# Object Detection and Tracking in Discriminant Subspace

**Amith. R[1], V.N. Manjunath Aradhya[2]**
[1]Research Scholar, Department of Computer Science and Engineering, Jain University, Bengaluru, India.
[2]Associate Professor, Department of Master of Computer Applications, Sri Jayachamarajendra College of Engineering, Mysuru, India.

---

**How to cite this paper:**
**R, Amith., & Aradhya, V**. (2017). Object Detection and Tracking in Discriminant Subspace. *IRA-International Journal of Technology & Engineering* (ISSN 2455-4480), 7(3), 62-75. doi:http://dx.doi.org/10.21013/jte.v7.n3.p2

---

**ABSTRACT**

*Detection and tracking of moving objects in video is essential for many computer vision applications and it is considered as a challenging research issue due to dynamic changes in background, illumination, object size and shape. Many traditional algorithms fails to detect and track the moving objects accurately, this paper proposes a robust method, to detect and track moving objects based on the combination of background subtraction and Orthogonalized Fisher's Discriminant (OFD). Background subtraction detects the foreground objects on subtracting frame by frame basis and updating the background model recursively. Orthogonalized Fisher's Discriminant projects high dimensional data onto a one dimensional space with the highest recognizability, which speedup the detection and tracking process and also preserves the structure of the objects resulting high accuracy. The proposed method is tested on standard datasets with complex environments and experimental results obtained are encouraging.*

## 1. Introduction

Object detection in videos is a highly progressive and challenging research area of computer vision for long time due to its day to day importance for various video analysis applications [14] like: Surveillance and security to perceive individuals, to contribute exceptional sense of security using evident data; Video abstraction to obtain short summary of videos, to produce object based outline; Traffic monitoring to inspect stream, to identify accidents; Video editing to eradicate unwieldy individual administrative interplay, to shape up innovative video effects; Intuitive games to offer common way of interplay with perceptive systems. Object detection and tracking in video is a highly progressive and challenging research area of computer vision due to day to day importance for various video analysis applications such as human machines interaction, traffic monitoring, video surveillance and indexing of video data. Detection and tracking can be characterized as the issue of evaluating the movement of an object in the image plane. The object detection and tracking are intently associated in light of the fact that tracking normally begins with detecting the objects, although detecting an object over and over again in consecutive video sequence is usually significant to benefit and verify tracking. Challenges in detection and tracking moving objects include partial and full occlusion handling, complex background, small/distant objects with low contrast, illumination problems, and noisy videos.

Background subtraction approach suffers from false alarms mainly because of illumination changes and dynamic background [17, 21, 23]. Background subtraction based moving object detection methods heavily depends on the background model. A good background model should be robust to illumination changes, weather conditions, etc. Many background modeling methods are proposed in literature, Basic Background Modeling makes use of average or median or analysis of histogram over time. Statistical Background Modeling models the background using Single Gaussian [1] or Mixture of Gaussians [2] or Kernel Density Estimation [3]. Fuzzy Background Modeling makes use of fuzzy mixture of gaussians, In Background Clustering each pixel in the video frame is represented in terms of clusters and are matched against corresponding cluster group. Classification of cluster is done based on whether the cluster is matched with the cluster which is part of background; usually K-means algorithm is used for clustering foreground and background pixels. Neural Network Background Modeling learns to classify pixels as background and foreground by calculating the mean of weights of neural network, which usually represents the background. Background model in Wavelet Background Modeling is defined in the temporal domain which preserves the information and is retained in the wavelet coefficients. The multi-scale property of discrete wavelet transform delivers detail parameters of each frame decomposed. And Background Estimation estimates the background and calculates if there are any slight deviations of pixels in the current frame when compared with the predicted value, then it is considered as foreground. As seen in the literature, statistical background methods are robust to the dynamic background and

illumination changes. Statistical background modeling methods are categorized into three categories [19], First category pertains Single Gaussian [1], Mixture of Gaussian [2], Kernel Density Estimation [3] and Principal Components Analysis [3]. The second category uses support vector models: Support Vector Machine [9], Support Vector Regression [11] and Support Vector Data Description [17]. Third category consists of Single General Gaussian [16], the Mixture of General Gaussians [12] and subspace learning using Independent Component Analysis [13], Incremental Non-negative Matrix Factorization [15],Linear Discriminant Analysis [8] and Canonical Correlation Analysis [9].

In literature, lots of work has been done using PCA and Fisher's Linear Discriminant Analysis (FLDA). It shows the difference in working procedure of PCA and FLDA. The FLDA seeks a projection vector that maximizes Fisher's ratio whereas the PCA searches the projection vector that has highest eigen value which has higher discriminative power for separating two classes [2, 10, 18]. Both PCA and FLDA are linear transformation technique used for dimensionality reduction. Being unsupervised nature, PCA ignores class labels but focus on directions that considers maximum variance. In contrast, the FLDA is supervised technique that finds the directions (*i.e.* linear discriminant) [6, 22]. These directions represent the axes that could maximize the separation between classes. According to Martinez *et. al.* [6], FLDA is superior to PCA for classification of different multiclass problem where class labels are known. It is to be noted that, PCA computes the component axis which has maximum variance for whole dataset. But in our proposed method, FLDA tries to maximize the component axis for best class separability.

