Journal of Scientific & Industrial Research
Vol. 81, February 2022, pp. 166-172

Multi-class SVM based C3D Framework for Real-Time Anomaly Detection

Vishnu Priya Thotakura and N Purnachand*
School of Electronics Engineering, VIT-AP University, Amaravati 522 237, Andhra Pradesh, India

Received 28 May 2021; revised 03 January 2022; accepted 19 January 2022

The conventional multi-class anomaly detection models are independent of noise elimination and feature segmentation due to large number of feature space and training images. As the number of human anomaly classes is increasing, it is difficult to find the multi-class anomaly due to high computational memory and time. In order to improve the multi-class human anomaly detection process, an advanced multi-class segmentation-based classification model is designed and implemented on the different human anomaly action databases. In the proposed model, a hybrid filtered based C3D framework is used to find the essential key features from the multiple human action data and an ensemble multi-class classification model is implemented in order to predict the new type of actions with high accuracy. Experimental outcomes proved that the proposed multi- class classification C3D model has better human anomaly detection rate than the traditional multi-class segmentation models.

Keywords: Convolutional neural network, Multiple instance learning, Region of interest, Support vector machine

Introduction

Human anomaly detection is one of the most promising areas of research in the recent past. Automatic detection of multi-class human anomalies will facilitate in understanding more complex actions and its variations. Most of the traditional machine learning techniques have the responsibility to identify different non-linear traditional approaches depend upon the basic concepts of a supervised learning scheme. These methods need manually annotated Regions of Interest (ROIs). There exist another group of techniques that includes the concepts of Multiple Instance Learning (MIL). In this work, the authors have emphasized on weakly-supervised classification scheme. Most traditional approaches detect inappropriate patterns on high-dimensional data sets that are computationally infeasible. Therefore, all patterns that are not required during the classification process are difficult to process. As a result, the total computing overhead is also significantly increasing.[1] Therefore, selecting essential features that usually take part during the classification process is very critical. Traditional techniques of image segmentation depend on two standard features; these are-similarity and differentiation. The approach to similarity is the most popular and widely accepted. It uses the index of similarity between numbers of different picture objects. Some of the most common techniques of similarity are provided below: thresholds, region growth, regional decomposition, and regional integration. This technique is more complex than a histogram with a two-model gray level.[2] First of all; there are large springs due to various objects, such as the sky, the branch and the leaf. Secondly, in these very short variations where there is local intensity variations due to a fluttering input when the pixel is projected from the same object.

Related Work

A hierarchical Segmentation model has been implemented using local features.[3,4] The Global pattern matching in the hierarchical partitioning algorithm may result in noisy pattern in the segmentation results. To overcome this problem, a partitioning based hierarchical segmentation was developed to find the anomaly patterns using the global pattern matching.[5] It is reported that the background subtraction is a strategy commonly used for segmentation of movement as part of static scenes. The background image formation is known as background evidence, e.g., after an initial period by means of averaging images.

In addition, morphological post handling activities such as disintegration, extension and shutting are carried out following a delineating of the foreground pixels to reduce noise effects and upgrade the distinguished areas.[6] Different

*Author for Correspondence
E-mail: chanduinece@gmail.com

authors from literature proposed and proved different models.[7–13] PSO based Bayesian technique is proposed in this study.[14] In this paper, the correlation based anomaly detection is performed on the groups. Here, Bayesian model is used to predict the action of the group using the PSO features for the action prediction process. As the number of classes in the action groups is increasing in size, proposed prediction model has less prediction rate due to high dimensionality and memory constraints.

Materials and Methods

Proposed Model

The proposed framework in Fig. 1, describes for the multi-class anomaly detection model implemented in four phases i.e., multi-class feature extraction, multi-class feature ranking, multi-class segmentation and multi-class classification model. In the first phase, a hybrid ensemble feature extraction model is designed and implemented on the input training data. In the second phase, multi-class feature ranking approach is used to find the feature ranking process for the C3D framework. In the third phase, multi-class feature segmentation is used to filter the noise in the edges. This segmentation process is used to filter the over-segmented regions in the edges. Finally, in the fourth phase, a multi-class SVM classification model is implemented on the optimal C3D features for anomaly prediction.

