

Humanrecognition Using ‘Pso-Ofa’ In Low Resolution Videos

K.Ranga Narayana¹,G.Venkateswara Rao²

¹Research Scholar, Department of IT GITAM Deemed to be University Visakhapatnam, India.

²Professor, Department of CSE GITAM Deemed to be University Visakhapatnam, India.

Katakam916@gmail.com¹, dr.vgurrala@gmail.com²

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract: In present scenario, tracking of target in videos with low resolution is most important task. The problem aroused due to lack of discriminatory data that have low visual visibility of the moving objects. However, earlier detection methods often extract explanations around fascinating points of space or exclude mathematical features in moving regions, resulting in limited capabilities to detect better video functions. To overcome the above problem, in this paper a novel method which recognizes a person from low resolution videos is proposed. A Three step process is implemented in which during the first step, the video data acquired from a low-resolution video i.e. from three different datasets. The acquired video is divided into frames and converted into gray scale from RGB. Secondly, background subtraction is performed using LBP and thereafter Histogram of Optical Flow (HOF) descriptors is extracted from optical flow images for motion estimation. In the third step, the eigen features are extracted and optimized using particle swarm optimization (PSO) model to eliminate redundant information and obtain optimized features from the video which is being processed. Finally to find a person from low resolution videos, the features are classified by Support Vector Machine (SVM) and parameters are evaluated. Experimental results are performed on VIRAT, Soccer and KTH datasets and demonstrated that the proposed detection approach is superior to the previous method.

Keywords: Low Resolution videos, LHOF, LBP, PSO, SVM, VIRAT dataset, Soccer, KTH Dataset.

1. Introduction

Over the past few decades, many efforts are made in the area of motion detection and compliance to make the following reliable, robust and effective: video surveillance, robots, authentication system, media production, biological research, etc [1]. But there are many challenges that create barriers to improving these applications. These challenges can include lighting change, dynamic background, concealment, space closure, shadow etc as in [2]. For object tracking in low resolution videos these barriers become more severe. In a low-resolution video it is very difficult to find exactly what you like because a lot of discriminatory information such as visual and original material is lost. It leads to the wrong pursuit that continues to lead to the discovery of a malfunctioning event. However, there are some advantages to using low-resolution video as it requires less storage, time of transfer and time interval [3]. Most common tracking methods are based on high resolution video (HR) to extract a straight line as in [4] and stand as in [5] target books. But these methods require additional calculation costs because they work in high-resolution frames. Other methods in the literature use videos with lower resolution as input but over time these videos are developed to higher resolution with the help of higher resolution techniques, proving that they are less expensive. In unconventional event detection documents, many methods such as [6], [7] use classification categories to identify events and do not use low-resolution video input. These dividers require study time and careful attention to the training database. Other methods such as [8] require manual setup at the start of the event's default program and have high computer costs. From the above literature it is observed that a new effective algorithm to be developed for detection of objects in low resolution videos which help in fully automated monitoring system.

2. Related Work

Unsaved field monitoring systems are required to have the ability to process accurate image and real-time to detect and track the moving objects. For tracking a system the accurate acquisition process of moving objects is one of the necessary requirements. While monitoring performance is guaranteed, it is challenging to meet real-time requirements in highly customized squares with a large viewing platform. In addition, in active field conditions, complex background, lighting changes, local movements such as tree shaking, dust tracking, concealment and more make the system problematic.

With the acquisition of moving motions, existing technological approaches in particular include light flow [9], [10], background output [11], frame differences [12] and in-depth learning methods [13]. In many low resolution video surveillance systems these modes can only detect moving objects without having to find a category and direct links to each object which is in moving state. In [14],removal of features and classifications techniques was included to address the above problem. However, the critical reversal of this approach cannot be overlooked is that

the regions of the object obtained by the moving acquisition cannot be classified as a single object or group of objects. The effect of separation of foreground and background images is disturbed due to non availability of direct access to the moving objects.

Typically, each video frame is split front and back to get moving objects depending on the type of pixel or color distribution. In the literature [15], [16], researchers proposed a variety of artificial extraction methods. As a widely used method, to judge the movement the frames are considered, the difference between the frames are based on gray level difference the video frames which are adjacent to each other. Apart from modeling the background, it is easy to use, as the local movements in the backgrounds are prone to noise or complex scenes. Meanwhile, the Gaussian composite (GMM) model [17] is very powerful, but requires a lot of modeling and rear modeling frames respectively, which has high computer cohesion and difficulty handling video frames with varying brightness, something that moves frequently and hides.