## 2. Proposed Method

### 2.1 Background Subtraction

Background subtraction algorithm detects the foreground object in a video by calculating the difference between the current image and the background image or often called reference image or background model. Background model is constructed using initial few frames. An averaging method is applied to construct background model that has updated recursively. This background model (BGM) is constructed as:

$$B_N(x,y) = B_{N-1}(x,y) + \frac{F_N(x,y)}{T} \tag{1}$$

Where, $B_N(x,y)$ is the intensity value of pixel at location $(x,y)$ in background model. $F_N(x,y)$ is the intensity value of pixel at location $(x,y)$ in $i^{th}$ frame.$N$ is the total number of frames used for background model construction, where *N* lies between *1* to *T* where, *T=200*. The pixels constituting the regions undergoing change are marked for further processing and morphological filters are applied to remove noise and classify pixels for object detection in video.

### 2.2 Orthogonalized Fisher Discriminant Analysis

Fisher Linear Discriminant Analysis (FLDA) maps high dimensional dataset onto one dimensional space with the greatest discrimination and the new subspace obtained affirms maximum between classes distance and minimum within classes distance.

Fisher Linear Discriminant Analysis is defined as,

$$J(\phi) = \frac{\phi^T S_B \phi}{\phi^T S_W \phi} \tag{2}$$

Where $\phi$ represents n-dimensional vector, $S_B$ denotes the between classess catter matrix and $S_W$ represents the within classes scatter matrix. Scatter matrix $S_B$ and $S_W$ are represented as Eq. (3) and Eq. (4) respectively.

$$S_B = \sum_{i=1}^{c} n_i (u_i - \bar{X})(u_i - \bar{X})^T \tag{3}$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in class\ i}^{c} n_i (u_i - x_k)(u_i - x_k)^T \tag{4}$$

Where $\bar{X}$ denotes the absolute mean of the sample cases, $u_i$ denotes the mean of the $i$ class, $c$ is the number of the classes, $n_i$ is the number of samples belongs to the $i$ class and $x_k$ is the k-th sample.

From Eq. (2) it is known that when $J(\phi)$ is maximized, it gives $\phi$ which is the optimal projection vector $W_{opt}$ and $W_{opt}$ is computed by calculating eigenvectors and eigenvalue as shown in Eq. (5).

$$[V, D] = eig(S_W^{-1} S_B) \tag{5}$$

where $V$ and $D$ are the eigenvalue and eigenvector of $S_W^{-1} S_B$ respectively. Then $W_{opt}$ is calculated by arranging the eigenvector according to their maximum eigen value. The Fisher classifier $F(X)$ is defined as Eq. (6).

$$F(X) = \begin{cases} class_1, & (W_{opt})^T X > \theta \\ class_2, & (W_{opt})^T X < \theta \end{cases} \tag{6}$$

Where $X$ is the input sample, and $\theta$ is the classifier threshold defined in Eq. (7).

$$\theta = \frac{n_1 \widetilde{u_1} + n_2 \widetilde{u_2}}{n_1 + n_2} \tag{7}$$

FLDA has major shortcomings; feature extraction in FLDA is optimal when only few projection vectors are used for that purpose. When huge number of projection vectors is considered, the classification performance of FLDA will be significantly depreciated. This is because, when more number of projection vectors are used, there will be redundancy among the fisher discriminant vectors. Selecting optimal number of projection vectors is highly subjective task and it will reflect on the recognition accuracy of samples. However in order to overcome the drawbacks, a simple yet effective process called orthogonalization is used, which removes the redundancy among the discriminant vectors by following Gram-Schmidt decomposition process as described below.

Suppose $u_1, u_2, \ldots, u_d$ are linearly independent Fisher's discriminant vectors. Then the Gram-Schmidt orthogonalization process uses the vectors $u_1, u_2, \ldots, u_d$ to construct the corresponding orthogonalized projection vectors $v_1, v_2, \ldots, v_d$. Let $v_1$ be $u_1$ and assume that k vectors $v_1, v_2, \ldots, v_k (1 \leq k \leq d - 1)$ have already been calculated. Then $(k + 1)^{th}$ orthogonalized vector $v_{k+1}$ is calculated as:

$$v_{k+1} = u_{k+1} - \sum_{i=1}^{k} \frac{u_i^T u_{k+1}}{v_i^T v_i} v_i \tag{8}$$

Instead of using original discriminant vectors $u_1, u_2, \ldots, u_d$ for feature extraction, we use discriminant vectors $v_1, v_2, \ldots, v_d$ which are orthogonal and this step is used as a preprocessing step before extracting the features obtained from fisher linear discriminant. Connected component analysis is used to track the foreground mask utilizing the spatio-temporal information and features extracted from Orthogonalized Fisher's Discriminant (OFD) of the detected object.