Feature Extraction for C3D Model: Ensemble Feature Extraction Models for Multi-class Anomaly Detection

In the proposed Multi-class feature extraction, kernel probability, Gaussian and non-linear exponential estimators are used to filter the ensemble features from the input data. These filters are used to find the essential key features for the Multi-class feature ranking process.

Filter 1: Kernel Probability-Based Feature Extraction

This work emphasized on weakly – supervised classification scheme and higher order interactions among independent variables. This Kernel probability-based feature extraction approach detects appropriate patterns on high – dimensional datasets which are computationally feasible. In the Filter 1, the Gaussian probability estimation is proposed to find the important key motion features in the given input training data. A hybrid kernel probability estimator is implemented in order to predict the essential motion vectors to the given input data.

HIf = HistIntensities (I),

$$\text{Gaussian Kernal Estimator: } GKE(\varphi, \theta) = \frac{e^{-\theta^2}}{\eta^* \varphi^2} \quad \ldots (1)$$

$\Psi = GKE(\sum HI_f, \sum V_f)$; η: Scaling factor; $\delta(V_f) =$

$$\frac{\partial V_f}{\partial t} = div(H(\|Gr(V_f)\|)(Gr(V_f)) \quad \ldots (2)$$

where $Gr(V_f) = \nabla$: gradient, div: divergence, $\|$: Eucledian norm

$$\text{Kernel Probability Estimator} =$$
$$KPE(D) = |HI_f \,/\, (\sum \max\{\Psi, \delta(V_f)\} * HI_f)| \quad \ldots (3)$$

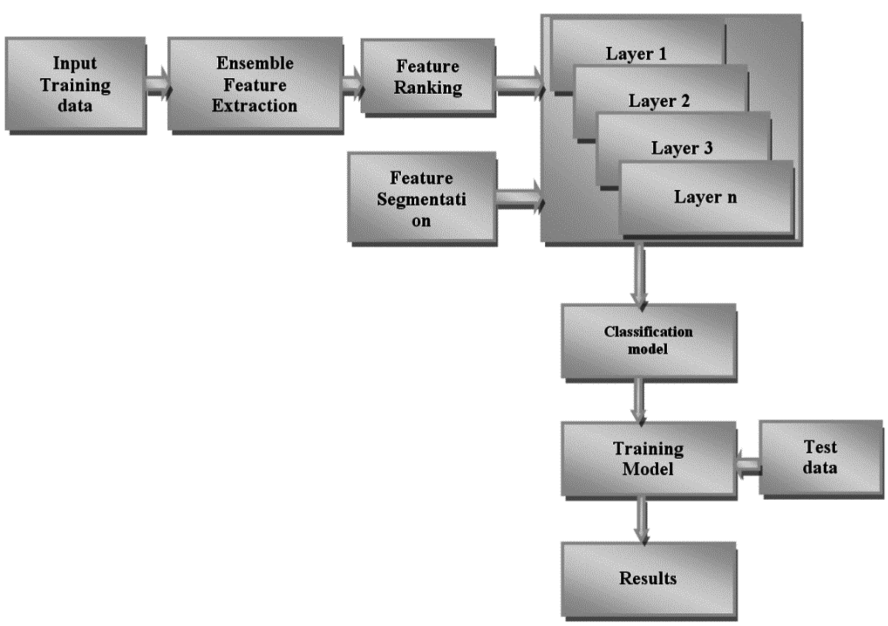

Fig. 1 — Proposed Model

The Gaussian probability estimation in Filter 1 is proposed to find the essential key motion features in the given input training data. A hybrid kernel probability estimator is implemented in order to predict the essential motion vectors to the given input data.

Filter 2: Gaussian Ranking Measure for Outlier Detection

To improve the process of ranking features in high – dimensional space, Gaussian ranking measure is employed. This ranking measure is helped in ranking univariate scoring metrics. Input: Training frames T, Features space FS and Compute the proposed Gaussian probability measure to find the key object in the given training data.

Input: Training frames T, Features space FS and Compute the proposed Gaussian probability measure to find the key object in the given training data.

$$\mathrm{G\_Pr\ (F)} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \left\{ \mathrm{e}^{2*\mu(\mathrm{i})\mathrm{KPE}\left(\mathrm{F}(\mathrm{I}(\mathrm{i})-\mu(\mathrm{i}))\right)} / \sigma(\mathrm{i}) \right\} \quad \ldots (4)$$

where G_Pr (F) defines the Gaussian Probability measure of each image block features F. F (I (i)) defines histogram value of ith frame.

