International Journal of Computer Engineering and Information Technology

VOL. 9, NO. 2, February 2017, 30–36
Available online at: www.ijceit.org
E-ISSN 2412-8856 (Online)

A Novel Real-time Human Activity Based Anomaly Detection Model Using Graph Based Clustering and Classification Model

D.Ratna Kishore[1], Dr. M. Chandra Mohan[2] and Dr.Akepogu. Ananda Rao[3]

[1] Asst.Prof, Department of Computer Science & Engineering, Andhrapradesh, India.

[2] Professor, Department of Computer Science & Engineering, JNTUH, Hyderabad, Telangana State, India

[3] Professor in CSE, Director,Academic & Planning, JNTUA,anantapuramu, Andhrapradesh, India.

ABSTRACT

Detecting online abnormality in the video surveillance is a challenging issue due to streaming, video noise, outliers and resolution. Traditional trajectory based anomaly detection model which analyzes the video training patterns for anomaly detection. This paper aims to solve the problem of video noise and anomaly detection .In this paper, a novel filtered based ensemble clustering and classification model was implemented using the threshold based method, graph based clustering algorithm and classification model. Experimental results proved that the proposed model has high computation detection rate compared to traditional real-time anomaly detection models.

Keywords: *Anomaly Detection, Video Anomaly, Graph Based Clustering Model.*

1. INTRODUCTION

In the current trends of digitalization of technology Human activity recognition (HAR) has become most demanding technology. With the exponential growth of electronic devices ,human activity-based processing system requires to perform various activities efficiently. Human activity recognition technology involves video or sequence of images to automatically identify the human activity which is supposed to be recognized by digital electronic devices. Over the last decades human activity recognition and their processing has become the most important area of research and same has been implemented in many applications of digital devices using pattern recognition and computer vision like illegal car parking, analysis of athletes performances, surveillance, security, diagnostics of orthopedic patients etc. Human activity recognition based on videos signifies to and algorithm that a digital device recognizes the human activity is being or was performed by analyzing sequence of images (video frames).An action or activity can be recognized as a set of features. Patterns of visible movement of subjects, edges, surfaces in a visualized scene is due to relative motions occur between a viewer/observer (a camera or an eye) and scene this can be called as an optic flow. In case of low resolutions where limbs could not be identified and the flow of fields are discriminative within a range of action. In case of higher spatial resolutions, configuration of physical body can be identified and recovered. It is identified that 3D configuration can be extracted from 2D images which signifies appearance for body configuration. Many methods of extraction have shown to be successful in analyzing the activity such as: spatio-temporal interest points, characterize spatio-temporal volumes and silhouette histogram of oriented features. These features are synthesized in a descriptor.

A typical HAR system can be classified in two categories: First one is the sequence-based classification, where the geometrical displacement of the feature points is calculated among current frame and initial frame. The second one is the frame-based classification where only current frame is utilized. It can be utilized individually or along with the image frames of the human activity in the past or incoming videos. Frame-based methods do not featured with the quality of geometrical displacement among multiple frames.

In general HAR system processing consists of three steps: pre-processing, feature extraction, and recognition. To execute pre-processing module, few well-known techniques such as histogram equalization (HE), homo-morphic filter, and median filter have been employed in order to improve the quality and accuracy of video frames. While on other hand, a lot of efficient effort has been made for feature extraction module in literature. However each of them possessing some limitations.

Related to the feature extraction, few very well-established methods like space-time volume (STV) has been evolved. However, in SVT approaches, a traditional sliding window technique has been used which consumes a large amount of computation for the localization of actions very accurately, and also it has difficulty in efficient recognition of the actions which can not be spatially segmented. Similarly, local binary pattern (LBP) method is being is being exploited for feature extraction. However Local binary pattern methodology is very sensitive to noise, occlusions, noise and viewpoint which could cause misclassification. It uses 3 * 3 operator for pixel comparison. The typical features can not be extracted by this small operator and also it the directional information of the frame as it only captures the relation surrounding with eight neighbour pixels. To resolve the LBP limitation local ternary pattern (LTP) has been implemented which combines the LBP description property with methods based on patch matching adaptability and appearance invariance. Disadvantage of LTP is, it is non invariant in grey-scale transformation of intensity which is based on a predefined fixed threshold value.

