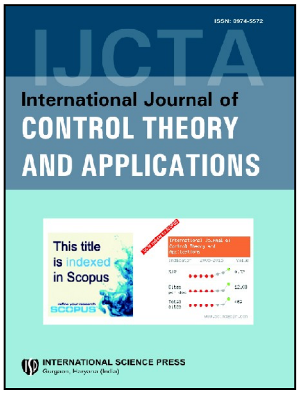

International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 10 • 2017

Content Based Video Feature Extraction and Classification Using Perceived Motion Energy Spectrum-SVM Classifier

N. Brindha[1] and P. Visalakshi[2]

[1] *Assistant Professor (Senior Grade), Department of Applied Mathematics and Computational Sciences, PSG college of Technology, Coimbatore, Tamilnadu, India, E-mail: brindhacsepsgtech@gmail.com*
[2] *Professor, Department of ECE, PSG college of Technology, Coimbatore, Tamilnadu, India, E-mail: spv@ece.psgtech.ac.in*

Abstract: The availability of high end technologies and internet has added to the abundant generation of data online, videos are not an exception for it. This accumulation of huge amount of videos paves way for concern about storage and retrieval. In order to aid faster retrieval the classification of video is essential. The proposed system employs content as the criteria for classification of the video. The term content refers to colour, shape, and texture. In the proposed system, the key frames are generated using block matching algorithm. Later, motion feature is extracted using PMES, this motion feature is estimated by extracting a macro block from a frame and then seeking its best match in another frame. Finally the video is classified using Support Vector Machine. SVM analyze the data and recognise patterns, used for classification purpose. The proposed system is implemented in a real time basis and is compared with the existing systems. The experimental results are also presented for discussion.

Key Words: Classification, Feature Extraction, PMES, SVM

1. INTRODUCTION

Video is the technology of electronically capturing, recording, processing, storing and reconstructing a sequence of images representing scenes in motion. The goal of content- based video classification is to identify and localize specified motion in videos. It relies on the processing of a set of features extracted from a video sequence**.** The video sequence is segmented into several frames. The segmentation step requires detecting all transitions in the video sequence and separating the key frame contained in it using block matching algorithm. The video frames thus segmented can be processed, and various temporal features can be extracted from them. SVM is to make the comparison between the testing and training databases to obtain the motion features. Hence by using PMES and SVM classifier, it is possible to achieve high generalization results for content- based video classification for purposes of retrieval. The concepts and the technologies used in the proposed system are briefly discussed below.

2. OVERVIEW OF CONTENT BASED VIDEO CLASSIFICATION

Content Based video classification is to segment and classify video into meaningful category.

COLOUR HISTOGRAM

In image processing, a colour histogram is a representation of distribution of colours in an image. It represents the number of pixels which have their colour present in a fixed set of colour ranges. The main drawback of colour histogram is that other visual features are not taken under consideration while classifying. The same object may possess different colours in different frames. The colour histogram will not be able to classify both objects under the same category. Another case is that two different objects having same colour would be classified under the same category. This is the major drawback of colour histogram.

EDGE DETECTION

Edge detection is an image processing technique. It figures out the boundaries of objects within images. It works by detecting the discontinuities in brightness. Edge detection is used for image segmentation and data extraction in areas such as computer vision, machine vision. Common edge detection algorithms include Sobel, Canny, Prewitt, Roberts, fuzzy logic methods. It tracks only dots, there is no provision for locating them with a scene.

3. OVERVIEW OF PMES REPRESENTATION

As years have elapsed, the basket for the visual features proposed for content-based video classification has over flown. The pity side of it is that motion information, the most crucial visual feature of human perception of video is not utilized to its core. This back drop is due to the bundle of obstacles that lay in the path of describing motion in videos, since motion is composite mix of camera and object motions. And adding to the hurdles, object motions include both human regarded and disregarded motions. To bring about a break through, Ma and Zhang proposed a shot based motion energy representation termed as PMES. In this method, two types of filters are brought into picture namely, temporal energy filter and a global motion filter. Temporal energy filter is used to eliminate disregarded object motion in scene and global motion filter acts as a shield for object motions from camera motions.

