International Journal of Communication Networks and Information Security

ISSN: 2073-607X, 2076-0930

*Volume **14** Issue **1s** Year **2022** Page **91:99***

A Grey Wolf Intelligence based Recognition of Human-Action in Low Resolution Videos with Minimal Processing Time

Ranga Narayana[1*] and G. Venkateswara Rao[2]

[1]Research Scholar, Department of IT GITAM Deemed to be University, Visakhapatnam, India. corresponding.katakam916@gmail.com*

[2]Professor, Department of CSE, GITAM Deemed to be University, Visakhapatnam, India. E-mail: dr.vrgurrala@gmail.com

Article History

Received: 13 July 2022
Revised: 20 September 2022
Accepted: 26 October 2022

Abstract

The usage of video cameras for security purposes has grown in recent years. The time for recognition of human plays an important role in solving many real time problems. In this paper, the process for identifying human action is done by separating the background using local binary pattern (LBP) and features extracted using faster histogram of gradients (FHOG) and Eigen values based on power method. The features are combined and optimized using grey wolf optimization (GWO) and finally classified using support vector machine (SVM). The experimental results are compared with existing methods in identifying the human action. The time factor is evaluated and compared with different optimization techniques like particle swarm optimization (PSO), Firefly algorithm (FA) and grey wolf optimization. The entire process is performed on three well known datasets like VIRAT dataset, KTH dataset and Soccer dataset. The comparison results prove that the recognition is done in quick time i.e. 10.28sec with improved rate of accuracy 93.35% for soccer dataset using proposed method.

Keywords: SVM classification, Grey Wolf Optimization, Eigen Values, Human action, faster histogram gradients.

1. Introduction

Human recognition and human action recognition is quite possibly the main exploration area in PC vision because of its usefulness in genuine applications like video observation, human PC collaboration, and video documented frameworks. In any case, action recognition actually stays a troublesome issue when managing unconstrained recordings, for example, web recordings, film and TV shows, and observation recordings. There are several difficulties in identifying the humans. In which some of the main reasons are appearance, position of viewing, and obstructions, relative same scenes of videos, fluctuation in brightness, shadowing and cam movements. While substantial progress has been made in resolving these issues, the problem of poor quality in has received far less focus during the research phase [1]. It is extremely difficult to identify human actions from videos of lower quality because critical visual data is degraded by a number of internal [2] and external factors such as reduced resolution, rate of sampling, compression of artifacts and blurriness in motion, cam instability, and tremor. Sample video frames of different datasets are shown in Figure 1 which has very low visual quality. Numerous surveillance frameworks necessitate additional video inquiry on minimally saved video data [3], while cell phones endeavor to fuse most significant semantic level into ongoing streaming [4, 5]. Hence, for reasons some of the examples of this action identification in inferior quality recordings ought to be additionally explored as it offers new experiences and difficulties to the exploration local area.

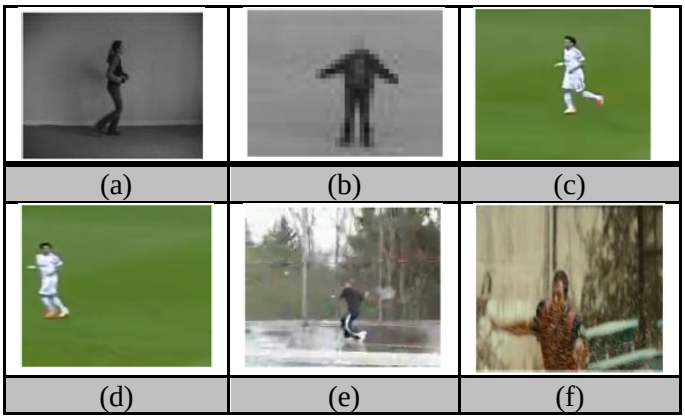


Figure 1. Example of Sample data collected low-quality datasets like KTH (a,b), Soccer(c,d) and VIRAT (e,f)