Filter 3: Non-Linear Exponential Gaussian Estimator

In this filter, a non-linear exponential Gaussian estimation is performed on the training image frames. In this filter, histogram is computed to each block in the ith video frame in order to compute the Gaussian motion vector as kernel filter in the C3D framework.

Non-linear Gaussian estimation using histogram H is computed as:

$$\text{Nonlinear Gaussian Estimator} = \mathrm{NGE(F)} = \sum_{i=0}^{N} \frac{1}{\sqrt{2\pi \mathrm{e}^{\mathrm{F(H[i])}}}} \,.\, \mathrm{e}^{-|\mathrm{H[i]} - \mu(\mathrm{H})|} \quad \ldots (5)$$

Multi-class Feature Ranking

In this work, a hybrid feature ranking measure is used to find the essential key features among the high dimensional space. Proposed ranking measure is used to find the maximized probabilistic entropy to each feature in the given feature space. A hybrid entropy measure to compute the feature rank in the given feature space is given as

$$\mathrm{HybridEnt(FS)} = \prod_{i=0}^{|\mathrm{FS}|} \mathrm{P}(\mathrm{F_i}/\mathrm{MC_k}) . \frac{\log(\mathrm{F_i}).\log(\mathrm{Max}[\mathrm{F_i}])}{\sqrt{\mathrm{Ent}(\mathrm{F_i})}} \quad \ldots (6)$$

where $\mathrm{P}\left(\frac{F_i}{MC_k}\right) = -F_i \log\left(P(MC_i)\right).prob(F_i), \mathrm{Ent}(F_i) = -\sum_i P\left(\frac{F_i}{MC_k}\right).Prob\left(\frac{F_i}{MC_k}\right)$ … (7)

Multi-class Thresholding-Based Segmentation Algorithm

In this work, a hybrid and faster approach is developed in order to identify the over-segmented outliers in the edges during the feature ranking and classification process. The proposed scheme includes the basic concepts of a global optimization algorithm which is usually used to maintain a balance between a probabilistic localization map and spatial consistency. Thresholding based segmentation is one of the key concepts in multi-class anomaly detection and masking. Most of these approaches give better results in case of video frames without noise distortion on the edges. On the contrary, these algorithms are incapable to handle noisy images.

$$\min H(V,C) = \sum_{i=1}^{M} \sum_{k=1}^{N} (\mu_{pq})^r \left\| X_q - C_p \right\|^2 \quad \ldots (8)$$

Subject to $\sum_{k=1}^{N} \mu_{pq} = 1{,}0 \leq \mu_{pq} \leq 1$

where, V and C are graph vertices and central regions,

X_q represents each pixel in the region; Cp is the central pixel of each region.

μ is the mean of the region.

In this work, an efficient evaluation criterion depending on the Gini index and entropy evaluation. Apart from this, it also measures the pixel’s homogeneity inside a particular region.

$$k_0 = \{m(a,b)\varepsilon H/\, 0 \leq m(a,b) \leq r_1 - 1\}, k_1 = \{m(a,b)\varepsilon H/\, r_1 \leq m(a,b) \leq r_2 - 1\}$$
$$k_N = \{m(a,b)\varepsilon H/\, r_n \leq m(a,b) \leq G - 1\}$$

H: Hypothetical region, r_1 Segmented region one, r_2 Segmented region two, G Gaussian bounded value

$$VarL(a,b) = \sum_{(i,j)\varepsilon s}[L(i,j) - \overline{L_c}]^2, VarW(a,b) = \sum_{(i,j)\varepsilon s}[W(i,j) - \overline{W_c}]^2 \quad \ldots (9)$$

$$m(a,b) = \sum_{(i,j)\varepsilon s} \frac{\left[[L(a,b) - \overline{L_c}(a,b)] X\, [W(a,b) - \overline{W_c(a,b)}]\right]}{\sqrt{VarL(a,b)\ X\ VarW(a,b)}} \ldots (10)$$

$$Correlation_e = L(a,b) + C_1\,[L(a,b) - \overline{L_c}(a,b)] + C_2\,[W(a,b) - \overline{W_c(a,b)}]^{*} .\mathrm{m\,(a, b)} \quad \ldots (11)$$

where, a, b are level wise pixel values, L (i, j) is level wise pixel regions, W(i, j) is weight of the ith level regions.