In order to recognition, few well-recognized classifiers like Support vector machine (SVMs), artificial neural networks (ANNs), Gaussian mixture models (GMMs) and hidden Markov model (HMMs) has been utilized for recognition purpose. Out of them HMM is used widely in sequence-based analyze in FER systems as HMM is widely incorporated for sequential data processing while using frame-level features, where as the vector-based classifiers like SVMs, ANNs and GMMs failed in order to processing of sequential feature vectors.

2. RELATED WORKS

H. Su, J. Zou and W. Wang tried to analyze activities in outdoor and identified numbers of issues such as poor resolution, shadows, distances and segmentation issues [1]. In order to overcome these issues, they presented a new algorithm by considering silhouette width and LBP. The authors transformed the activity vectors to gray value. Both static and dynamic features of silhouette width are presented by gray values. Activity width sequence image is represented as texture and merged with LBP (Local Binary Patterns). LBP can be used as pattern extraction tool in spatio-temporal environment. The researchers validated their theory by experiments on outdoor database. Further research can be done on this methodology in order to extend this approach by implementing envelope of activities on the said database for optimized results.

K. G. Chathuramali and R. Rodrigo proposed a new SVM classification for human activity recognition to achieve faster activity recognition [2]. They used SVM over metric learning because of better performance and less cost. This approach performs best when implemented in high dimensional vector space. Here the computation time is directly proportional to numbers of training data (Decrease in training data results reduced computation time). They simulated and identified some features in their presented approach, those are:- 1) Imbalanced training data results very poor performance rate. 2) Decrease in training data results reduced computation time. 3) On increasing training set, accuracy rate increases. They concluded that, their method outperforms other existing approaches.

W. Lin, M. Sun, R. Poovandran and Z. Zhang proposed a new technique for human activity recognition for video processing automatically [3]. They sub-categorized features having high correlations into Category Feature Vectors (CFVs). Numbers of GMMs (Gaussian Mixture Models) are merged together to form activity and GMMs are represented by CFVs. The authors demonstrated that their presented technique is more flexible than that of other pre-existing methods. To enhance the recognition accuracy rate, a new algorithm known as CFR is formed. Here in the process of human activity recognition, video frames with high confident frames are considered. The researchers validated their theory with experiments and showed that, their approach is more effective than that of other approaches.

B. Chakraborty, A. D. Bagdanov, J Gonz`alez and X. Roca tried to merge probabilistic optimization with human activity recognition [4]. They used HMM (Hidden Markov Model) for their approach. The authors suggested that their technique is eligible to differentiate similar actions. The researchers used HMMs to monitor stochastic movements of body parts. The above said algorithm used detectors to find the view point changes and self-occlusions by using sub-classifiers. Every sub-classifier is represented by a SVM. Numbers of detections are merged by simple geometric constraint model. The researchers experimented their method on KTH, Weizmann and HumanEva datasets. They achieved robust human activity recognition through their work.

J. Hernández, R. Cabido, A. S. Montemayor and J.J. Pantrigo introduced a new method to identify human activities in terms of 2D sequences [5]. They integrated real-time visual tracking and feature extraction concepts to present their approach. The researchers divided their concept into three important modules, those are:- tracking, feature extraction and activity recognition. Tracking modules merges particle filter and local seach procedure in order to fasten the computation. Feature extraction module split silhouette into smaller rectangular boxes. Statistics of these rectangular boxes are monitored in the process of evolution. In the last module, it transfers these statistics to SVM for classification. The researchers demonstrated that their models work in real-time environment with better performance.