### 3. Experimental Results and Discussion

To assess the performance of proposed method, openly accessible standard PETS, CAVIAR, ViSOR, iLIDS and other datasets are used and each videos consists of 500 frames (Average). The frequency of complete occlusion among the objects in video is often and the challenging feature of these datasets is detecting and tracking the objects in low contrast, complex background and varying light conditions. From Figure 3.1 to 3.6 we can see that, objects are detected and tracked successfully with various video sequences, which consists of distant/small objects, illumination variances, low contrast sequences, complex background, noisy video sequences, occlusion, etc., both in indoor and outdoor environment. An efficient background subtraction method is proposed to detect foreground objects by calculating the difference between the current image and the background image, and the background model will be updated accordingly. Due to this, any environmental changes and complex background will not affect the detection process. The Orthogonalized Fisher's projection maximizes the scatter between mean values of two different classes, while variance of each class is minimized. This literally speedup the detection and tracking process and also improves accuracy. The proposed method outperforms the traditional PCA, Locality preserving projects (LPP) and kalman filter method as demonstrated in Table I and Table II. As shown in figure 3.1 to 3.6 Image sequences are tested with the proposed method and traditional PCA approach by detecting multiple objects (distant/small) including car, humans and bicycle in both indoor and outdoor environment under complex background. We can see that in most of the cases the results obtained from traditional PCA and LPP based approach fails to preserve the object structure which leads to failure in future steps like erroneous tracking and recognition. Also the false positive rate is high in PCA and LPP, when there is no object in the frame. False Alarm Rate, Detection Rate, Accuracy, Specificity, False Positive Rate and False Negative Rate achieved from the proposed method, traditional PCA and locality preserving projections is demonstrated in Table I.

As shown in the figure 3.7, Image sequences from PETS dataset are tested with the proposed method and Kalman filter based method by tracking multiple objects (distant/small) including car, humans and bicycle in outdoor environment under complex background, figure 3.8 shows, successfully tracking multiple moving objects (vehicles) even though trees swings in highway under sunny day condition, figure 3.9 shows tracking a person walking in outdoor environment with low contrast and handling occlusion in complex background, figure 3.10 shows tracking a person walking in indoor environment with different light conditions and complex background, figure 3.11 shows tracking moving objects including humans walking and bicycle in outdoor environment under bright sunny/shady weather condition, figure 3.12 shows tracking small/distant moving objects in the indoor environment under complex background. Also we can see that results obtained from kalman filter method have more false positives when compared with proposed method and in most of the cases bounding boxes are overlapped with other objects while tracking using kalman filter based method. Performance analysis of the proposed method and kalman filter method is demonstrated in Table II.
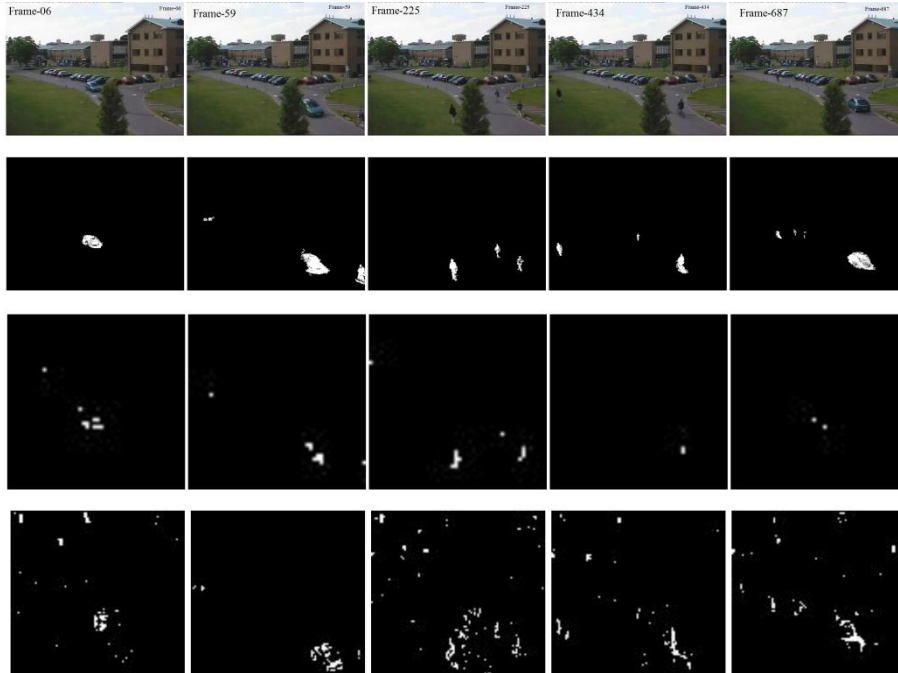
Figure 3.1: 1st Row: Original Sequence; 2nd Row: Detection Results of Proposed Method; 3rd Row: Detection Results of Traditional PCA Approach; 4th Row: Detection Results of LPP.
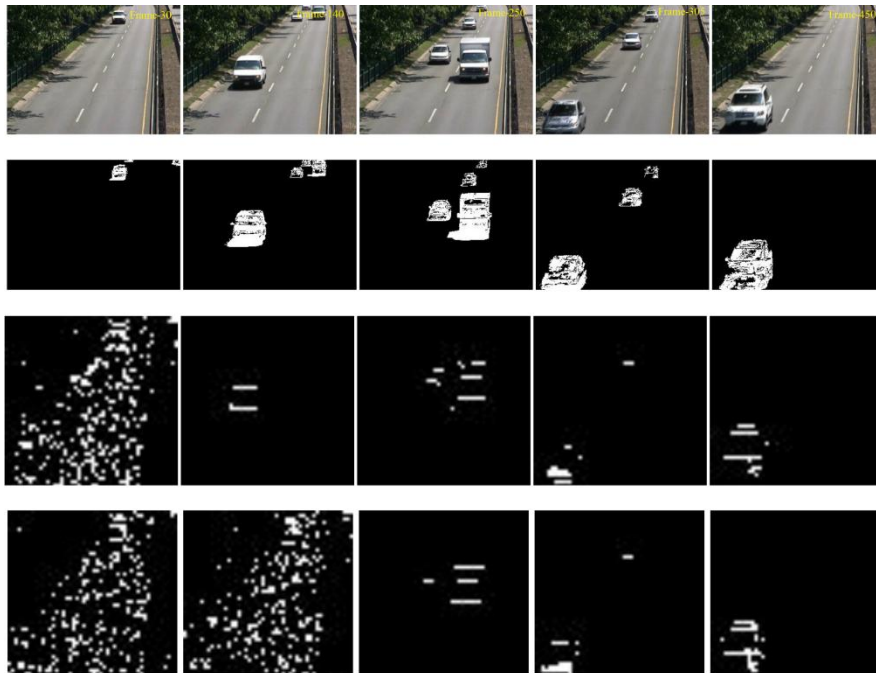