Shape and movement are the most well-known components for human activity identification in videos with lower quality [6, 7]. Mirjalili et al., [8], further improved it by presenting two fundamental kinds of neighborhood highlights: shape as histogram of gradients (HOG), and movement as histogram of optical flow (HOF). Laptev [9] stretched out the HOG shape descriptor to three measurements i.e. histogram of 3D inclination directions (HOG3D), which quantizes 3D angle directions on ordinary polyhedrons. Narayana [10] suggested utilizing part-based portrayals, for example, interest focuses t defeat foundation and impediment related issues. They utilized Dollár's element locator [11] is used to extract rectangular shapes from footage, and each rectangular shape is represented by a widened LBP-TOP (an augmentation of LBP-TOP to nine cuts, three for each plane) descriptor Besides straightforwardly applying LBP on picture outlines, there were elective techniques in writing that pre-owned it to remove surfaces from different types of pictures. Klaser et al., [12] identified the action of humans based on the texture type of descriptors. The utilization of nearby twofold examples to depict movement history and movement energy pictures which develops the data of shape and movement individually.

Wang & Mori [13] introduced two productive calculations for movement discovery in low resolution recordings caught from surveillance information in order to give security. Their research focuses on identifying the action of humans from a wide field of viewing, where the size of frame is often less than forty numbers of pixels. Many optimization techniques like genetic algorithm, particle swarm optimization [14], and firefly optimization were used for feature optimization for identifying of human-action recognition [15].

In this paper, fast HOG and Power method to determine Eigen values and grey wolf optimization (GWO) algorithm is used to reduce the features by which significantly speeds up the run-time of recognition without sacrificing accuracy.

2. Grey Wolf Optimization

The Grey Wolf Optimization (GWO) algorithm mimics grey wolves' strong hierarchy of leadership and mechanism of hunting. Four types of grey wolves are utilized to illustrate the leadership structure: alpha, beta, delta, and omega. This GWO computation mimics the administration order and wolf pursuing process. In hunting process the main steps are categorized into three, one is searching for prey, second is encircling the prey and finally attacking of the prey. These three steps are used for the implementation of the GWO technique. The three position vectors of prey are denoted as $P(\alpha), P(\beta) and P(\delta)$. The velocities are termed in equations (1), (2) & (3). The equations can be shown as

$$V_\alpha = |C.P_\alpha(t) - P(t)| \quad (1)$$
$$V_\beta = |C.P_\beta(t) - P(t)| \quad (2)$$
$$V_\delta = |C.P_\delta(t) - P(t)| \quad (3)$$

$V_\alpha, V_\beta and V_\delta$ are the updated velocity of wolf w.r.t to best position and initial position P(t) of the grey wolf. The positions are updated as eqs. (4), (5) (6):

$$P_1 = P_\alpha - B.V_\alpha \quad (4)$$
$$P_2 = P_\beta - B.V_\beta \quad (5)$$
$$P_3 = P_\delta - B.V_\delta \quad (6)$$

$P_1, P_2 and P_3$ are the updated position of wolfs.

The vector coefficients are termed to be B and C and are calculated as Eq. 7 and Eq. 8

$$B = 2b.r_1 - b \quad (7)$$

$$C = 2.r_2 \quad (8)$$

where b elements drop linearly from 2 to 0 during the period of iterations and r_1,r_2 are the random form of vectors in [0,1]. The positions of each wolf's are illustrated in Figure 2. The final best position of wolf is given as $P(t+1) = P_1 + P_2 + P_3/3$