K. K. Htike, O. O. Khalifa, H. A. Ramli and M. A. Abushariah presented a new approach for human activity recognition for video processing by using sequences of postures [6]. The authors used one static camera for their approach. They used K-means, fuzzy C means, multilayer perceptron self organizing maps and FFNNs classifiers in both training and testing phases. Then the accuracy rate is calculated. Maximum accuracy is achieved in case of self-organizing maps. The researchers also mentioned that supervised learning classifiers outperform unsupervised classifiers. Recognition rate is directly proportional to the numbers of training sets. They compared their proposed model with other models and achieved better results.

H. Kuehne, J. Gall and T. Serre mentioned a new end-to-end generative method for human activity recognition [7]. They integrated the idea of Fisher vector-based visual representation with structured temporal model. Fisher vector adds the advantage of front-end generative models (like Gaussian). They validated their technique for recognition and parsing into action units. The authors also analyzed various feature representations and their effect on the overall performance of the system. They used a HTK framework for evaluation of their technique. The implementation shows that, they are able to achieve significant gains in accuracy for both recognition and parsing. The said models works at its best when sufficient training set data are available.

A. K. Kushwaha and R. Srivastava proposed a new framework for human activities recognition from video sequences [8]. The said framework is invariant in nature and is categorized into three modules, those are:- 1) detection and location by background omission, 2) creation of view invariant spatio-temporal templates, 3) view activities recognition through template matching. In this approach, the background is omitted by change detection and background modelling techniques. Creation of view invariant templates are done. Moment invariants and Mahalanobis distance are considered by the researchers for template matching of different activities. They evaluated on their own dataset, KTH activities recognition datasets, i3Dpost dataset, MSR dataset, video web multiview dataset and WVU human activities recognition datasets. They concluded from their experiments that, their suggested model is efficient, flexible and robust in nature.

S. Maćkowiak, P. Gardziński, Ł. Kamiński and K. Kowalak introduced a new algorithm for human activities recognition on multiview video [9]. The main aim of this approach was to develop a system that will automatically detect human behaviors (such as medical emergencies) and call for emergency help. They picked directed graphical model to implement the behaviors. The said graphical model can be based on propagation nets and dynamic Bayesian networks. They selected voxel reconstruction in order to reconstruct a 3D scene. The researchers demonstrated that, the presented system can achieve human activities recognition. Further future works may be carried out to enhance the interaction between objects.

N. Robertson and I. Reid proposed a new technique for human activities recognition in video processing [10]. They accepted human activities as stochastic sequences and the activities are represented by feature vector. The feature vector is responsible for detecting and monitoring trajectory information. Probabilistic search method is used for activities recognition and HMM is used to smoothen activities sequences. By considering the similarities between predefined HMMs, high-dimensional recognition is gained. The whole phenomenon is carried out by a hierarchy of actions. The higher levels of hierarchy use Bayesian networks whereas, the lower levels use non-parametric sampling technique. On merging both, they developed a new framework.

3. PROPOSED MODEL

Online Video Filtering Algorithm
Input: Video Sequence Images
Output: Preprocessed Video Sequences
Step 1:

For each frame F
do

Divide the human action frame into two classes Cls_1 in X-axis and Cls_2 in Y-axis directions with foreground level fg , background level bg and the block mean value Th such that $fg = \{0,1,2...Th\}$ and

$$bg = \{Th+1, Th+2...Th+N\}$$

where N is the total number of levels in each direction

$$Th = \min\{\sum f_{xy}(x,y))/|f_x(x,y)|, \sum f_{xy}(x,y))/|f_y(x,y)|\}$$

Step 2:

Determine the foreground level lower threshold using C-V model T_1.