In this work, a level set based segmentation method, can segment every individual image using intensity homogeneity. In this model initially, the actual image is transformed into a new modality in which object and background can be distinguished properly. The process of multilevel image thresholding segmentation is responsible for splitting a particular image into different non-overlapping regions. In this work, a new non-local means two-dimensional histogram and an extended version of the gravitational search approach is implemented in order to detect the optimal thresholds.

Ensemble Filter based Multi-class C3D Classification Model

In the low-end systems, conventional methods can be processed in real time, and in limited conditions such as simple context and setting, fixed lighting, and so on, these methods achieve good efficiency. However, in complex environments, the efficiency of these methods can dramatically decrease. Deep approaches based on CNN are able to identify single or multi-class anomalies. In order to locate objects in each image, conventional object detection approaches are typically focused on multiple features. Feature descriptors are used on the basis of low-level visual indications, making it difficult in complex circumstances to capture representative semantic information. Finally, each stage of the detection pipeline is separately built and configured so that it is unable to obtain the optimum global solution for the entire system. One or more convolution layers concatenated with a regular neural network consisting of one or more hidden layers called completely connected (FC) layers can be present in CNN architecture. The new FC layer is the output layer that determines the making of the diction. The convolution layers are normally accompanied by a sub-sampling layer and a nonlinear layer. The role of visual cortex cells in the brain corresponds to the function of the convolution layer in CNN. Convolution layers are used to create feature maps of the input image as a feature extractor. This means that the CNN convolution layer is synonymous with local filters added to the input space and with the coefficient of the filter kernel calculated during the training process. A series of primitive patterns; the low-level features represented in the input images, such as edges and lines; the first convolution layer can be removed. By integrating these primary features, such as corners, the second convolution layer defines patterns of patterns. By integrating these secondary features obtained from the previous layer, the third convolution layer extracts higher-level characteristics based on detecting patterns of certain patterns, and so on. A non-linear "trigger" function used to differentiate signal apart from useful features on each hidden layer in a general neural network as well as CNNs.

In this framework, a hybrid ensemble feature extraction models are used as kernels in order to filter the motion vectors in the training frames. In this framework, 3 × 3 × 3 ensemble kernel estimator is used as filter with 5 pooling layers and 2 connected fully layers. The segmentation model is integrated in the C3D framework in order to remove the noise in the feature identification process. The C3D network has 8 converting layers, five max pooling layers and two fully connected layers, followed by the softmax output layer. Finally, an ensemble multi-class SVM is proposed in the C3D framework for anomaly prediction on the features.

Non-Linear Multi-class CNN-SVM Classifier Algorithm

In the hybrid non-linear multi-class SVM approach, an efficient non-linear kernel function is used to transform the feature space and to predict the anomaly class in each video frame.

Input: C3D positive and negative multi-class feature bags and its class labels

Step 1: Initializing the non-linear SVM hyper parameters and kernel function.

Step 2: To each feature set in the feature extraction process, apply a non-linear SVM function and its kernel function on the input video frame.

Step 3: Construct a non-linear optimization function for the SVM model. In this proposed model, the optimization function is minimized to improve the error rate of the prediction. The boosted CNN-SVM classifier and its non-linear kernel function are applied on the feature sets for class prediction.

For each feature set do

Apply SVM multi-class optimization models as

$$min_{W_k,a_k} \frac{1}{2}\|W_k\|_1^2 + \tau_m + sig(i).\textstyle\sum_{i=1}^{l} a_i(y_i[ker\langle x,y\rangle.w + b]1 + \xi_i) - \sum_{i=1}^{l}\gamma_i\xi_i \quad \ldots (12)$$

$$s.t\ sig(i).ker\langle x,y\rangle.w + b \geq 1 - \xi_i^n - \tau_m$$

where $\xi_i^n > 0, \tau_m > 0;\ m = 1 \ldots\ldots classes$

Here kernel function $\ker\langle x,y\rangle$ represents the kernel functions defined from C3D features space to classes.