Figure 3.2: 1st Row: Original Sequence; 2nd Row: Detection Results of Proposed Method; 3rd Row: Detection Results of Traditional PCA Approach; 4th Row: Detection Results of LPP.
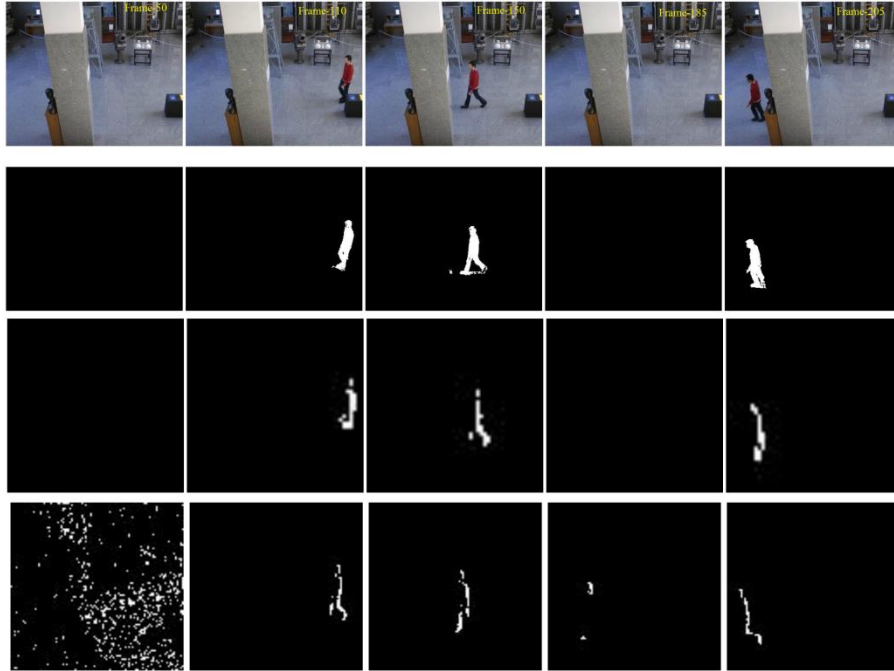
Figure 3.3: 1st Row: Original Sequence; 2nd Row: Detection Results of Proposed Method; 3rd Row: Detection Results of Traditional PCA Approach; 4th Row: Detection Results of LPP.



Figure 3.4: 1st Row: Original Sequence; 2nd Row: Detection Results of Proposed Method; 3rd Row: Detection Results of Traditional PCA Approach; 4th Row: Detection Results of LPP.