The variance of foreground level $fl = \{0,1,2...Th\}$ is

Compute quantization of color component

$f_t^c(t(x)) = \left[l * f_o^c(t(x)) / \mathrm{L} \right] \ \forall \mathrm{x} \in \mathrm{I}$

$step2$: compute $\mathrm{G(t,x)}=[\mathrm{g}_0(t,x), \mathrm{g}_1(t,x)\langle \mathrm{g}_2(t,x)\rangle ... \mathrm{g}_{l^2-1}(t,x)]$

where $\mathrm{g}_j(t,x) = \mathrm{MCcor}_t^X([j/l], j-1*[\mathrm{j}/\mathrm{l}])... \mathrm{j}=0,1,...\mathrm{i}^2-1.$

$\mathrm{q}_i(\mathrm{b,x}) = (1-\alpha)\mathrm{q}_i(\mathrm{b,x}) + \alpha.\mathrm{q}_i(\mathrm{t,x})$ if L(t,X)=0

$\mathrm{q}_i(\mathrm{b,x}) = (1-\alpha)\mathrm{q}_i(\mathrm{b,x}) + \beta.\mathrm{q}_i(\mathrm{t,x})$ if L(t,X)=0

$i = 0,1....k-1$

$$\partial\varphi / \partial \mathrm{t} = \mathrm{g_j}(\mathrm{t,x})(\frac{\mathrm{df}(\mathrm{x,y})}{\mathrm{dx}}[\mu_X \mathrm{div}(\nabla\varphi / |\nabla\varphi|)] + \frac{\mathrm{df}(\mathrm{x,y})}{\mathrm{dy}}[\mu_Y \mathrm{div}(\nabla\varphi / |\nabla\varphi|)])$$

$$T_1 = \partial\varphi / \partial t * \sum (\mathrm{x}* prob^2(x) - x^2 * prob) / |N|$$

Step 3: Determine upper threshold T_2.

$$T_2 = \sum_{x=Th+1}^{N-1} prob(x) * \sum_{x=Th+1}^{N-1} (\mathrm{x}* prob^2(x) - x^2 * prob(\mathrm{x})) / |\mathrm{N}|$$

Step 4:

Check the foreground and background noise using the following condition:

if $(\mathrm{T}_1 > \lambda_1)$

remove gaussian noise(fg);

else if$(\mathrm{T}_2 > \lambda_2)$

remove guassian noise(bg);

else

continue;

Step 5:

Repeat steps 2 to 4 until the video sequence is empty.

After enhancing the given video sequences using a pre-processing algorithm, the next step is to find the human pose in the enhance image using proposed chaining graph model.

Apply Adaptive Local binary patterns

LBP partitioned image into feature segments.

Given an image I, the LBP computes the local patterns of the image texture, which is computed at each pixel value by evaluating the differences between it and its neighbor pixels.

$$\mathrm{LBP}(\mathrm{m}=8, \mathrm{r}=1) = \sum_{\mathrm{i}=0}^{\mathrm{m}-1} \mathrm{f}(\mathrm{M_i} - \lambda).2^{\mathrm{i}}$$

where f(x)=1, $\mathrm{x} \geq 0$

and f(x)=0, $\mathrm{x} \prec 0$

We improved the traditional LBP model with two scheme adaptive LBP model.

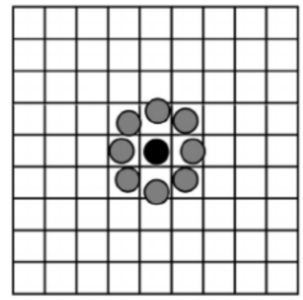
The two scheme LBP model with threshold process is given as:

First threshold LBP scheme as:

$$\mathrm{LBP}(\mathrm{m}=8, \mathrm{r}=1) = \sum_{\mathrm{i}=0}^{\mathrm{m}-1} \mathrm{f}(\mathrm{M_i} - \mathrm{T_1}).2^{\mathrm{i}}$$

Second threshold LBP scheme as

$$\mathrm{LBP}(\mathrm{m}=8, \mathrm{r}=1) = \sum_{\mathrm{i}=0}^{\mathrm{m}-1} \mathrm{f}(\mathrm{M_i} - \mathrm{T_2}).2^{\mathrm{i}}$$



P=8, *R*=1

BLH(i=1,2,…k) are the LBP generated feature histograms.