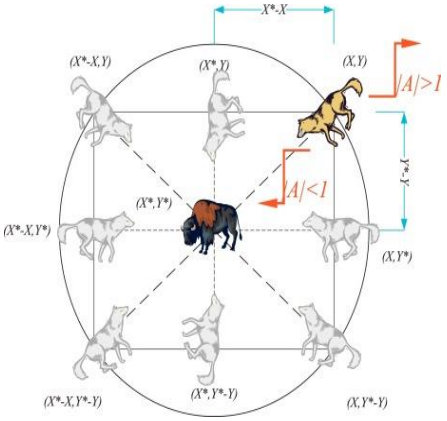


Figure 2.Position Vectors and Their Possible Next Locations

3. Methodology

3.1 Input Dataset

VIRAT video dataset is one of the most important datasets utilized in the improvement of the vision-based community. For undertaking experimental assessments, the VIRAT dataset provides video sequences ranging in duration of$0.5\ \text{sec}\ to\ 5min$. In this work, we looked at 10000 video sequences with a frame rate of 30 fps. The size of the frame considered is 1280×720 pixels with width and height.

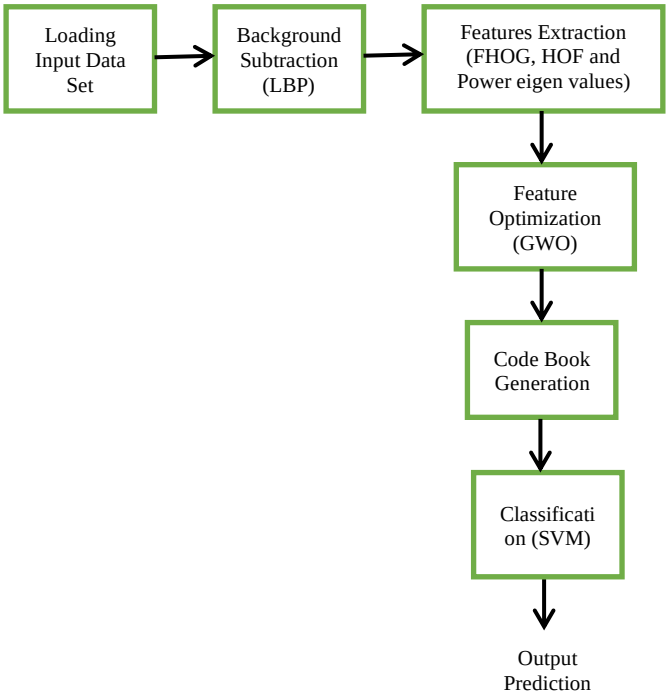


Figure 3.Processed Block Diagram

Second dataset considered is KTH. The videos of human available in this dataset are walking, running, jogging etc. Every category available in dataset has nearly 600 videos. This dataset maintains unique background with a length of 5sec and has 25 fps. The size of every frame considered is 160×120 pixels with width and height. Block diagram of the dataset process is illustrated in Figure 3.The third dataset which is considered is soccer dataset. Some of the action videos are available such as kicking, running, walking and dribbling etc... Around 255 videos are available for conducting experimental evaluation. The length of available videos is 3 sec with 28 numbers of frames in one second. The size of the frame considered is 150×150 pixels with width and height.

3.2 Subtraction of Background

In the paper, input video is divided into frames and is given for Local binary pattern (LBP) for background subtraction. LBP is an important attribute descriptor in computer vision for texture matching. It is an effective method of describing texture. The pixel identification in a scene using LBP operator is done by thresholding the neighbor of each pixel, middle value of pixels and by arraying the obtained results to a binary code [16] LBP has several distinguishing characteristics, including invariance in grey scale, simplicity in computational of non-parametric, invariance in illumination, and high power discrimination. After this LBP process, the features need to be extracted for further processing.

3.3 Extraction of Features

Fast HOG, HOF, and power-based Eigenvalues are used to extract the features. HOG is a feature descriptor that is frequently employed in the extraction of features from picture data. HOG is also responsible for directing the edge. This is performed by identifying the gradient and orientation (or magnitude and direction) of the edges. They are several steps for evaluating HOG descriptors. Firstly the responses of gradient magnitude and flow vector displacement need to be calculated in horizontal and vertical direction for HOG and HOF [17]. In this paper, FHOG features are calculated wherein the gradient histograms are calculated before computing the final features [18]. In FHOG, the down-sampling of image and histogram is applied. At higher levels, the performance of extracted FHOG features is dependent on the picture quality, while at lower levels, it is resistant to extreme down-sampling. The Eigen values of the features must be calculated using their power. The number of blocks each frame and the number of frames per volume are used to calculate Eigen values. By computing (scaled) vectors in the sequence, the "power method" attempts to determine the largest magnitude Eigen value and corresponding Eigen-vector of a matrix [19]:

$$X^{(k)} = AX^{(k-1)} \qquad (9)$$

The superscript (k) refers to the k^{th} vector in a sequence. To calculate the Eigen values, it is necessary to identify the changes detected in the pixels of a given direction. For every frame of the video the Eigen values and Eigen vectors are computed in all the directions. For identification of human these Eigen features are very much useful. The features obtained using FHOG, HOF and power based Eigen values are combined together and given to optimization technique.

3.4 Optimization of Features

The obtained features are optimized using GWOwhich will improve the rate of accuracy in identifying the human and human action. Rey wolf optimization functioning and behaviors of wolfs is discussed in section 2. Based on the process the features will be optimized to global best features. The correct characteristics that are necessary to identify the person and human activity are assessed based on the optimization method.

3.5 Generation of Code Book

In this step, bag of words (BoW) are considered. This BoW model which uses a histogram of visual words to represent the distribution of local motion patterns in a video. In recent years the bag-of-words method is mostly used and has good attention towards the proposed application. These methods begin by detecting local salient areas with spatiotemporal interest point detectors. Around each 3D interest point, features such as gradient and optical flow are retrieved. K-means clustering or Fisher vector can be used to compute the visual words. Consideration of k mean clustering for getting bag of words is suggested in this article. The bag-of-words methods have been demonstrated to be insensitive to changes in look and posture. Finally, the representation of human action is done by histogram of visual words and SVM is used as a classifier to identify the action.

3.6 Support Vector Machine

In the last stage, the employment of machine learning technique aids in the improvement of identification of human action. A support vector machine (SVM) classifier is suggested in this paper [20, 21]. The hyper-plane is used by SVM classifiers to execute their operations. The dataset frames are split using the hyperplane and trained with both positives and negatives. Following that, the SVM makes a choice by supplying a test dataset. The parameters are chosen using K-fold cross-validation. Utilization of data for training is 70% and for testing the utilization of data is 30%. As a result, human-action identification is performed [22].

4. Results and Discussions

The implemented methodology is executed in MATLAB tool. The experimental evaluation is conducted by using an Intel corei5 processor with 8GB RAM and windows OS. To identify the efficiency in functioning of proposed methodology the trails have been carried outon three different datasets one is KTH, other is VIRAT and Soccer. Identification of human action over several datasets is shown by using the proposed methodology in below figures.

4.1 Case 1: KTH Dataset

For conducting experimental findings, we explore several forms of low resolution videos from the KTH dataset. Various action videos are used as input to recognize human actions and offer results. One of the video results is processed and shown in figure 4, figure 5 & figure 6.