$$\ker\langle x,y\rangle = \begin{cases} e^{-\,sig(i).\xi_i^n \log(\sum\|x-y\|^2\ if\ x=y} \\ e^{-\,sig(i).\xi_i^n \log(\sum\|x-y\|^{\frac{1}{2}}\ if\ x<y} \\ e^{-\,sig(i).\xi_i^n \log(\sum\|y\|^2\ if\ x>y} \end{cases} \quad \ldots (13)$$

$$sig(i) = \frac{m_{inp}^{(mx)}}{1+e(-\lambda(i-\phi))}\ where\ \phi = \frac{1}{2}\left(m_{inp}^{(mx)} - m_{inp}^{(mn)}\right) \quad \ldots (14)$$

$m_{inp}^{(mx)}: mean\ value\ of\ maximum\ likelihood\ function\ in\ the\ input\ image\ inp.$

$m_{inp}^{(mn)}: mean\ value\ of\ minimum\ likelihood\ function\ in\ the\ input\ image\ inp.$

In this algorithm each multi-class is predicted using the polynomial similarity index as

$$C(px, py) = v_0px^3 + v_1py^3 + v_2px^2py + v_3px^3 + v_4py^3 + v_5$$

where, px and py represent the horizontal and vertical coordinates of image pixels and v0 to v5 are coefficients to be updated for polynomials by using the Eigen values of the multi-class feature vectors. The $\boldsymbol{\lambda}$ is scaling factor.

f(px, py) represents the gray value at the coordinate (px, py) are calculated. Set $\boldsymbol{\mu}_{Ci}$ is the membership degree of ith pixel to the C^{th} class, where, i [1, N], and membership degree needs to satisfy the constraint.

$$E = \sum_{i=1}^{N}\sum_{c=1}^{C}\frac{\boldsymbol{Prob((px_i,py_c)_iB_c).e}^{-\left(\frac{D(px_i,py_c)}{\min(px_i,py_c)}\right)}}{\sum_{c=1}^{k} e^{-\left(\frac{D(px_i,py_c)}{\max(px_i,py_c)}\right)}}$$
$$\|v_0px^3 + v_1py^3 + v_2px^2py + v_3px^3 + v_4py^3 + v_5 - p_i\|^2 \quad \ldots (15)$$

Step 4: Test data is predicted to the class y based on the largest decision values as $argmax\{W_k^TE_i + b_k\}$ … (16)

Results and Discussion

Experimental results are simulated in python and third-party libraries on the different databases. In this work, different multi-class datasets have been referred.[15–16]

Different classes of anomaly are taken to predict the multi-class test data using the proposed filtered based C3D framework. In the framework, all the multi-class features are filtered and extracted in order to predict the class label with high computational accuracy and less error rate. In this section, the runtime computation of the proposed models, feature extraction measures, classification accuracy, recall, precision and error rates are computed and compared to the conventional models.

All the input video frames are trained using the multi-class deep classification process for anomaly prediction. As represented in the Fig. 2, the proposed approach has better feature extraction runtime than the conventional approaches.

The accuracy of the proposed approach from Table 1 has better than the conventional approaches for multi-class action prediction. In Fig. 3(a), the

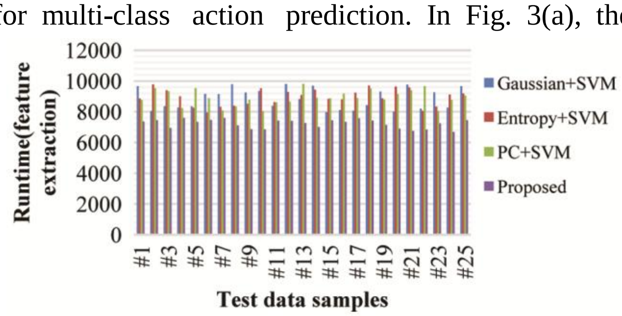

Fig. 2 — Comparative analysis of Multi-class Anomaly prediction model to the conventional models using feature extraction runtime (ms)

Table 1 — Performance result of average probabilistic Multi-class Anomaly detection to the conventional models in terms of accuracy on two training datasets