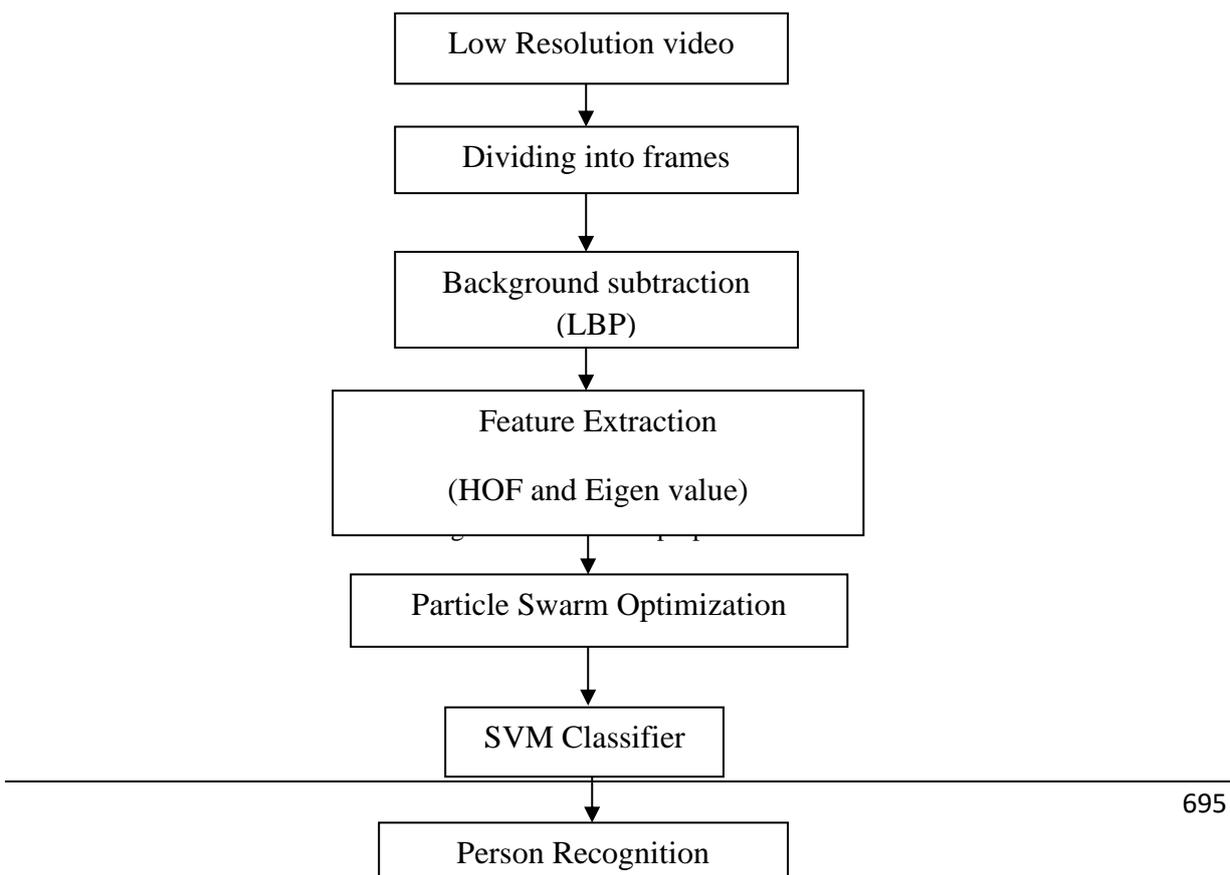
After the acquisition process, regional links can be obtained by connecting regional algorithms. By taking a photo input several times, scan face techniques [18] paste the label on each pixel and separate the target according to the label. For faster detection, methods designed to prevent [19] were introduced. However, these algorithms still use a lot of computer resources and are unable to integrate a broken object into sound.

In [23] Ranganarayana et al., presented two efficient algorithms for motion detection in low resolution videos captured from surveillance data so as to provide privacy. But these algorithms have their own drawbacks. One being the Standard FCM algorithm doesn't consider any spatial information in the image context due to which the data becomes sensitive to noise and other imaging artifacts. Secondly the mean shift algorithm is only based on colour feature, which exhibits performance limited to partial occlusions.

In [24] the authors made a detailed comparative study on various approaches, datasets and applications with respect to human identification in low resolution videos which reveals certain facts and challenges to be addressed. Methods like spatio-temporal based and Optical Flow analysis were also discussed which play a vital role and could be used in various applications.

3. Proposed Methodology

The proposed framework comprises various activities i.e. dividing the video into frames, background subtraction, Feature extraction, feature processing, finally classification and human recognition. The activities and associated techniques are presented in figure 1.