PMES work flow is depicted in fig 1. Since the mixture energy measure expresses the energy resulting from both camera motion and object motion, its ability to describe a video shot is limited. The blocks with low entropy will have consistent vector angle values throughout the shot, suggesting that their motion is the result of camera movement instead of local movement. PMES measure is calculated for every macro block in the shot, and this provides us with a set of descriptor features equal to the number of macro blocks, regardless of the length of the shot and the varying number of the resulting vector fields. PMES describes the magnitude of the local motions in each part of the scene, and in this sense, it seems to provide a unique signature for each video, which could be used for classification based solely on motion.

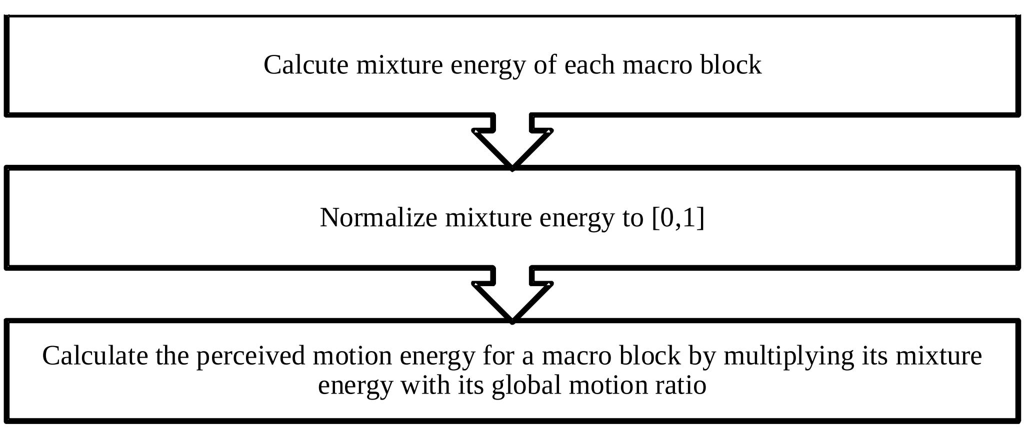

Figure 1: Work Flow of PMES

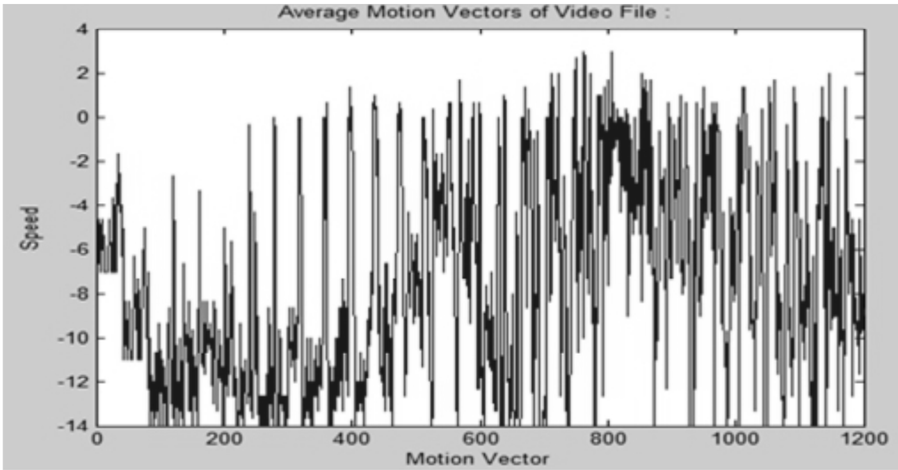

Figure 2: Average Motion Vector

Motion is the essential characteristic distinguishing dynamic videos from still images. Motion information represents the visual content with temporal variation. In PMES, the main objective is to extract the motion feature from the key frame. Motion energy at each macro block position(i,j) can be presented by average of motion. Computation speed is calculated by the average number of points searched for a macro block. The average motion vector obtained from Perceived Motion Energy Spectrum (PMES) is displayed as shown in fig 2.