Test Data	PSO + Bayesian Net	Gaussian +SVM	Entropy + SVM	PC + SVM	Proposed Model
Image1	0.92	0.93	0.91	0.93	0.98
Image2	0.92	0.93	0.94	0.93	0.97
Image3	0.92	0.93	0.94	0.94	0.98
Image4	0.92	0.93	0.94	0.93	0.97
Image5	0.93	0.94	0.94	0.93	0.97
Image6	0.92	0.93	0.94	0.95	0.97
Image7	0.92	0.94	0.93	0.92	0.98
Image8	0.93	0.93	0.91	0.92	0.97
Image9	0.93	0.93	0.93	0.92	0.97
Image10	0.92	0.93	0.91	0.94	0.97
Image11	0.93	0.93	0.95	0.94	0.98
Image12	0.92	0.93	0.92	0.95	0.98
Image13	0.92	0.94	0.91	0.93	0.97
Image14	0.92	0.93	0.92	0.93	0.97
Image15	0.92	0.93	0.93	0.92	0.97
Image16	0.93	0.93	0.94	0.94	0.97
Image17	0.92	0.94	0.91	0.94	0.97
Image18	0.92	0.93	0.93	0.93	0.98
Image19	0.92	0.93	0.93	0.94	0.97
Image20	0.93	0.93	0.95	0.94	0.97
Image21	0.92	0.93	0.94	0.93	0.97
Image22	0.92	0.93	0.93	0.94	0.97
Image23	0.93	0.94	0.95	0.94	0.98
Image24	0.92	0.93	0.94	0.93	0.97
Image25	0.93	0.94	0.92	0.93	0.97

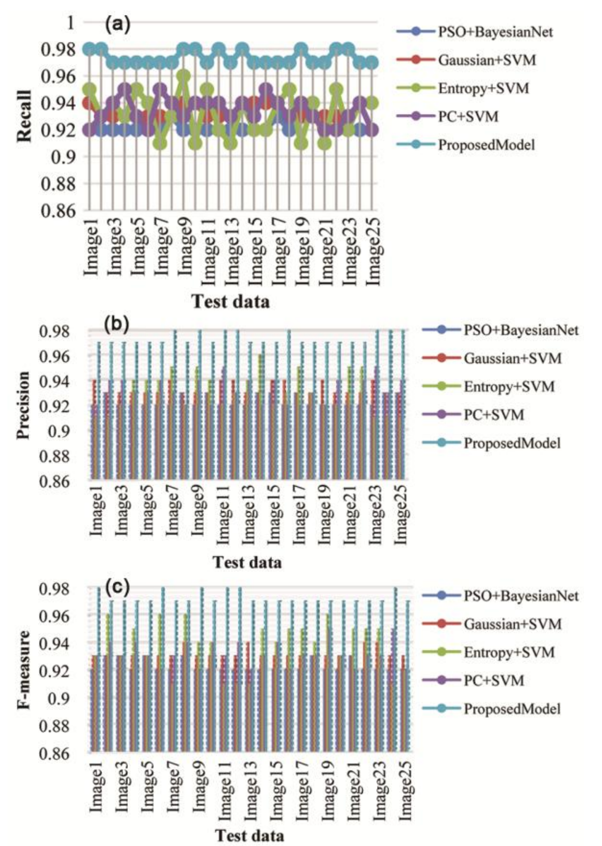

Fig. 3 — Performance of the model in terms of (a) Recall on two training datasets (b) Precision on two training datasets (c) F-measure on two training datasets

Table 2 — Comparative result of proposed multi-class anomaly classification approach to the conventional approaches for Error rate measure

Test Data	PSO + Bayesian Net	Gaussian +SVM	Entropy +SVM	PC +SVM	Proposed
Image1	0.08	0.07	0.08	0.07	0.03
Image2	0.07	0.06	0.05	0.06	0.03
Image3	0.07	0.06	0.08	0.06	0.03
Image4	0.08	0.07	0.05	0.07	0.03
Image5	0.08	0.07	0.08	0.06	0.03
Image6	0.08	0.07	0.09	0.06	0.03
Image7	0.08	0.07	0.08	0.07	0.02
Image8	0.08	0.06	0.05	0.06	0.03
Image9	0.08	0.07	0.06	0.07	0.03
Image10	0.08	0.07	0.08	0.07	0.03
Image11	0.08	0.07	0.08	0.07	0.02
Image12	0.08	0.07	0.07	0.06	0.03
Image13	0.07	0.07	0.07	0.07	0.02
Image14	0.07	0.07	0.05	0.06	0.03
Image15	0.08	0.07	0.05	0.07	0.03
Image16	0.08	0.06	0.09	0.08	0.02
Image17	0.08	0.07	0.08	0.07	0.03
Image18	0.08	0.07	0.08	0.07	0.03
Image19	0.07	0.06	0.08	0.05	0.02
Image20	0.08	0.07	0.08	0.08	0.03
Image21	0.07	0.06	0.05	0.07	0.02
Image22	0.08	0.07	0.07	0.07	0.03
Image23	0.08	0.07	0.06	0.07	0.03
Image24	0.08	0.07	0.06	0.07	0.03
Image25	0.07	0.07	0.05	0.07	0.03