A.Dataset

One of the most important dataset which is used in improving of vision based community is VIRAT video dataset. The dataset is designed very realistic and challenging for video processing domains, as it provides low resolution videos and recognition of persons in video task will be performed. Virat dataset provide videos sequences of length from 0.5sec to 5min for performing experimental evaluations. In this paper, we have considered 10000 video sequences with a frame rate of 30F/sec. The size of frame is 1280pixels in width and 720pixels in height [22].

The KTH database [23] is the largest and most common data commonly used in the signaling classification of human actions. The database contains activities like walking, running, jogging etc which are performed on several topics and different situations. The dataset consists of 600 videos in every category, six different actions and four situations mentioned. Every sequence as a unique background with vertical camera of 25fps (frames per second). The sequence of video is reduced to 5sec of length. The size of frame is 160pixels in width and 120pixels in height.

The soccer dataset consist of four activities i.e. kicking, running, walking and dribbling. The available videos in this dataset are 255. The sequence of video is reduced to 3sec of length and 28 frames per second. The size of frame is 150pixels in width and 150pixels in height.

B. Back ground Subtraction (BS)

The videos obtained from the datasets are divided into frames and perform background subtraction. The construction of background is one of the important steps while performing background subtraction. The performance of BS purely depends on the features which are selected for background modeling. This back ground subtraction is done using local binary pattern (LBP) operator.

LBP is operated by taking difference of center pixel with the other neighbor pixels. This helps in extracting the visual descriptors. The mathematical evaluation is given as,

$$LBP_{Rr,Sp} = \sum_{p_i=0}^{Sp-1} (N_{p_i} - C_c) \cdot 2^{p_i} \tag{1}$$

Here, Rr is --the radius and Sp is the sampling point. N_{p_i} is the neighbour pixel and C_c is the center pixel. The C_c coordinates is taken as (0,0). The N_{p_i} coordinates is taken as $(x + RrCos(2\pi p_i/Sp), y - RrSin(2\pi p_i/Sp))$

In the process of BS, the first frame of video is used as background model and for every pixel in the frame we calculate LBP using equation 1. The next frame LBP is compared with background model LBP. If the value is similar for any of the background model LBP then the pixel is classified as background else the pixel will be classified as foreground.

C. Feature Extraction

The motion of objects in a video sequence is characterized and unified using optical flow estimation method. It is widely used technique for tracking of systems, detection of objects and specific needs that are required. In this HOF features are extracted from the motion range images which grasp the information that is available in form of dynamic. The algorithm also evaluates the vectors fields which are very much useful in identifying the spatial movements of images over a period of time and provides necessary information for motion analysis [20].

In order to identify the required object in the low resolution videos, the flow constraint evaluation done using following equations:

$$VI_x(a_x b_x) \cdot dx(a_x b_x) + VI_y(a_x b_x) \cdot dy(a_x b_x) = -Int_i(a_x b_x) \tag{2}$$

From the above equation, the spatial derivatives of the image are termed as $VI_x(a_x b_x)$ and $VI_y(a_x b_x)$ for the position of pixels in spatial domain $[a_x b_x]^T$. The change in intensity between the images is given as $Int_i(a_x b_x)$. The pixel values are assigned for the images based on the colour of images. Here we consider grey scale images so the value of pixel is 0 for background objects and is 1 for foreground objects. After assigning the values the images are filtered with a moving average filter so that the binary objects in the borders are transferred into close

linear edges. The displacement vector is considered as $Dis(a_x b_x)$ at a position $[a_x b_x]^T$. The pixels transform from one position to other position and position of pixel in current frame $[a'_x b'_x]^T$ is

$$dis = [a'_x, b'_x]^T - [a_x b_x]^T \tag{3}$$

Equation (1) has no solution as they are two unknowns. Additional constraints can be given by smoothing constraints on the field of optical flow motion model. Minimizing the errors using,

$$E = \sum \|VI_x(a_x b_x). dx(a_x b_x) + VI_y(a_x b_x). dy(a_x b_x) + Int_i(a_x b_x)\|_2 \tag{4}$$