SUPPORT VECTOR MACHINE (SVM)

SVM are based on the concept of decision planes that define decision boundaries. A decision plane separates between a set of objects having different class memberships. In machine learning, SVM are associated with learning of certain algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of the two categories, an SVM training algorithm builds a model that assigns new examples into one of the categories, making it a non-probabilistic binary linear classifier. A support SVM model is a representation of examples as points in space, mapped so that examples of separate categories are divided by a clear gap. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

4. OVERVIEW OF PROPOSED SYSTEM

The proposed system enhances PMES by employing Support Vector Machine, a very popular and well-established binary classifier. The generalization capabilities of Support Vector Machines make them an excellent classifier. However the back drop is that in the present scenario of Content-Based image and video retrieval its place is often limited to relevance feedback.This system aims to demonstrate that an SVM can be trained with enough videos from a certain class to discriminate between videos that belong to that class (positive) and videos that don’t (negative). This means that, after training, the SVM will be able to correctly classify every new video it is given, based on the parameters estimated from the training set. The concept of temporal segmentation is used to focus on the feature extraction and classification.

Key Frame Extraction

Key frame is extracted using block matching algorithm. This algorithm includes certain morphological processes like morphological reconstruction, dilation, erosion and opening. The block algorithm is depicted in the figure, fig 3. And the application of the algorithm to the sample picture is shown in fig 4.

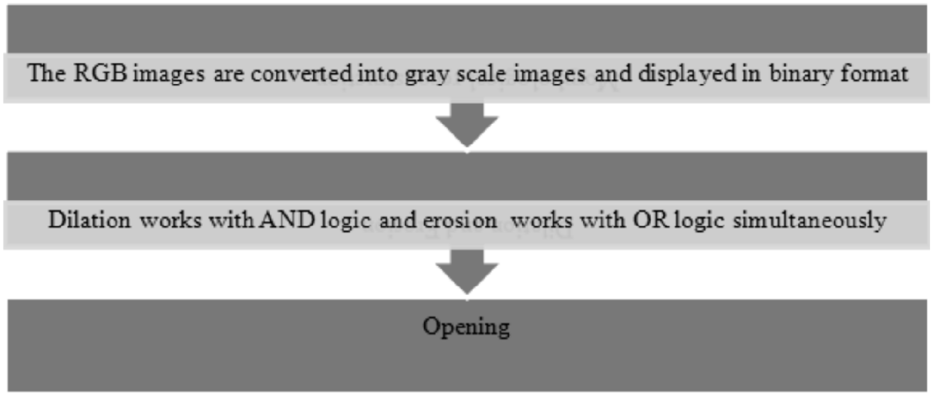

Figure 3: Working of Block Algorithm

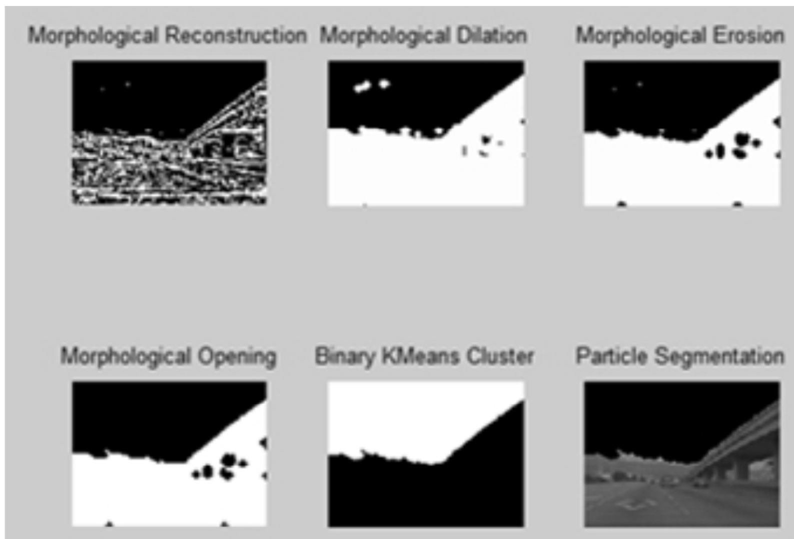

Figure 4: Morphological Process

(a) Feature Extraction

To calculate the motion features, we have to extract the motion fields from the videos. To this end, we have to apply a block-matching algorithm, over temporal features. A large temporal distance for the motion fields means that the motion vectors have correspondingly increased magnitudes. This helps to eliminate potential camera shakes, and also significantly reduce the effects of noise, since both of these will remain small in magnitude and thus easily detectable. The motion vectors are applied a threshold of 7 pixels and the features are extracted, its shown in fig 5.

