple. In [6] Oncel et al. have proposed a region descriptor, known as Covariance based method, which can be applied for detection and classification of objects. The Covariance matrices developed are a fusion of various image features which consists of spatial and statistical information [7, 8]. Pedro Cortez-Cargill et al. have discussed the accuracies of different feature vectors and also they have suggested the features which yield optimum results [9]. In [10,11] the authors have proposed an algorithm using Kalman Filter for tracking of multiple objects. The Kalman Filters are optimal estimators and it is recursive so that measurements can be processed as they arrive. Also an exact data association is necessary in order to track the objects correctly. Kalman Filter and HOG method can be combined for detection and tracking which substantially minimizes the HOG computation in each and every frame [12]. In [13], the authors have combined HOG and Optical flow for detection and tracking, and in [14], the authors have given a combined algorithm using HOG and Covariance based method for pedestrian detection thus emphasizing the scope for algorithm combination.

3. OVERVIEW OF THE PROPOSED METHODOLOGY
The overview of the proposed pedestrian detection and tracking system based on HOG, Covariance and Kalman Filter is described in Figure 1. The initial step is to perform background subtraction and some morphological operations to obtain the foreground objects. As the background and foreground objects are filtered out, region of interest (ROI) is extracted. ROI gives only the area where the objects are

potentially detected. For the extracted detection window, HOG features and Covariance matrices are calculated. The computed HOG features are then fed to a linear Support Vector Machine (SVM) which classifies the given input as either human or non-human. For the same detection window extracted, the Covariance features are also calculated simultaneously. Covariance matrix is obtained by fusing varieties of image features such as gradients, norms, coordinates etc. The dissimilarity measure is calculated using the difference between the image stored in the database and the given image. If the obtained dissimilarity score is less it indicates that there exist a desired object.

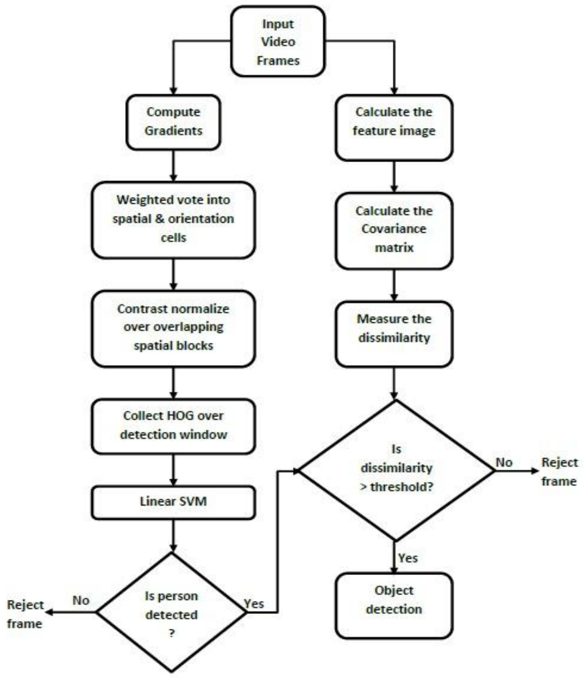


Fig 1: Overview of the proposed detection and tracking system

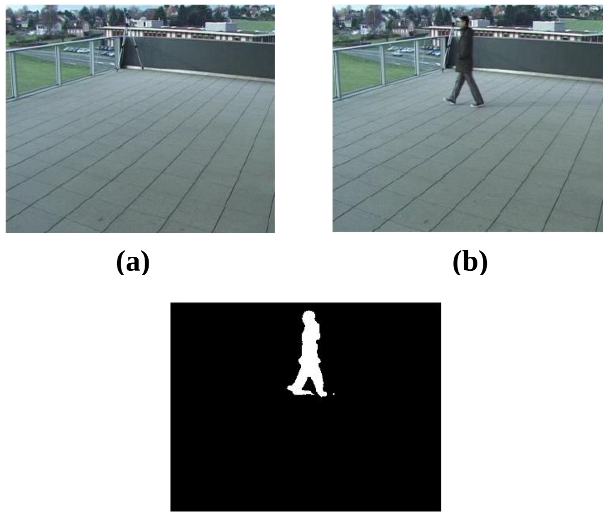
If both the algorithm outputs as human, only then the algorithm proceeds for detection. The detected values are fed to Kalman Filter for tracking.

4. DETECTION AND TRACKING SYSTEM

The work carried out in this methodology is for stationary camera. The object in the foreground has to be localized to obtain the region of interest. Background subtraction algorithm is used for foreground localization. A frame which does not have any objects in it is considered as the reference frame. The difference between two frames, reference frame and current frame, is calculated using the equation (1) given below:

$$\text{Current\_frame} - \text{Reference\_frame} > \text{Threshold} \quad (1)$$

Background subtraction results in only foreground objects. Background subtracted image will be converted into binary image using thresholding and some morphological operations as well.