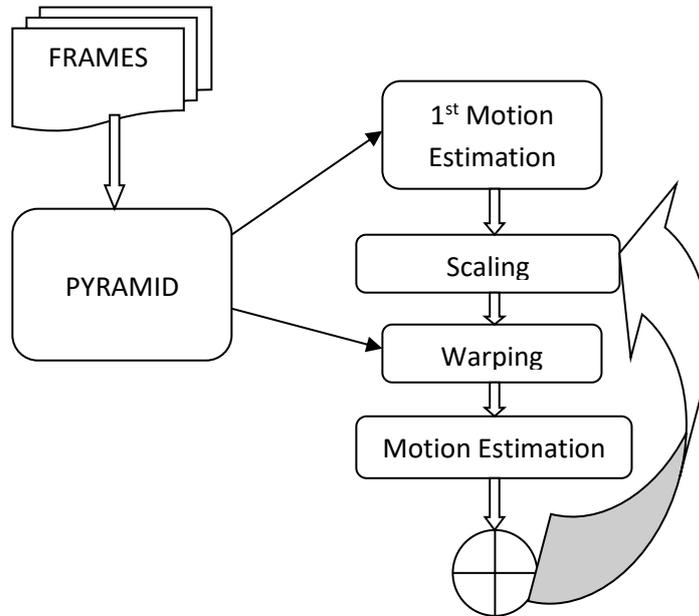


Fig2. Process of Optical Flow motion Estimation

The Eigen values of the features need to be determined. Based on number of blocks per frame and number of frames per volume Eigen values are evaluated. The changes observed in the pixels of a particular direction need to be identified to calculate the Eigen values. The highest change observed in Eigen values is termed as Eigen vector in that direction. Hence these Eigen value and Eigen vectors are calculated in every direction for all the frames in the video. These eigen features helps in identifying the person.

D. Feature Processing

In this preprocessing technique modified particle swarm optimization (MPSO) is considered for optimizing the feature obtained using feature extraction process. The features are selected randomly using modified PSO. These randomly selected features will be given to fitness function calculated using equation 2.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) , \gamma > 0 \tag{5}$$

Where K is a radial basic function (RBF) SVM kernel x_i and x_j are training vectors ($i \neq j , x_i \neq x_j$) and γ is kernel parameter.

$$arg \max_{S \in \varphi} Acc\{K(S_1, S_2)\} \tag{6}$$

Where φ is Set of original features, S is set of Selected features, K is RBF SVM kernel. The positions of the features are updated based on swarm technique. The position are calculated using below equations

$$P_1 = P_Q - B.V_Q \tag{7}$$

$$P_2 = P_R - B.V_R \tag{8}$$

P_1 and P_2 are the position of swarms(features) which are updated w.r.t P_Q and P_R . Here P_Q and P_R are the initial position of the swarms (features).

Where V_Q and V_R are the velocity calculated using below equation

$$V_Q = |C.P_Q(t) - P| \tag{9}$$

$$V_R = |C.P_R(t) - P| \tag{10}$$

Where B and C are the vector coefficients which are calculated as,

$$B = 2b.r_1 - b \tag{11}$$

$$C = 2 \cdot r_2 \tag{12}$$

b is linearly decreased from 2 to 0 over the course of iterations and r1, r2 are random vectors.

To generate a better position, the information of best position P is used in PSO. In this the best position is considered as best features.

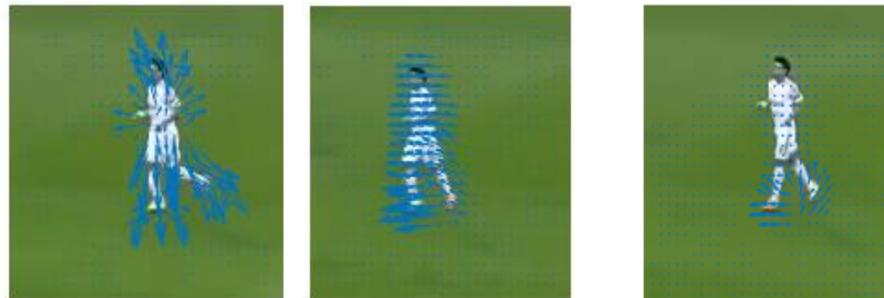
$$P = \frac{P_1 + P_2}{2} \tag{13}$$

E. Classification Technique

The features obtained using optimization technique is divided into training and testing data. The data is retrieved for n number of iterations. The use of machine learning technique in the final stage helps to improve the recognition system. In this paper, proposed support vector machine (SVM) classifier. The SVM classifiers use hyper-plane for performing its operation. The dataset frames are trained with both positives and negatives and are divided using the hyper plane. Later the SVM makes decision by providing test dataset. Parameters are selected based on K-fold cross-validation. Division of dataset considered for training is 70% and 30% for testing using SVM non-linear RBF kernel. Hence, person identification is done. The performance metrics are evaluated from the output retrieved from SVM.

4. Results and Discussion

The evaluation of proposed methodology is done using KTH Dataset, Soccer Dataset and VIRAT Dataset. Three datasets are mostly used in recognition of person in low resolution videos. The dataset are framed for performing experimental results. The dataset consists of different videos of different lengths. These videos are taken from cameras which have lower resolution. The parameters are evaluated used the proposed methodology. The person recognition of different datasets is shown in fig 3, fig 4 and fig 5.