Figure 5: Feature Extraction using PMES

(b) Classification

The SVM application is preferred because it's fast, efficient and user-friendly. Concerning the classes used, two different ones were used: sports and non- sports video. These did not seem to provide any problems for the features. Since shows like that are usually captured in the same colour patterns and are mostly identical. As a result, successful classification results after training were expected be quite high. The content varied from interviews sample of each class as well as videos that didn't belong to any class. The classes were at some cases overlapping, and the SVM for each one was trained and evaluated independently to impose a weight factor on the positive examples. The dataset was populated with instances of each positive training video in the training set until the number of positive examples became at least 1.2 times the number of negative examples. The small advantage on the positive examples reflected the fact that false positives are more serious than false negatives, and it is thus preferable to slightly favour positive examples over negative. One such example is shown in fig 6.

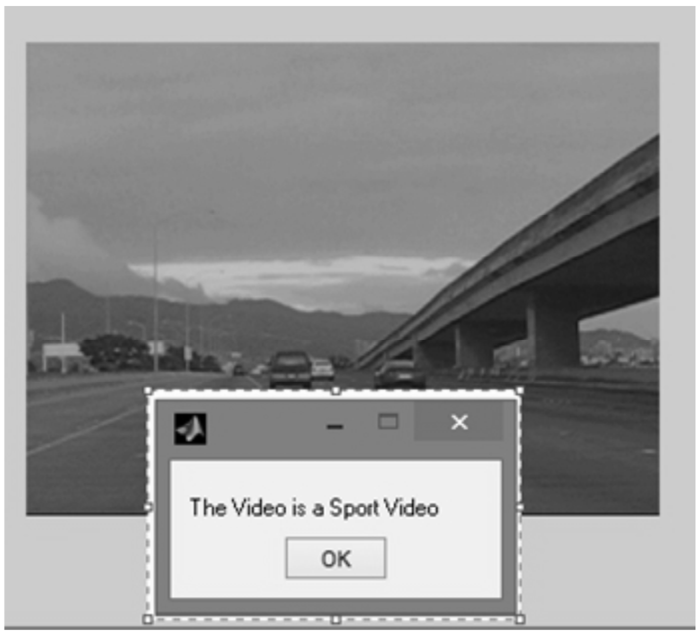

Figure 6: Classification of Videos

5. EXPERIMENTAL RESULTS

The proposed system was implemented using matlab and the experimental results are discussed below. The performance is calculated based on the existing algorithms like edge detection, colour histogram and proposed algorithm Perceived Motion Energy Spectrum (PMES). The comparison result is high when obtained with PMES. Three performance metrics are used. They are Precision, Recall and F-score.

Precision (positive predictive value) is the fraction of retrieved instances that are relevant to the video. Precision values are calculated for performance by the formula,

$$\boxed{Recall = \frac{True\ positive}{True\ positive + False\ negative}}$$

True Negative **:** motion feature was not established and not predicted

True Positive **:** motion feature was established and predicted

False Negative **:** motion feature was established and not predicted

False Positive **:** motion feature was not established and predicted

VIDEO 1 indicates Sports video, VIDEO 2 indicates Education video, VIDEO 3 indicates News video, VIDEO 4 indicates Action video, VIDEO 5 indicates Wildlife video and VIDEO 6 indicates Entertainment video.

Table 1
Performance Analysis Using Precision

	EDGE-DETECTION	*COLOUR-HISTOGRAM*	*PMES*
VIDEO1	0.75	0.65	0.72
VIDEO2	0.73	0.68	0.75
VIDEO 3	0.77	0.72	0.82
VIDEO 4	0.72	0.77	0.87
VIDEO 5	0.63	0.73	0.92
VIDEO 6	0.65	0.75	0.86

Precision calculation for performance graph is displayed by stating that the values in PMES are higher as shown in figure 7.