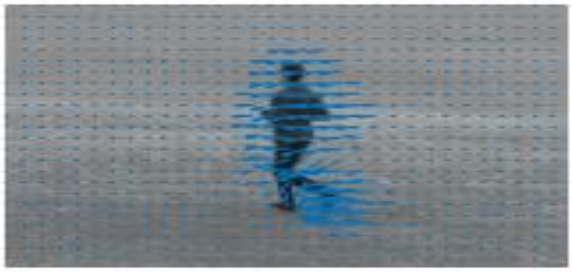


Figure 4. Extraction of Frames for Given Input Video

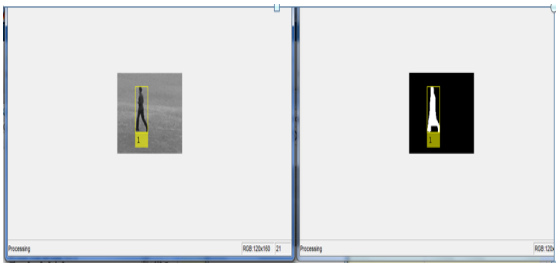


Figure 5. Frames Processing

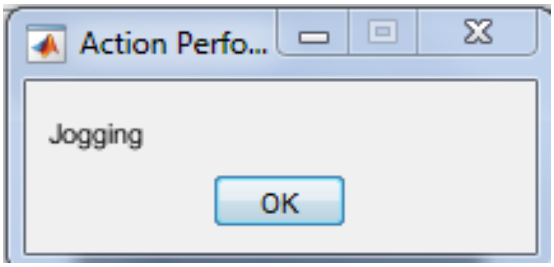


Figure 6. Output Prediction Result

4.2 Case 2: Soccer Dataset

In this section, we investigate several forms of low resolution videos from the soccer dataset in order to execute experimental findings. Various action videos are used as input to recognize human actions and offer results. Figure 7, Figure 8 and figure 9 illustrates the Extraction of frames for given input video, Frames processing and Output Prediction result respectively.

Figure 7. Extraction of Frames for Given Input Video

Figure 8. Frames Processing

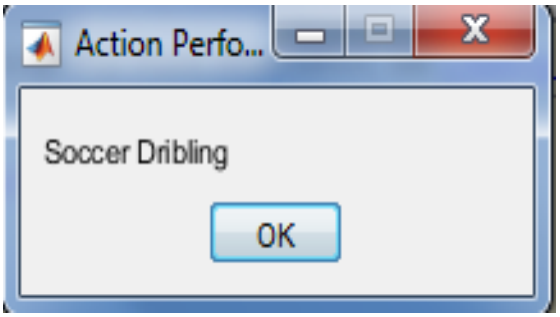


Figure 9. Output Prediction Result

4.3 Case 3: VIRAT Dataset

In this section, we investigate several forms of videos with lower resolution from the VIRAT dataset in order to perform experimental outcomes. Various action films are used as input to recognize human actions and produce results as shown in Figure10& Figure 11.

Figure 10. Extraction Of Frames For Given Input Video

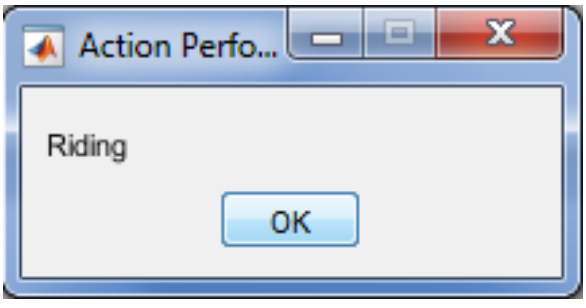


Figure 11. Output Prediction Result

The chosen dataset consists of video sequences and processed with our proposed methodology and compared with existing methods. The parameters like accuracy, detection rate, false detection rate, miss rate, F1 score and precision are identified. The experimental results are evaluated and tabulated. Table 1 shows the Parametric evaluation using different techniques.

The rate of accuracy using the proposed methodology is compared with various other existing techniques on different low resolution datasets and is shown in Table 2. The implemented work produces good accuracy results compared to other techniques. The achievement of good accuracy is due to involvement of the optimization technique. The

accuracy rate in identifying the human action in lower resolution videos obtained using grey wolf optimization is 93.35%. The author details and their value of accuracy in identifying human from low resolution videos are shown.

Table 1. Performance Evaluation Results

		KTH Dataset	***Soccer Dataset***	***VIRAT Dataset***
Detection Rate	PSO-OFA	90.02	88.56	88.48
	FO-SVM	90.50	91.84	89.70
	Proposed Method	**90.82**	**92.88**	**90.98**
False Detection rate	PSO-OFA	9.98	10.37	12.57
	FO-SVM	7.58	7.78	9.58
	Proposed Method	**5.98**	**6.72**	**9.38**
FPPI	PSO-OFA	9.04	9.50	11.27
	FO-SVM	9.01	8.15	10.29
	Proposed Method	**8.77**	**7.12**	**9.01**
Miss Rate	PSO-OFA	9.98	11.43	11.51
	FO-SVM	7.05	7.07	8.74
	Proposed Method	**5.63**	**6.19**	**8.46**
F1 Score	PSO-OFA	90.12	89.08	89.95
	FO-SVM	91.14	92.03	90.05
	Proposed Method	**92.08**	**93.09**	**90.80**
Precision	PSO-OFA	90.23	89.62	89.425
	FO-SVM	92.41	92.21	90.41
	Proposed Method	**94.01**	**93.25**	**90.61**