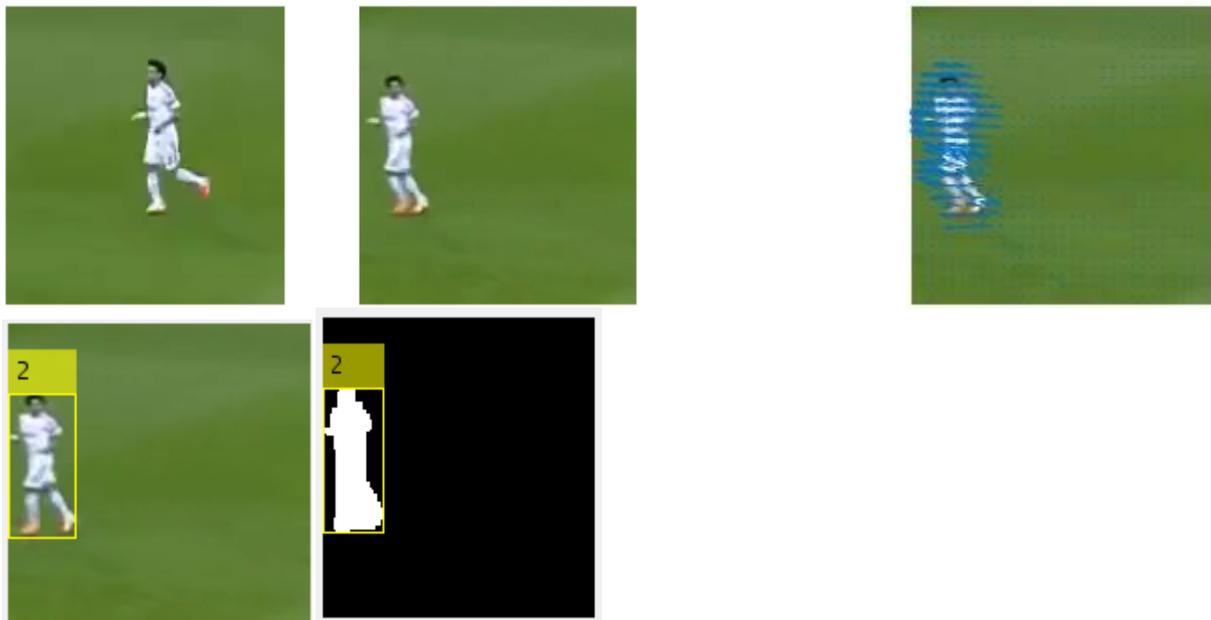


Fig3. Person Recognition for Soccer dataset

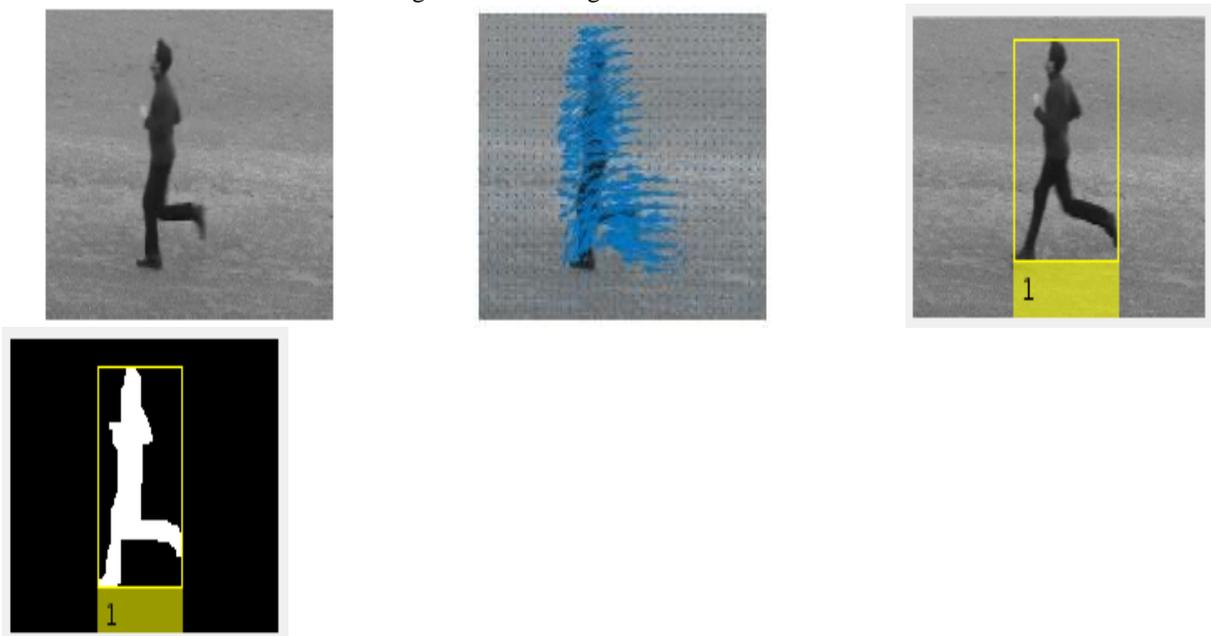


Fig4. Person Recognition from KTH Dataset

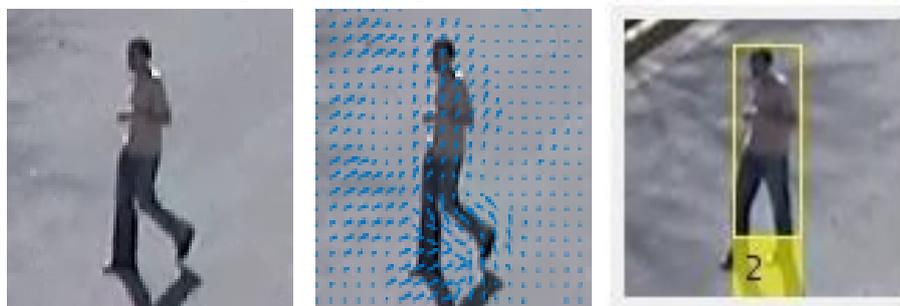


Fig5. Person Recognition from VIRAT Dataset

Evaluation Metrics

The evaluation metrics are used to differentiate the existing technique and proposed technique. The terms which are used to evaluate the proposed work is shown in table 1. The training and testing of the dataset is performed. First the training of SVM is performed and a model is created, one for each dataset. For the rest of our experiments, we test our person recognition on the reasonable subset of videos from VIRAT, Soccer and KTH test sets.

Table1. Parameters Evaluated

Metrics	Formula
Miss Rate	1-Recall i.e. No of FP in positive samples/Total no of positive samples
FPPI	False positives/Total images
Detection Rate	Number of correctly classified person recognition
Precision	Fraction of true positives /All positives
F1 Measure	$TP/(TP+ (0.5*FP+FN))$
False Detection Rate	1-Precision

As shown in table 1 all the parameters are evaluated using proposed methodology. The results obtained using Particle swarm optimization and without using optimization techniques are compared.