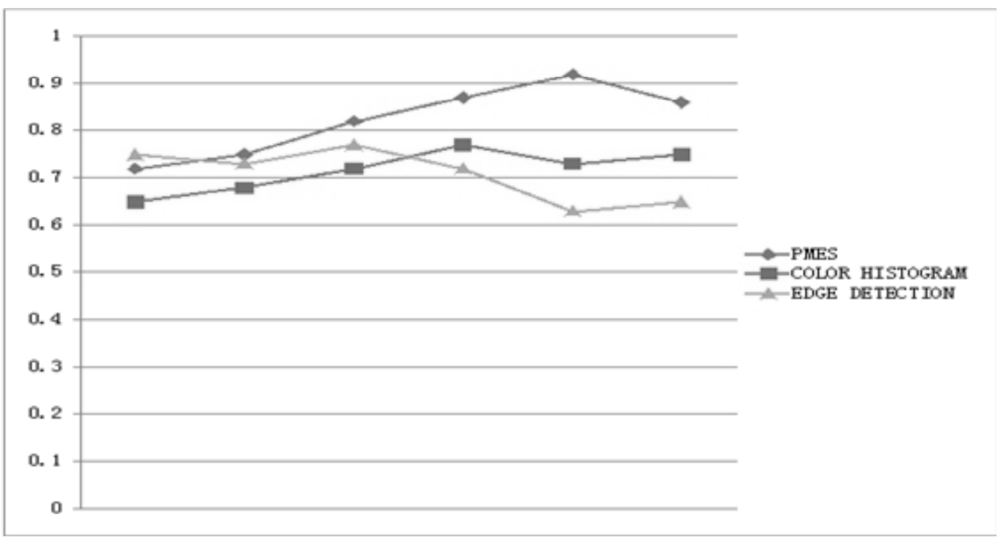

Figure 7: Precision Performance Analysis of PMES, Colour Histogram and Edge Detection

Recal**l** (sensitivity) is the fraction of relevant instances that are retrieved. Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved. Recall values are calculated for performance by the formula,

$$\boxed{Recall = \frac{True\ positive}{True\ positive + False\ negative}}$$

True Negative **:** motion feature was not established and not predicted

True Positive : motion feature was established and predicted

False Negative : motion feature was established and not predicted

False Positive : motion feature was not established and predicte.

Table 2
Performance Analysis Using Recall

	EDGE-DETECTION	*COLOR-HISTOGRAM*	*PMES*
VIDEO1	0.65	0.72	0.75
VIDEO2	0.68	0.75	0.73
VIDEO3	0.72	0.82	0.77
VIDEO4	0.77	0.87	0.72
VIDEO5	0.73	0.92	0.63
VIDEO6	0.785	0.97	0.645

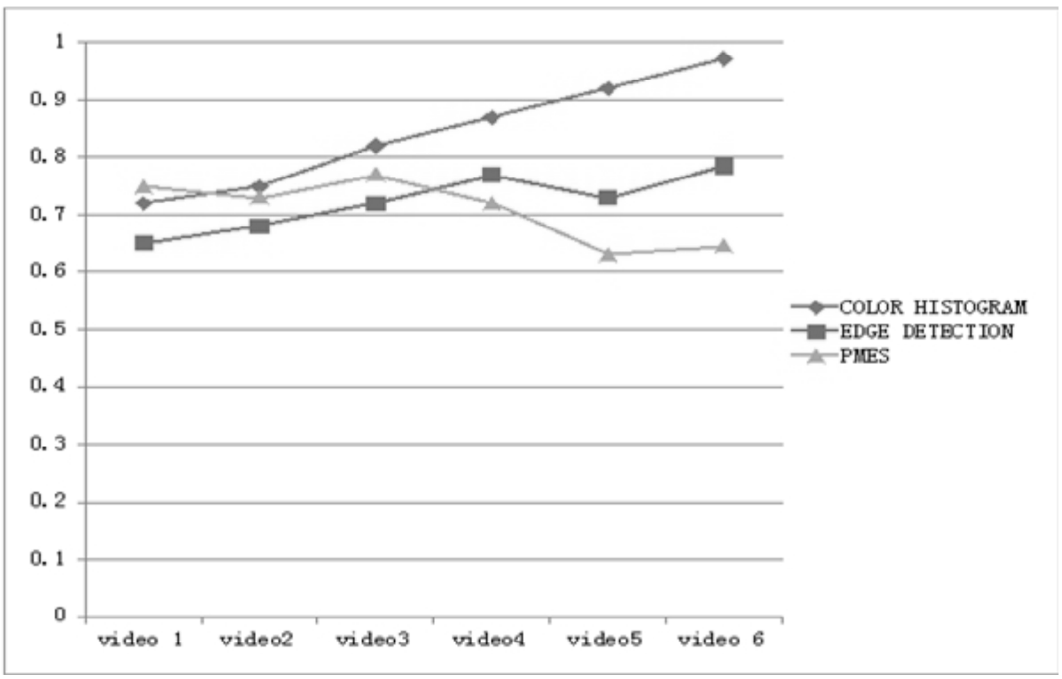

Figure 8: Recall Performance Analysis of PMES, Colour Histogram and Edge Detection