Once the LBP feature histograms are ready, the next key step is to filter the feature histograms using normalized controlled similarity measure.

For each block BLH (i=1,2,…r) in r blocks

Determine block similarity as

Let $\mathrm{B_i}$ and $\mathrm{B_j}$ are the blocks in the BLH (i=1,2…r); where i!=j

$$\mathrm{NCSim}\left(\mathrm{B_i}, \mathrm{B_j}\right) = \sum \mathrm{CL}(\min(\mathrm{B_i}, \mathrm{B_j}), \mathrm{T_1}, \mathrm{T_2}) / (\mathrm{T_1} + \mathrm{T_2})$$

Where NCSim is the normalized controlled similarity, CL is the control limit of $\mu_{\min(\mathrm{B_i},\mathrm{B_j})} \pm 3\sigma$ limits and $\min(\mathrm{B_i}, \mathrm{B_j})$ is the minimum intensity values of $\mathrm{B_i}$ and $\mathrm{B_j}$.

Enhanced Anomaly Classification Model:

For each dimension (Y,P,R) axis

Construct graph model as

$\mathrm{PG_Y} = \mathrm{ProbChainGraph(Y)};$

$\mathrm{PG_P} = \mathrm{ProbChainGraph(P)};$

$\mathrm{PG_R} = \mathrm{ProbChainGraph(R)};$

Extract $\mathrm{E(PG_Y), E(PG_P), E(PG_R)}$ as Edges sets in Y,P,R direction .

End for

For each dimension (Y,P,R)

Do

$\mathrm{Seg\_PG_Y}[\,] =$ Apply enhanced chaining graph model in Y direction for $\mathrm{PG_Y}$;

$\mathrm{Seg\_PG_P}[\,] =$ Apply enhanced chaining graph model in P direction for $\mathrm{PG_P}$;

$\mathrm{Seg\_PG_R}[\,] =$ Apply enhanced chaining graph model in R direction for $\mathrm{PG_R}$;

Done

Let r_k be the randomly selected region in each direction.

For each region r_i in $SPG_Y[]$

Do

Find center of region in Y direction as

$$c_y = \min\{\frac{|r_k - r_{i+1}|}{|SPG_Y[]|}\} \text{ where } i=\{\{1,2...|SPG_Y[]|\}-k\}$$

Set $\lambda_y = c_y$

Choose k random trees as initial tree growing.

for each tree in the random trees

Randomly select n features from the $Seg\_PG_Y[]$.

For each feature A

Compute conditional probability to each attribute C.

$$P(A): \prod_{i=1,j=1}^{n,m} P(A(v_n, C_m)) \ .$$

Compute the Mutual Information to each attribute MI) : $-prob_i \sum_{i=1}^{m} l\,og \sqrt[n]{prob_i}$

Where

$$prob_i = prob(i)/\text{prob}(D_i)//i=1..m(classes)$$

Class predicted gain measure to each attribute is given as:

$$CPGain_A(D_m) = \max(Corr(A), prob(D_i/D)) * \sum_{i=1}^{m} |D_i|/|D| \times MI(D_i)$$

End for

Create a node with the highest CPGain measure.

End for

Select the majority voting as class prediction from the base classifiers and proposed ensemble learning.

Calculate misclassified error rate and statistical true positive rate;

4. EXPERIMENTAL RESULTS

The evaluation of the proposed system is carried out using standard performance measures which are of falsepositiverate(fpr), truepositiverate(tpr), accuracy and errorrate, true positives(tp), false positives(fp), true negatives(tn), false negatives(fn). (fp+tn) indicates the total number of pixels represented foreground and background.