The results obtained using various datasets using optimization technique and without using optimization technique is shown in table 2

Table 2. Evaluation of Datasets without using optimization technique

	KTH Dataset	Soccer Dataset	VIRAT Dataset
Accuracy	85.29	84.15	82.63
Detection Rate	83.78	82.61	81.67
False Detection rate	14.37	15.56	18.16
FPPI	13.28	14.39	16.48
Miss Rate	16.21	17.38	18.32
F1 Score	84.69	83.514	81.754
Precision	85.628	84.431	81.836

The use of optimization technique helps in improving the rate of accuracy and other parametric values for person recognition in low resolution videos. The parameters values obtained are shown in table 3.

Table 3. Evaluation of Datasets without using Swarm optimization technique

	KTH Dataset	Soccer Dataset	VIRAT Dataset
Accuracy	90.512	89.56	88.61
Detection Rate	90.019	88.56	88.48
False Detection rate	9.98	10.37	12.57
FPPI	9.04	9.50	11.27
Miss Rate	9.98	11.43	11.51
F1 Score	90.12	89.087	89.951
Precision	90.23	89.620	89.425

The comparison of the results obtained are shown using bar graphs. Here, the accuracy, decision rate, F1 sco and Precision comparison is shown in fig 6 and False detection rate, FPPI and miss rate comparison is shown in fig 7.