(a) **(b)**

(c)

Fig 2: Background Subtraction (a) Reference frame (b) Current frame (c) Binarized image

Region of interest is calculated for the obtained binary image. The corner points obtained in both horizontal and vertical directions, which are nothing but variations in brightness levels, are the interest points in the image. To have good corners, the minimum of the two Eigen values of autocorrelation matrix of the second derivative of the image should be greater than some predefined threshold, given by

$$R = Min\ (|\lambda_1|, |\lambda_2|) \quad (2)$$

Region of interest is created by plotting the interest points over a blank frame and bounding these points, a rectangle is drawn. If the average intensity values are greater than the threshold then the coordinates of the region of the interest is mapped to the original grayscale image. The algorithm of detection is applied to this window. The coordinates of region of interest can be seen in Figure 3.

4.1 Histogram of Oriented Gradients (HOG)

HOG [2] features are best suited for pedestrian detection as they are robust and give reliable performance. The HOG features are computed only for the detection window, of size 64 x 128, not for the entire image. The detection will be moved throughout the image. In order to calculate HOG features, point derivatives are calculated along both x and y directions, i.e., in horizontal and vertical directions, 1-D derivative masks given by $M_x=[-1,0,1]$ and $M_y=[-1,0,1]^T$, are applied to the image. The derivative masks which are convolved with image I, give gradients in x and y direction.

$$G_x = I * M_x \quad (3)$$

$$G_y = I * M_y \quad (4)$$

The magnitude of gradient and angle of orientation are calculated as:

$$|G(x, y)| = \sqrt{(G_x^2 + G_x^2)} \quad (5)$$

$$\varphi(x, y) = \arctan \frac{G_y}{G_x} \quad (6)$$

The orientation bins are spaced evenly over 0^0 – 180^0 unsigned gradient values. Here 9 bins are created with 20^0 for each bin. The 64 x 128 image is divided into a grid, where each unit is called a cell which has 8 x 8 cells. Four adjacent cells correspond to a block. Magnitude and gradients are calculated for each cell. In each cell weighted votes for each orientation angle is calculated. This orientation gives the direction of majority of the edges passing through that pixel. For each cell, a 9-D histogram is calculated. All the cell histograms in a block are concatenated to create a block histogram, which is 9 x 4 = 36-D. The process is carried out for all the blocks. Each block is considered for feature extraction by moving the blocks over adjacent blocks by 50%. There will be a total of 7 horizontal blocks and 15 vertical blocks. Thus there will be 7 x 15 = 105 blocks. The corresponding figure is shown in Figure 4.

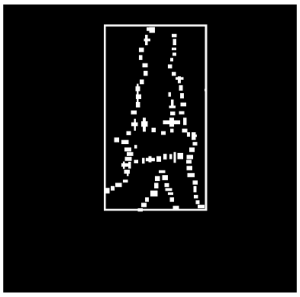


Fig 3: Region of interest around corner points

The illumination abnormalities can be easily handled by normalizing the histogram. The length of feature vector is 105 x 36 = 3780, which is obtained by concatenating all the histograms in the whole image for all the blocks. L2 norm is used for histogram normalization.

4.2 Covariance Method

Covariance method [6] is a simple and efficient method used for object detection. The covariance method provides a natural way of fusing various kinds of information such as the features and modalities of the given image. In this method the given image is represented in terms of its spatial and statistical information. Let I be the given image. Let F be the feature image with W x H x d dimension, extracted from I. F is given as:

$$\mathrm{F(x, y)} = \phi(\mathrm{I, x, y}) \tag{7}$$

where the mapping function can be color, norm, coordinates etc which is denoted by ϕ.

For a given region of rectangular shape $\mathrm{R} \subset \mathrm{F}$, let $\{z_k\}_{k=1,2...n}$ be the d-dimensional feature points inside R. The region R with d x d matrix of covariance of the feature points is given as:

$$\mathrm{C_R} = \frac{1}{n-1}\sum_{k=1}^{n}(z_k - \mu)(z_k - \mu)^T \tag{8}$$

where μ represents the mean of the points.