Table 1. True Positive rate of the existing and proposed model on online video sequences

Online Capture Samples	Avg TPR (EHPM)	Proposed ECGPM	Proposed Model
Home security	0.9463	0.9684	0.9762
Outside Security	0.9486	0.9691	0.9813
Public Security	0.9524	0.9704	0.9892

Table.1. shows the average true positive rate of the enhanced hybrid probabilistic model, ensemble based chaining graph probabilistic model and proposed anomaly detection model. By observing the table it is clear that the proposed technique can perform better true positive rate compare to the existing technique .

Fig.13., shows the average true positive rate of the enhanced hybrid probabilistic model, ensemble based chaining graph probabilistic model and proposed anomaly detection model.

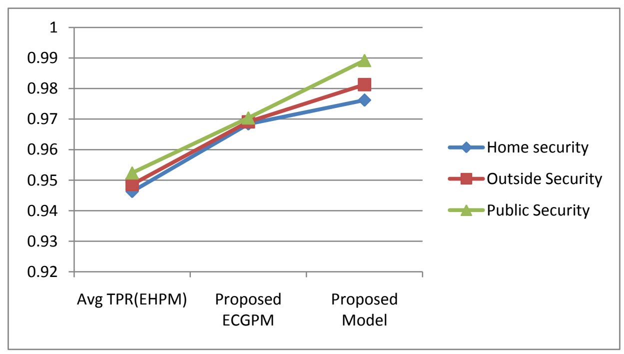


Fig.1. Avg True Positive Rate of the existing and proposed model

By observing the table it is clear that the proposed technique can perform better true positive rate compare to the existing technique .

Table. 2. RunTime Anomaly detection comparison of proposed and existing models on realtime datasets.

Algorithm	RunTime(Secs)
Texture based Pose Estimation	135
JBPCG	175
EJBPCG	116
Proposed Model	98

Table . 2. Describes the runtime comparison of proposed and existing model on the anomaly detection video datasets evaluation. From this table, it is observed that proposed model has less runtime compared to traditional models.

Table 3: Segmentation quality measures with outliers

Algorithm	Avg Segmented Regions	Avg Ouliers
Texture based Pose Estimation	25	14
JBPCG	16	8
EJBPCG	13	8
Proposed Model	8	7

Table 3. Describes the segmentation measures comparison of proposed and existing model on the anomaly detection video datasets evaluation. From this table, it is observed that proposed model has high segmentation quality index compared to traditional models.

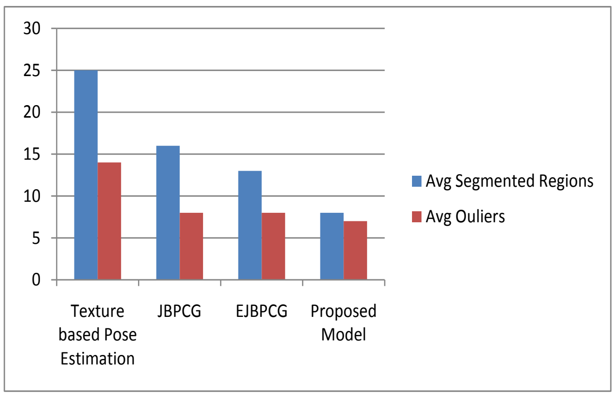


Fig. 2. Segmentation Quality index measures on anomaly data

Table 4: Segmentation results with quality metrics with quality index

Algorithm	Avg Segmentation QualityIndex
Texture based Pose Estimation	0.94
JBPCG	0.97
EJBPCG	0.973
Proposed Model	0.981

Table 4, Describes the segmentation measures comparison of proposed and existing model on the anomaly detection video datasets evaluation. From this table, it is observed that proposed model has high segmentation quality index compared to traditional models.