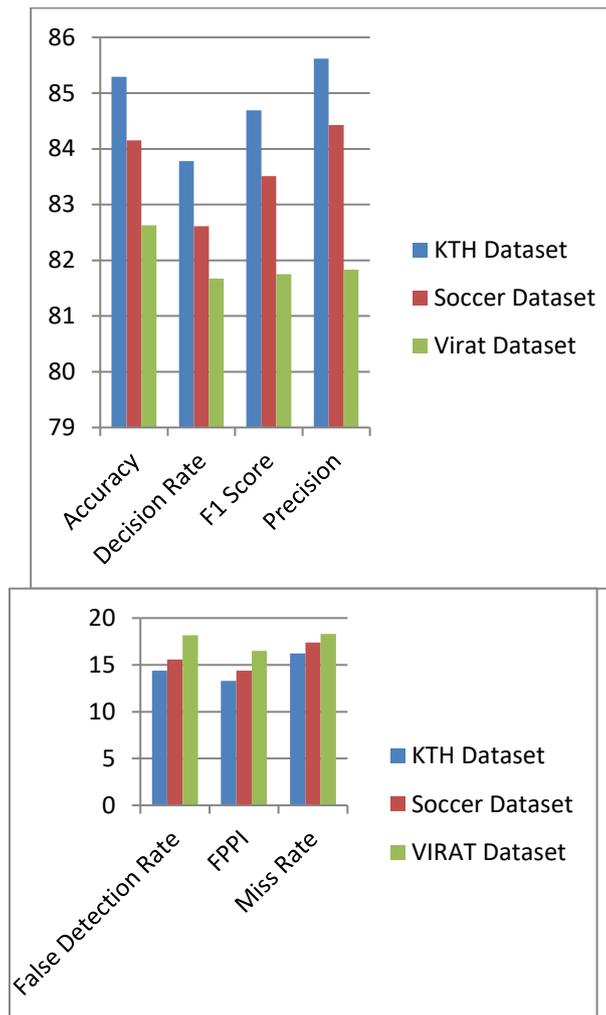


Fig 6. Comparison results obtained using Optical flow algorithm.

The comparison shown in figure 6 are evaluated using optical flow algorithm and classified using support vector machine. Among the three datasets KTH dataset results having an accuracy of 85.29. The other parameters like F1 score and precision are better for KTH dataset..

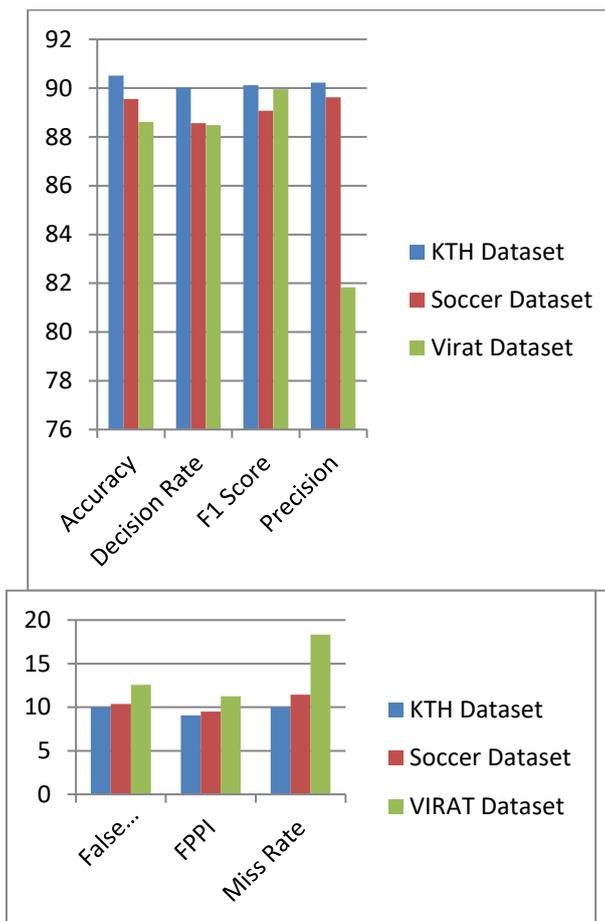


Fig 7. Comparison results obtained using proposed PSO-OFA.

From fig 7 it is shown that proposed PSO-OFA SVM techniques performs well in terms of all the parameters evaluated compared to OFA-SVM. The accuracy obtained for KTH dataset using optimization technique is 90.51% which is 5% higher than without optimization technique.

5. Conclusion

The problem of recognition of person in low resolution videos is studied in this paper as a part of our research work. By using optical flow motion estimation, Eigen values and particle swarm optimization techniques, the underlying problem of person recognition has been improved. Feature extraction technique plays a key role in identifying the person. The optimization techniques help in obtaining the global best features. The improvement is shown using the calculated parameters. The performance of proposed methodology is executed on KTH video dataset, VIRAT dataset and Soccer dataset. By considering different datasets, the accuracy of proposed aspect has been evaluated. Soccer dataset contains action videos and we performed recognition of the person. The proposed PSO-OFA technique outperforms well. Further we can combine more features with the existing ones for improving the performance of human recognition.