Table 2. Comparison of accuracy results

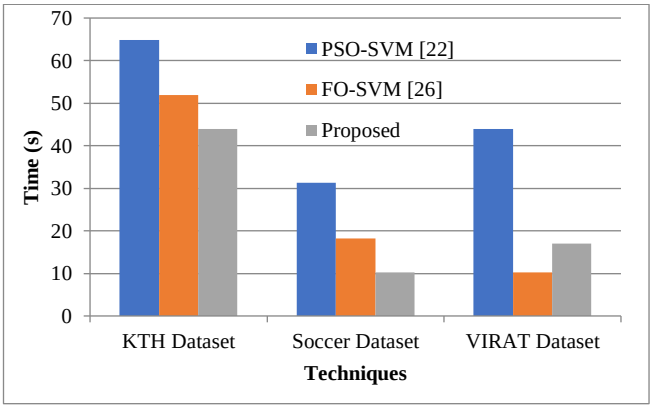
Accuracy (%)	**PSO-SVM**	**89.56**
	Supervised CNN-SVM [1]	91.70
	FO-SVM	92.40
	[3]	92.8
	[6]	92.6
	[22]	90.2
	[21]	91.30
	[20]	91.1
	Proposed GWO-SVM	**93.35**

One of the important constraints considered in this paper is time factor. The time for identification of human and action is observed. By using the proposed methodology the time taken for recognition is very low with higher rate of accuracy.

According to Table 3, the processing time for identifying human activity for the proposed approach is around 43.88 seconds for the KTH dataset. Using proposed method there is improvement seen for every dataset. The graphical representation of comparison of time factor is shown in Figure12.

Table 3. Comparison Of Accuracy Results

Time in seconds		KTH Dataset	Soccer Dataset	VIRAT Dataset
	PSO-SVM	64.890787	31.270154	53.059503
	FO-SVM	51.889344s	18.269328s	33.058520s
	Proposed Method	**43.887720s**	**10.268551s**	**17.057105s**



5. Conclusions

In this paper, the faster evaluation for extraction of features is done by using FHOG and Power based Eigen values. This stage helps to improve the processing time in recognition of human and human action in our experiments. The grey wolf optimization technique helps in obtaining the improved optimized features. Finally linear SVM classification is performed. The evaluation results prove that the proposed method have good rate of accuracy and high-speed processing in identification of human and human action in lower video resolution when compared with other existing techniques. For all the three datasets the proposed system gives good results in which the accuracy rate achieved for KTH dataset is 92.31%, VIRAT dataset is 91.27% and for soccer dataset is 93.35%. The average time considered is sec 10.28 for identification of human action using VIRAT dataset.

References

[1] N. AlDahoul, A. Q. Md Sabri, &A. M. Mansoor, "Real-time human detection for aerial captured video sequences via deep models", Computational intelligence and neuroscience, Vol. 2018, No. 1639561, pp. 1-14, 2018.

[2] E. Cantu-Paz, "Feature subset selection, class separability, and genetic algorithms", In Genetic and evolutionary computation conference, Springer, Berlin, Heidelberg, pp. 959-970, 2004.

[3] A. A.Chaaraoui, P. Climent-Pérez, & F. Flórez-Revuelta, "Silhouette-based human action recognition using sequences of key poses. Pattern Recognition Letters, Vol. 34, No. 15, pp. 1799-1807, 2013.