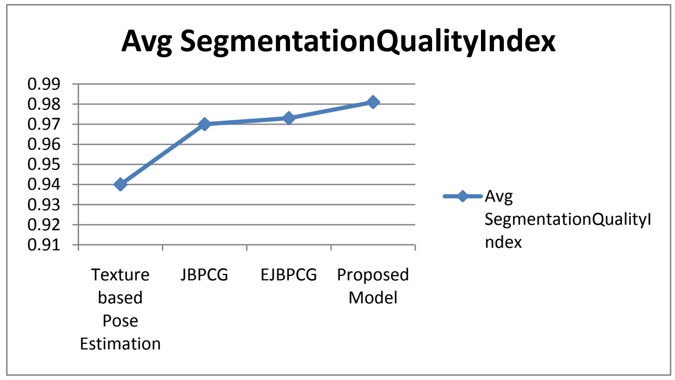


Fig 3,Describes the segmentation measures comparison of proposed and existing model on the anomaly detection video datasets evaluation. From this table, it is observed that proposed model has high segmentation quality index compared to traditional models.

Table 5: Anomaly detection F-measure

Algorithm	Anomaly F-measure
Texture based Pose Estimation	0.87
JBPCG	0.913
EJBPCG	0.934
Proposed Model	0.976

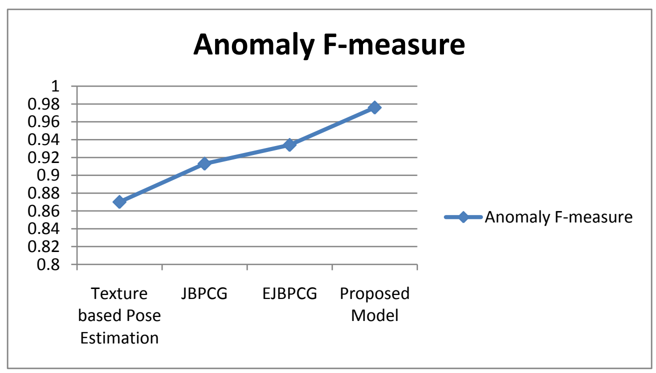


Fig. 4. Anomaly detection F-measure comparison

Fig 4 and Table 5,Describes the anomaly detection measure comparison of proposed and existing model on the anomaly detection video datasets evaluation. From this table, it is observed that proposed model has high anomaly detection rate compared to traditional models.

REFERENCES

[1] H. Su, J. Zou and W. Wang, "Human Activity Recognition Based On Silhouette Analysis Using Local Binary Patterns", "10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)", pp. 924-929, 2013.

[2] K. G. Chathuramali and R. Rodrigo, "Faster Human Activity Recognition with SVM ", "International Conference on Advances in ICT for Emerging Regions – ICTer 2012", pp.197-203, 2012.

[3] W. Lin, M. Sun, R. Poovandran and Z. Zhang, "Human Activity Recognition for Video Surveillance", "", pp.2737-2740, 2013.

[4] B. Chakraborty, A. D. Bagdanov, J Gonz`alez and X. Roca, "Human action recognition using an ensemble of body-part detectors ", "", pp..

[5] J. Hernández, R. Cabido, A. S. Montemayor and J.J. Pantrigo, "Human activity recognition based on kinematic features", "Expert Systems", pp.345-353, 2014.

[6] K. K. Htike, O. O. Khalifa, H. A. Ramli and M. A. Abushariah, "Human Activity Recognition for Video Surveillance using Sequences of Postures ", " ", pp.79-82, 2014.

[7] H. Kuehne, J. Gall and T. Serre, "An end-to-end generative framework for video segmentation and recognition"

[8] A. K. Kushwaha and R. Srivastava, "Multiview human activity recognition system based on spatiotemporal template for video surveillance system ", "Journal of Electronic Imaging", 2016.

[9] S. Maćkowiak, P. Gardziński, Ł. Kamiński and K. Kowalak, "Human Activity Recognition in Multiview Video", "11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)", pp.148-153, 2014.

[10] N. Robertson and I. Reid, "A general method for human activity recognition in video", "Computer Vision and Image Understanding", pp.232-248, 2006.