Figure 3.5: 1<sup>st</sup> Row: Original Sequence; 2<sup>nd</sup> Row: Detection Results of Proposed Method; 3<sup>rd</sup> Row: Detection Results of Traditional PCA Approach; 4<sup>th</sup> Row: Detection Results of LPP.
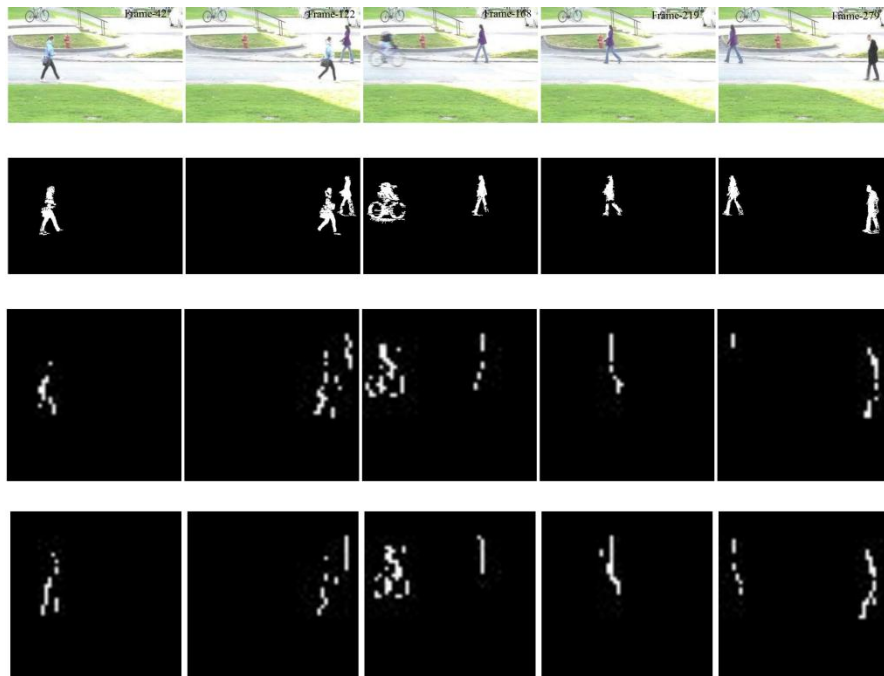


Figure 3.6: 1<sup>st</sup> Row: Original Sequence; 2<sup>nd</sup> Row: Detection Results of Proposed Method; 3<sup>rd</sup> Row: Detection Results of Traditional PCA Approach; 4<sup>th</sup> Row: Detection Results of LPP.
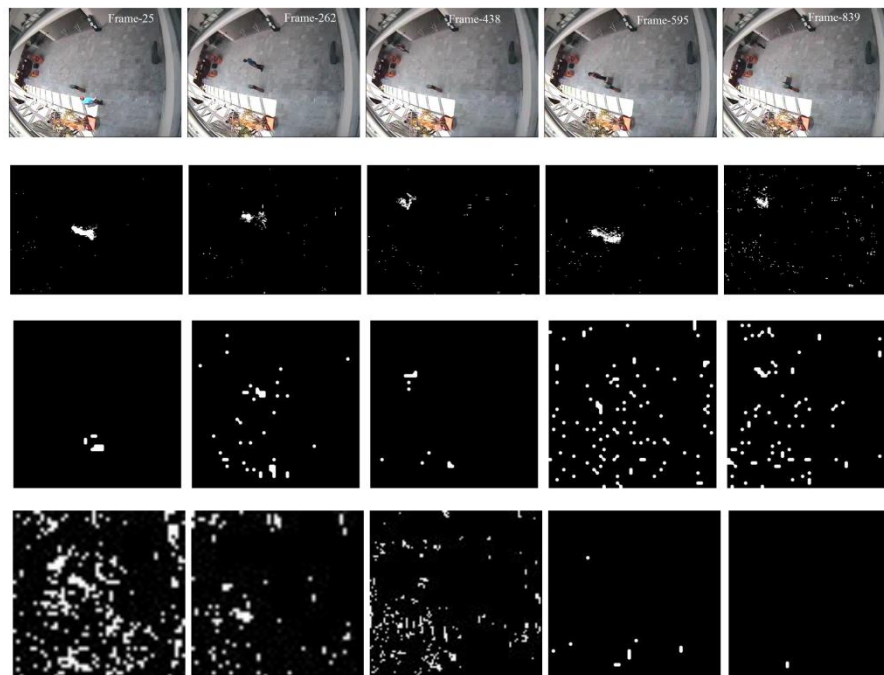
Figure 3.7: 1$^{st}$ Row: Original frames; 2$^{nd}$ Row: Tracking results of kalman filter;3$^{rd}$ Row: Results of proposed method tracking multiple objects (small/distant) including car, humans and bicycle in outdoor environment.
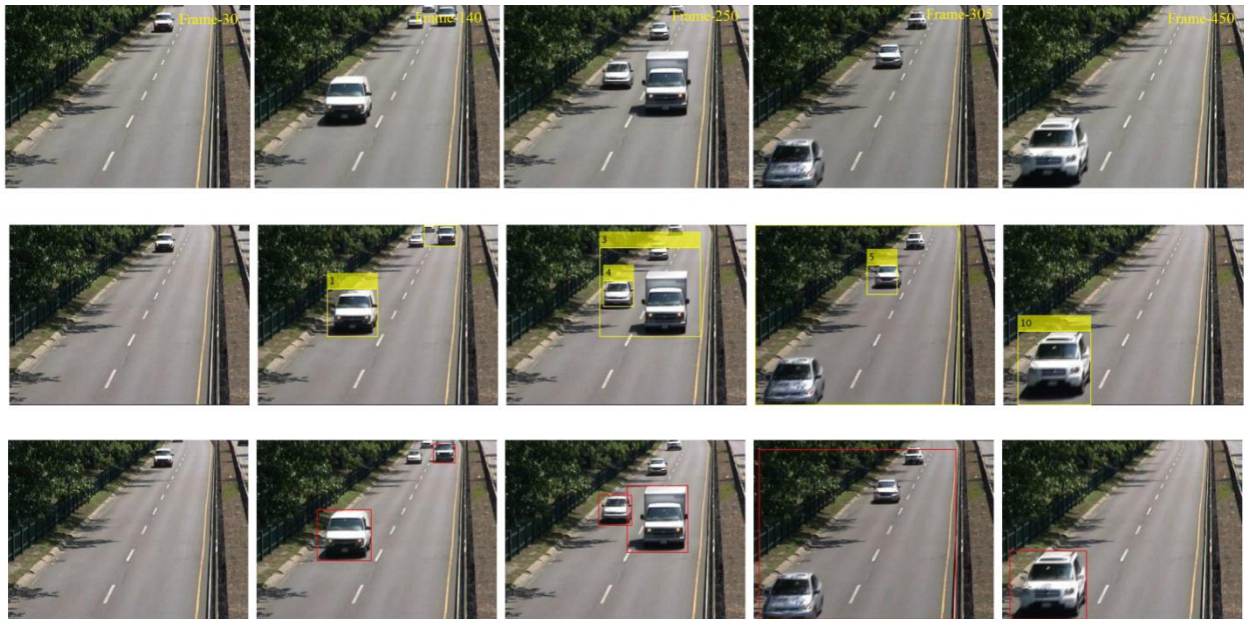