Since the Covariance matrices do not lie on Euclidean space, frequently used machine learning algorithms do not hold good. Nearest neighbor algorithm is used to calculate the distance between the feature vectors. The distance metric used to estimate the dissimilarity between the two covariance matrices is given by:

$$\rho(\mathrm{C_1, C_2}) = \sqrt{\sum_{\mathrm{i=1}}^{\mathrm{n}} \ln^2 \lambda_\mathrm{i}(\mathrm{C_1, C_2})} \tag{9}$$

where $\{\lambda_i(\mathrm{C_1, C_2})\}_{i=1,2...n}$ are the generalized eigen values of $\mathrm{C_1, C_2}$, computed from

$$\lambda_\mathrm{i}\mathrm{C_1 x_i} - \mathrm{C_2 x_i} = 0 \quad \mathrm{i} = 1...\mathrm{d} \tag{10}$$

and $\mathrm{x_i} \neq 0$ are the generalized eigen vectors.

The covariance matric has only (d^2 + d) / 2 different values, which signifies very low dimensionality. The non-diagonal values represent the correlation and the diagonal values represent the variance of each feature. In this paper we have used 7-D feature vector using coordinates, and norm of first and second order derivatives of the intensities using equation (7).

4.3 Kalman Filter

Kalman filter [10] is an optimal estimator, which uses a series of measurements acquired over time to produce the estimates of the unknown variables that tend to be more decisive than the ones based on single measurement. The name Kalman Filter is given because it finds the best estimates from noisy data inputs thus filtering out the noise. A tracking system is considered, in Kalman Filter approach, where x_k is the state vector represented by the object, which behaves dynamically. The subscript k represents the discrete time instant. The main objective here is to estimate x_k using the measurement z_k. The Kalman filter description is given below:

1) Process equation

$$x_k = \mathbf{A}x_{k-1} + w_{k-1} \tag{11}$$

where the transition matrix is given by $\boldsymbol{A}$ and x_k the state at time k-1 to k. Vector w_{k-1} is the Gaussian process noise $N(.)$ with the following normal probability distribution $p(w)$

$$p(w) \sim N(0, Q) \tag{12}$$

2) Measurement equation

$$z_k = Hx_k + v_k \tag{13}$$

where z_k is the measurement observed at time k-1 to k and measurement matrix is given H. v_k is the Gaussian measurement noise N(.) with normal probability distribution p(v)

$$p(v) \sim N(0, R) \quad (8) \tag{14}$$

3) Time update equations

Apriori estimate of state $\hat{x}_k^-$, and $\hat{x}_k$ covariance error estimate is obtained for the next time step k.

$$\hat{x}_k^- = A\hat{x}_{k-1} + w_k \tag{15}$$

(16)

$$P_k^- = AP_{k-1}A^T + Q$$

4) Measurement update equations

These equations are affiliated with system feedback. The main objective is aposteriori estimation. The corresponding equations are given below:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad (17)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H\hat{x}_k^-) \quad (18)$$

$$P_k = (1 - K_k H) P_k^- \quad (19)$$

Using the above measurement update equations, Kalman gain K_k is calculated. Also using the measurement z_k the aposteriori error estimate P_k and aposteriori state estimate $\hat{x}_k$. The new aprior estimate is predicted using the previous aposteriori estimate, calculated recursively using time and measurement equations. The main highlight of Kalman filter is the recursive behavior of estimating states.

5. EXPERIMENTAL RESULTS

In this section the experimental results of the proposed detection and tracking methodology has been discussed. The proposed algorithm is implemented using OpenCV 2. 4. 3 on Microsoft Visual Studio 2010 IDE. The work is carried using Second Generation Intel Core i3 CPU @ 2.13 GHz processor and 3GB RAM on a personal computer. The frame size used is 320 x 240 at 25 frames per second. The results obtained have proven to be satisfactory. The average time taken for detection is 120 ms using the proposed methodology and is sufficient for real time implementation. The result of the proposed detection system is as shown in Figure 5. The detection rate can be slightly increased in low illumination conditions using a 1-D kernel given by [-1, 1, 2]. This kernel gives a washed out look but it enhances the edge pixels. The result is shown in Figure 6. The detection rate of the proposed algorithm is provided in Table 1 where detection rate is calculated as the ratio of number of detections to the actual total number of detections. The proposed methodology shows the improvement in detection rate as a result of combination of contrastive algorithms.