References

1. K.Srinivasan, K.Pokumaran, G.Sainarayan, "Improved BackgroundSubtraction Techniques for Security in Video Application", in *Anti-counterfeiting, Security, and Identification inCommunication*, 2009, pp. 114-117.
2. Sugandi, B., Hyoungseop Kim, JooKooi Tan, Ishikawa, "Trackinglow resolution objects by metric preservation", in *Computer Visionand Pattern Recognition (CVPR)*, 2011, pp. 1329-1336.
3. Nan Jiang, Heng Su, Wenyu Liu, Ying Wu, "Tracking lowresolution objects by metric preservation", in *Computer Visionand Pattern Recognition (CVPR)*, 2011, pp. 1329 – 1336.
4. Y. Chen, Y. Rui, and T. Huang."Multicue hmm-ukf for realtimecontour tracking", *IEEE Transactions on Pattern Analysis andMachine Intelligence*, 2006, pp. 1525 –1529.

5. D. Cremers, "Dynamical statistical shape priors for level set based tracking" in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, pp. 1262–1273.
6. Kamijo, S., Ikeuchi, K.; Sakauchi, M., "Traffic monitoring and accident detection at intersections", in *IEEE Transactions on Intelligent Transportation Systems*, 2000, pp. 108-118.
7. Tian Wang, Snoussi, H., "Histograms of Optical Flow Orientation for Visual Abnormal Events Detection", in *IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2012, pp. 13-18.
8. Lili Cui, Kehuang Li, Jiapin Chen, Zhenbo Li, "Abnormal event detection in traffic video surveillance based on local features", in *Image and Signal Processing (CISP)*, 2011, pp. 362-366.
9. Z. Wang, X. Sun, W. Diao, Y. Zhang, M. Yan, and L. Lan, "Ground moving target indication based on optical flow in single-channel SAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1051-1055, Jul. 2019.
10. Y. Xin, J. Hou, L. Dong, and L. Ding, "A self-adaptive optical flow method for the moving object detection in the video sequences," *Optik*, vol. 125, no. 19, pp. 5690-5694, Oct. 2014.
11. [11] H. Sajid and S.-C.-S. Cheung, "Universal multimode background subtraction," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3249-3260, Jul. 2017.
12. G. Shi, J. Suo, C. Liu, K. Wan, and X. Lv, "Moving target detection algorithm in image sequences based on edge detection and frame difference," in *Proc. IEEE 3rd Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Oct. 2017, pp. 740-744.
13. X. Ou, P. Yan, Y. Zhang, B. Tu, G. Zhang, J. Wu, and W. Li, "Moving object detection method via ResNet-18 with encoder-decoder structure in complex scenes," *IEEE Access*, vol. 7, pp. 108152-108160, 2019.
14. C.-W. Liang and C.-F. Juang, "Moving object classification using a combination of static appearance features and spatial and temporal entropy values of optical flows," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3453-3464, Dec. 2015.
15. T. S. Haines and T. Xiang, "Background subtraction with Dirichlet process mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 670-683, Apr. 2014.
16. Z. Bian and X. Dong, "Moving object detection based on improved Gaussian mixture model," in *Proc. 5th Int. Congr. Image Signal Process.*, Oct. 2012, pp. 109-112.
17. J. Zuo, Z. Jia, J. Yang, and N. Kasabov, "Moving target detection based on improved Gaussian mixture background subtraction in video images," *IEEE Access*, vol. 7, pp. 152612-152623, 2019.
18. L. He, X. Zhao, Y. Chao, and K. Suzuki, "Configuration-transition-based connected-component labeling," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 943-951, Feb. 2014.
19. J. Chen, K. Nonaka, H. Sankoh, R. Watanabe, H. Sabirin, and S. Naito, "Efficient parallel connected component labeling with a coarse-to-fine strategy," *IEEE Access*, vol. 6, pp. 55731-55740, 2018.
20. D.X. Zhou and H. Zhang, "Modified GMM background modeling and optical flow for detection of moving objects," *Conference Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, 2005, pp. 2224-2229.
21. N. Dalal, *et al.* "Histograms of oriented gradients for human detection." In *CVPR*, pp. 886-893, 2005.
22. K. Ranganarayana and G. Venkateswara Rao, "Motion detection in low resolution video surveillance data to provide personal privacy", in *IJAER* Vol.14 No. 23 (2019) pp. 4251-4255.
23. C. Schuldt, I. Laptev and B. Caputo, "Recognizing Human Actions: A Local SVM Approach" *ICPR'04*.
24. K.RANGANARAYANA and G. VENKATESWARA RAO, 2020. A STUDY ON APPROACHES FOR IDENTIFYING HUMANS IN LOW RESOLUTION VIDEOS. *International Journal of Advanced Research in Engineering and Technology (IJARET)*. Volume:11, Issue:12, Pages:1665-1679.