Figure 3.8: 1$^{st}$ Row: Original frames; 2$^{nd}$ Row: Tracking results of kalman filter; 3$^{rd}$ Row: Results of proposed method tracking multiple moving objects (vehicles) in highway under sunny day condition and complex background
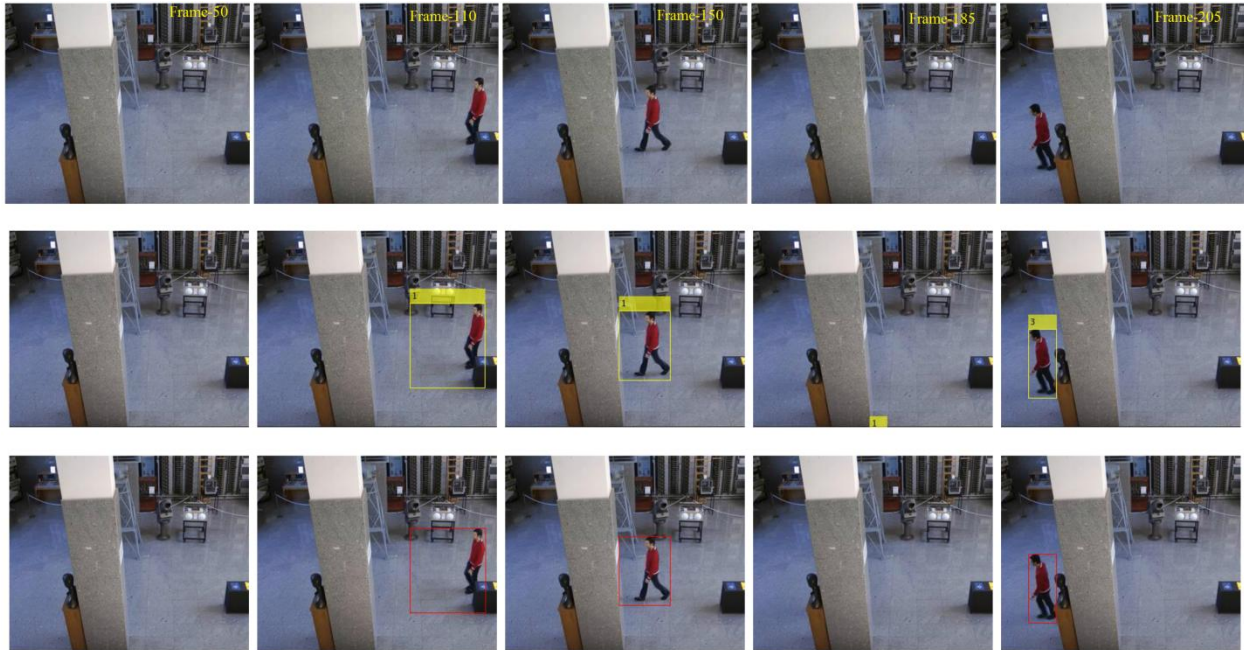
Figure 3.9: 1$^{st}$ Row: Original frames; 2$^{nd}$ Row: Tracking results of kalman filter; 3$^{rd}$ Row: Results of proposed method tracking a person walking in low contrast outdoor environment by handling occlusion under complex background.