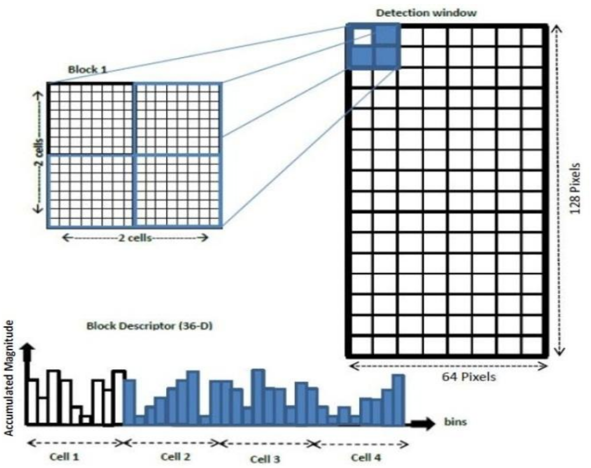


Fig 4: HOG computation block descriptor

Fig 5: Result of the proposed methodology

Fig 6: Result of proposed methodology using 1-D kernel

6. CONCLUSION

In this paper a pedestrian detection and tracking system using HOG, Covariance and Kalman Filter is implemented. The system is developed with minimum complexity. The developed system gives good performance for detection and tracking. The main contribution of this paper is combination of algorithms which reduces the false detection rate and improves the detection rate. The correctly detected persons are tracked in subsequent frames successfully. The experimental results suggest that the results obtained in detection and tracking are satisfactory.

The work can be still extended for moving cameras and for highly occluded and cluttered areas. The method can also be tried by replacing Kalman filter with Particle filter.

Table 1: Detection rates calculated for various detection algorithms using local database

Detection Algorithm	Detection Rate (in %)
HOG	90
Covariance method	89
HOG + Covariance method	92

7. REFERENCES

[1] Na Shoul, Hong Peng, Hui Wang, Li-Min Meng and K.-L. Du, "An ROIs Based Pedestrian Detection System for Single Images", *IEEE 5th International Congress on Image and Signal Processing*, 2012.

[2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *IEEE Conference on Computer Vision and Pattern Recognition,* vol. 1, pp. 886 - 893, 2005.

[3] M. Ishikawa et aI, "Selection of Histograms of Oriented Gradients Features for Pedestrian Detection", *ICONIP 2007, Part II, LNCS 4985, pp. 598-607,springer-V*erlag Berlin Heidelberg, 2008.

[4] P. Dollar, C. Wojek, B. Schiele, P. Perona, "Pedestrian detection: an evaluation of the state of the art," IEEE Trans. Pattern Anal. Machine Intell., vol. 34, pp. 743-761, April ,2012.

[5] Daimeng Wei, Yong Zhao, Ruzhong Cheng and Guoliang Li, "An Enhanced Histogram of Oriented Gradient for Pedestrian Detection", *Fourth International Conference on Intelligent Control and Information Processing (ICICIP),* Beijing, China, June 2013.

[6] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification", *In Proc. 9th European Conf. on Computer Vision*, Graz, Austria, volume 2, pages 589-600, 2006.

[7] Yaping Liu, Jian Yao, Renping Xie, and Sa Zhu, "Pedestrian Detection from Still Images Based on Multi-Feature Covariances", *Proceeding of the IEEE International Conference on Information and Automation*, Yinchuan, China, August ,2013.

[8] Prajna Parimita Dash, Santhosh Aitha and Dipti Patra," Ohta Based Covariance Technique for Tracking Object in Video Scene", *IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS),* Bhopal, India, Pages 1-4, March,2012.

[9] Cargill P. C, Rius C. U, Domingo Mery and Alvaro Soto, "Performance Evaluation of the Covariance Descriptor for Target Detection", *International Conference of the Chilean Computer Science Soceity (SCCC),* Santiago, Chile, , pages 133 – 141, Nov 10-12, 2009.

[10] Xin Li, Kejun Wang, Wei Wang, Yang Li, "A Multiple Object Tracking Method Using Kalman Filter", *Proceedings of 2010 IEEE International Conference on Information and Automation,* Harbin, China, June 20-23, 2010.

[11] Jong-Min Jeong, Tae-Sung Yoon, and Jin-Bae Park, "Kalman Filter Based Multiple Objects Detection-Tracking Algorithm Robust to Occlusion", *SICE Annual Conference*, Hokkaido University, Sapporo, Japan, September 9-12, 2014.

[12] Changyan Li and Lijun Guo, Yichen Hu, "A New Method Combining HOG and Kalman Filter for Video-based Human Detection and Tracking", *3rd IEEE conference on International Congress on Image and Signal Processing*-2010

[13] .Dayanand Kumar N. C, K. V Suresh, "HOG-PCA Descriptor with Optical Flow based Human Detection and Tracking", International Conf. on Communication and Signal Processing, April 3-5, 2014.

[14] Bhuvanarjun. K. M, T. C. Mahalingesh, "Pedestrian Detection in a video sequence using HOG and Covariance method", *International Conference on Recent Innovations in Science, Engineering and Management*,(IJEEE), New Delhi, vol. 7, issue 1,2015.