[4] C. C.Chen, &J.K. Aggarwal, "Recognizing human action from a far field of view", In 2009 Workshop on Motion and Video Computing (WMVC),Snowbird, UT, USA, pp. 1-7, 2009.

[5] B.Saghafi, & D. Rajan, "Human action recognition using pose-based discriminant embedding", Signal Processing: Image Communication, Vol. 27, No. 1, pp.96-111, 2012.

[6] N. Dalal, & B. Triggs,"Histograms of oriented gradients for human detection", IEEE computer society conference on computer vision and pattern recognition (CVPR'05), San Diego, CA, USA, pp. 886-893, 2005.

[7] N.Dalal, B. Triggs, &C. Schmid,"Human detection using oriented histograms of flow and appearance", In European conference on computer vision, Springer, Berlin, Heidelberg, pp. 428-441, 2006.

[8] S. Mirjalili, S. Saremi, S.M.Mirjalili, &L.D.S. Coelho, "Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization", Expert Systems with Applications, Vol. 47, pp. 106-119, 2016.

[9] I. Laptev, "On space-time interest points. International journal of computer vision", Vol. 64, No. 2, pp. 107-123, 2005.

[10] K.R. Narayana, "Humanrecognition Using 'PSO-OFA' In Low Resolution Videos", "Turkish Journal of Computer and Mathematics Education (TURCOMAT)", Vol. 12, No. 11, pp. 697-703, 2021

[11] P.Dollár, V. Rabaud, G. Cottrell, & S. Belongie, "Behavior recognition via sparse spatio-temporal features", In 2005 IEEE international workshop on visual surveillance and performance evaluation of tracking and surveillance, Beijing, China, pp. 65-72, 2005.
[12] A. Klaser, M. Marszałek, &C. Schmid, "A spatio-temporal descriptor based on 3d-gradients", In BMVC 2008-19th British Machine Vision Conference, Leeds, pp. 99.1-99.10, 2008.
[13] Wang, Y., & Mori, G. "Human action recognition by semilatent topic models", IEEE transactions on pattern analysis and machine intelligence, Vol. 31, No. 10, pp. 1762-1774, 2009
[14] K. Ranganarayana, &G.V. Rao, "A Study on Approaches For Identifying Humans In Low Resolution Videos", International Journal of Advanced Research in Engineering and Technology (IJARET), Vol. 11, No. 12, pp.1665-1679, 2020.
[15] B.Saghafi, & D. Rajan, "Human action recognition using pose-based discriminant embedding", Signal Processing: Image Communication, Vol. 27, No. 1, pp.96-111, 2012.
[16] S.Rahman, J. See, &C.C. Ho, "Action recognition in low quality videos by jointly using shape, motion and texture features", IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Malaysia, pp. 83-88, 2015.
[17] K. RangaNarayana, &G.V. Rao, "Action recognition in low resolution videos using FO-SVM", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 12, No.4, pp. 1149- 1162, 2021.
[18] K. Ranganarayana, &G.V. Rao, "A Study on Approaches For Identifying Humans In Low Resolution Videos", International Journal of Advanced Research in Engineering and Technology (IJARET), Vol. 11, No. 12, pp.1665-1679, 2020.
[19] K.K. Reddy, N. Cuntoor, A.Perera, & A. Hoogs, "Human action recognition in large-scale datasets using histogram of spatiotemporal gradients",IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance,Beijing, China, pp. 106-111, 2012.
[20] A. Iosifidis, A. Tefas, &I. Pitas, "Discriminant bag of words-based representation for human action recognition", Pattern Recognition Letters, Vol. 49, No. 11, pp. 185-192, 2014.
[21] J.See, &S. Rahman, "On the effects of low video quality in human action recognition", IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA),Gold Coast, QLD, Australia, pp. 1-8,2016.
[22] Wang, Y., & Mori, G. "Human action recognition by semilatent topic models", IEEE transactions on pattern analysis and machine intelligence, Vol. 31, No. 10, pp. 1762-1774, 2009