Figure 3.10: 1$^{st}$ Row: Original frames; 2$^{nd}$ Row: Tracking results of kalman filter; 3$^{rd}$ Row: Results of proposed method tracking a person walking in indoor environment under complex background.

Figure 3.11: 1$^{st}$ Row: Original frames; 2$^{nd}$ Row: Tracking results of kalman filter; 3$^{rd}$ Row: Results of proposed method tracking multiple moving objects including humans and bicycle in outdoor environment under bright sunny and shady condition.
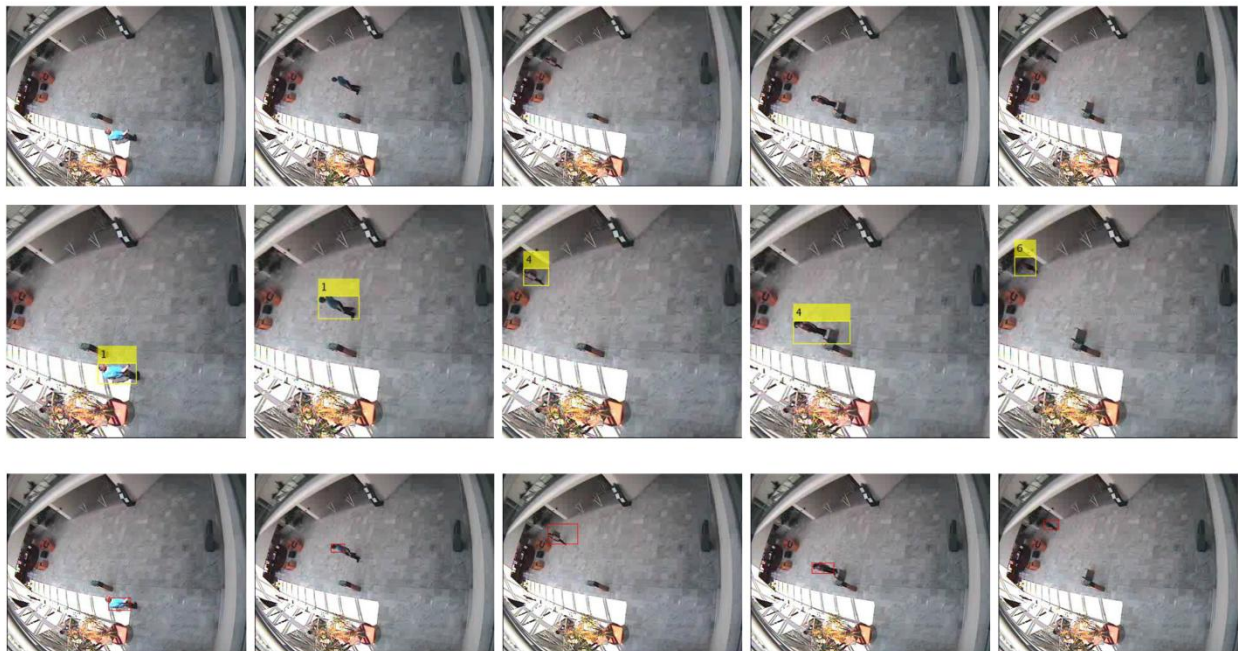


Figure 3.12: 1$^{st}$ Row: Original frames; 2$^{nd}$ Row: Tracking results of kalman filter; 3$^{rd}$ Row: Results of proposed method tracking small/distant moving objects in indoor environment under complex background.

Table I: Performance analysis of object detection

| Metrics | Proposed Method | Traditional PCA | Locality Preserving Projections (LPP) |
|---|---|---|---|
| Detection Rate (TP/(TP+FN)) | 98.22% | 73.26% | 82.80 |
| Accuracy ((TP+TN)/TF) | 89.85% | 69.29% | 78.27 |
| Specificity (TN/(FP+TN)) | 85.82% | 45.24 | 60.60 |
| False Alarm Rate (FP/(TP+FP)) | 12.80% | 38.28% | 26.50 |
| False Positive Rate (FP/(FP+TN)) | 14.18% | 54.76% | 39.40 |
| False Negative Rate (FN/(FN+TP)) | 12.70% | 25.82% | 25.44 |

Note: TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative; TF: Total number of frames in the video; TG: Total Ground truth object frames.