average recall of two datasets is computed for the proposed approach to the conventional approaches and the recall of the proposed approach has better than the conventional approaches for multi-class anomaly prediction. Fig. 3(b) presented the average efficiency of the multi-class anomaly prediction model to the conventional models using classification precision and the average precision of two datasets is computed for the proposed approach to the conventional approaches. In Fig. 3(c), the average F-measure of two datasets is computed for the proposed approach to the conventional approaches, the F-measure of the proposed approach has better than the conventional approaches for multi-class anomaly prediction. In Table 2 the efficiency of the multi-class anomaly prediction model to the conventional models for classification error rate are presented. These models are trained and tested on C3D framework for statistical analysis.

Conclusions

Video anomaly detection plays a vital role in the real-time surveillance systems. Most of the traditional anomaly detection models are based on fixed features size and computational memory. Also, these models use limited segmented features for multi-class anomaly detection process. As the number of human anomaly classes is increasing, it is difficult to find the multi-class anomaly due to high computational memory and time. In order to improve the multi-class human anomaly detection process, an advanced multi-class segmentation-based classification model is designed and implemented on the different human anomaly action databases. In the proposed model, a hybrid filtered based C3D framework is used to find the essential key features from the multiple human action data. In this work, a hybrid multi-class-based anomaly detection model is proposed in order to improve the detection rate in the real-time video datasets. Experimental results show that the present multi-class anomaly detection model has better detection rate and runtime compared to the conventional models in terms of accuracy, precision and recall. In future work, this can be extended to parallel processing on big video databases using the cloud computing environment.

References

1 Xiao Y, Yang X & Liu B, A new self-paced method for multiple instances boosting learning, *Inf Sci*, **515** (2020) 80–90.

2 Zhang X, Yang S, Zhang J & Zhang W, Video anomaly detection and localization using motion-field shape description and homogeneity testing, *Pattern Recognit*, **105** (2020) 107394.

3 Lv P, Yu, Y, Fan Y, Tang X & Tong X, Layer-constrained variational auto encoding kernel density estimation model for anomaly detection, *Knowl Based Syst*, **196** (2020) 105753.

4 Xu D, Song R, Wu X, Li N, Feng W & Qian H, Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts, *Neurocomputing*, **143** (2014) 144–152.

5 Behera S, Dogra D P, Bandyopadhyay M K & Roy P P, Estimation of linear motion in dense crowd videos using Langevin model, *Expert Syst Appl*, **150** (2020) 113333.

6 Qasim T & Bhatti N A, low dimensional descriptor for detection of anomalies in crowd videos, *Math Comp Simul*, **166** (2019) 245–252.

7 Javan Roshtkhari M & Levine M D, An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions, *Comput Vis Image Underst*, **163** (2013) 1436–1452.

8 Xie W *et al*., Unsupervised spectral mapping and feature selection for hyper spectral anomaly detection, *Neural Netw*, **132** (2020) 144–154.

9 Cong T & Baranowski J, Binary Classifier for Fault Detection Based on Gaussian Model and PCA, *IFAC-Papers On-Line*, **51(24)** (2018) 1317–1323.

10 Li S, Yang Y & Liu C, Anomaly detection based on two global grid motion templates, *Signal Process Image Commun*, **60** (2018) 6–12.

11 Jiang G *et al*., Fast inter-frame prediction in multi-view video coding based on perceptual distortion threshold model, *Signal Process Image Commun*, **70** (2019) 199–209.

12 Sabokrou M, Fayyaz M, Fathy M, Moayed Z & Klette R, Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes, *Comput Vis Image Underst*, **172** (2018) 88–97.

13 Aslan F, Durdu A, Sabanci K & Mutluer M A, CNN and HOG based comparison study for complete occlusion handling in human tracking, *Measurement*, **158** (2020) 107704.

14 Song W, Dong W & Kang L, Group anomaly detection based on Bayesian framework with genetic algorithm, *Info Sci*, **9** (2020) 138–149.

15 http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html

16 http://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html