Table II: Performance analysis of object tracking

| Metrics | Proposed Method | Kalman Filter |
|---|---|---|
| Tracker Detection Rate (TRDR) (TP/TG) | 87.76% | 77.48% |
| Accuracy ((TP+TN)/TF) | 87.85% | 76.17% |
| Specificity (TN/(FP+TN)) | 84.38% | 70.82 |
| False Alarm Rate (FP/(TP+FP)) | 11.10% | 42.56% |
| False Positive Rate (FP/(FP+TN)) | 15.62% | 29.18% |
| False Negative Rate (FN/(FN+TP)) | 14.88% | 21.15% |

## 4. Conclusion

Arobust method to detectand track moving objects in video is proposed grounded on the background subtraction and Orthogonalized Fisher's Discriminant Analysis (O*FD*). The proposed method detects the foreground objects using background subtraction method which reduces the effect of noise and variations in the background. O*FD* is proposed to improve the accuracy and speed of detection and tracking process. The proposed method is tested with publicly available standard PETS, CAVIAR, iLIDS dataset and other video sequences by considering challenges like illumination variance, low contrast sequences, occlusion, small/distant objects and complex background, which achieves fast and accurate detection and tracking of objects. Also proposed method outperforms traditional PCA and LPP method by effectively preserving the shape of objects which is vital in accurate tracking as demonstrated and proves an effective method for detection and tracking of moving objects. But the proposed method is prone to shadows and our future work focuses on tracking objects by eliminating shadows.

### References

[1] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, Pfinder: Real-time tracking of the human body, IEEE Trans on Pattern Analysis and Machine Intelligence 19 (7) (1997) 780–785.

[2] L. Li, J. Gong, and W. Chen, "Gray-Level Thresholding based on Fisher Linear Projection of Two - Dimensional Histogram", Pattern Recognition, Elsevier, vol. 30, issue 5, pp. 743-749, May, 1997.

[3] C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, in: IEEE Conf on CVPR, 1999, pp. 246–252.

[4] A. Elgammal, D. H. D, L. Davis, Non-parametric model for background subtraction, in: 6th European Conf on Computer Vision, 2000, pp. 751–767.

[5] N. M. Oliver, B. Rosario, A. Pentland, A bayesian computer vision system for modeling human interactions, IEEE Trans on Pattern Analysis and Machine Intelligence 22 (8) (2000) 831–843.

[6] M. Martinez, and A. C. Kak, "PCA versus LDA", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, issue 2, pp. 228–233, Feb. 2001.

[7] H. Lin, T. Liu, J. Chuang, A probabilistic svm approach for background scene initialization, in: International Conference on Image Processing, 2002, pp. 893–896.

[8] Morellas V., Pavlidis L., Tsiamyrtzis P. DETER: detection of events for threat evaluation and recognition, Machine Vision and Applications, Volume 15, (2003), 29-45.

[9] Lee D. Online Adaptive Gaussian Mixture Learning for Video Applications, ECCV Workshop on Statistical Methods for Video Processing, Prague, Czech, (2004).

[10] O. Veksler, "Notes on Pattern Recognition, CS434a/541a", University of Western Ontario, Lecture 8, Oct. 2004.

[11] J. Wang, G. Bebis, R. Miller, Robust Video-based Surveillance by Integrating Target Detection with Tracking, in: IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in conjunction with CVPR, (2006), 137–142.

[12] A. Tavakkoli, M. Nicolescu, G. Bebis, A novelty detection approach for foreground region detection in videos with quasistationary backgrounds, in: ISVC, 2006, pp. 40–49.

[13] M. Yamazaki, G. Xu, Y. Chen, Detection of moving objects by independent component analysis, in: ACCV, (2006), 467–478.

[14] A. Yilmaz, O. Javed, M. Shah, "Object Tracking: A Survey," ACM Computing Surveys, vol. 38, pp. 1-45, 2006.

[15] S. Bucak, B. Gunsel, O. Gursoy, Incremental Non-negative Matrix Factorization for Dynamic Background Modelling, International Workshop on Pattern Recognition in Information System, (2007).

[16] H. Kim, R. Sakamoto, I. Kitahara, T. Toriyama, K. Kogure, Robust Silhouette Extraction Technique using Background Subtraction, 10th Meeting on Image Recognition and Understand (MIRU), (2007).

[17] M. Allili, N. Bouguila, D. Ziou, A robust video foreground segmentation by using generalized gaussian mixture modelling, in: Fourth Canadian Conf on Computer and Robot Vision (CRV), 2007, pp. 503–509.

[18] G. A. Azim, and Z. A. Abo-Eleneen, "Thresholding based on Fisher Linear Discriminant", Journal of Pattern Recognition Research, vol. 6, no. 2, pp. 326-334, June, 2011.

[19] Thierry Bouwmans. Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey. Recent Patents on Computer Science, Bentham Science Publishers, 2011, 4 (3), 147-176.

[20] V. Reddy, C. Sanderson, and B. Lovell, "Improved foreground detection via block-based classifier cascade with probabilistic decision integration", IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 1, pp 175-181, Jan. 2013.

[21] T. S. F. Haines and Tao Xiang, "Background Subtraction with Dirichlet process mixture model", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 4, pp. 670-683, April, 2014.

[22] S. Raschka, 'http://sebastianraschka.com/Articles/2014_python_lda.html#a-comparison-ofpca-and-lda', Notes on FLDA, 2014.

[23] P. St-Charles, G. Bilodeau, R. Bergevin,SuBSENSE: A Universal Change Detection Method with Local Adaptive Sensitivity', IEEE Transactions on Image Processing, vol. 24, issue 1, pp. 359-373, Jan. 2015.