Recall calculation for performance graph is displayed by stating that the values in PMES are lower. The percentage of correct positives, correct negatives and overall instances appear in the training set as shown in figure 8

The F_1 score can be interpreted as a weighted average of the Precision and recall, where an F_1 score reaches its best value at 1 and worst score at 0.F-score values are calculated for performance by the formula,

$$F= \frac{2.precision.recall}{precision + recall}$$

Table 3
Performance Analysis Using F-score

	EDGE-DETECTION	*COLOR-HISTOGRAM*	*PMES*
VIDEO1	0.7	0.6	0.72
VIDEO2	0.71	0.62	0.75
VIDEO3	0.73	0.7	0.82
VIDEO4	0.7	0.73	0.87
VIDEO5	0.6	0.7	0.92
VIDEO6	0.62	0.71	0.86

F-Score calculation for performance graph is displayed by stating that the values in PMES are higher

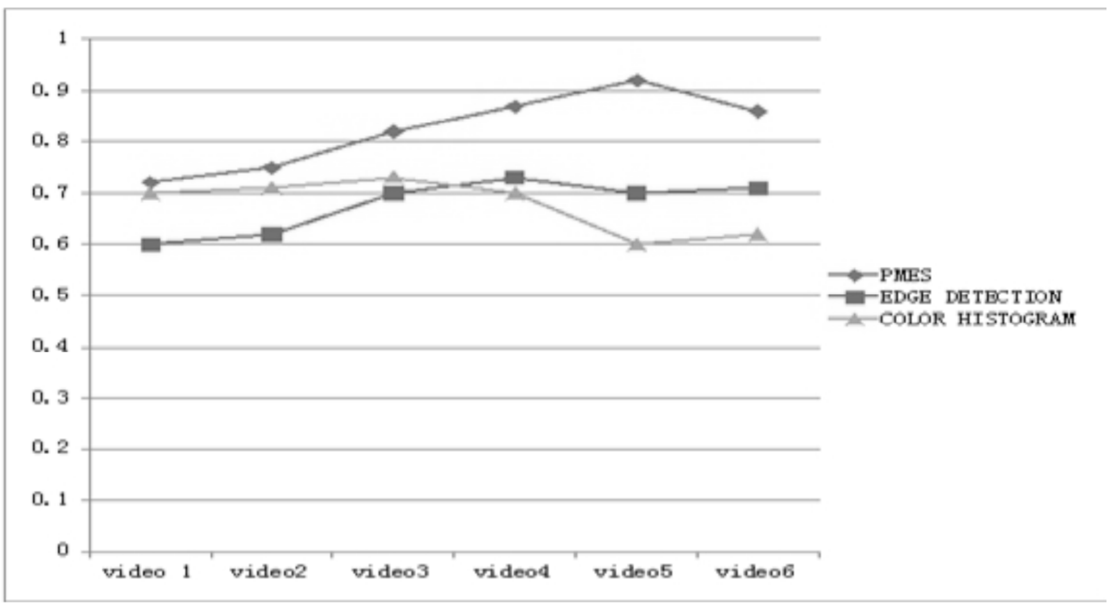

Figure 9: F-score Performance Analysis of Pmes, Colour Histogram and Edge Detection

The accuracy is calculated based on the existing algorithms like K-NN, Decision tree and proposed algorithm Support Vector Machine (SVM). The comparison result is high when obtained using SVM.

The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. Accuracy values are calculated for performance by the formula,

$$\text{Accuracy} = \frac{\text{number of true positives + number of true negative}}{\text{Number of (true positive +false positive+ false negative +true negative)}}$$

True Negative : not a sports video and not classified.

True Positive : sports video and classified.

False Negative : sports video and not classified.

False Positive : not a sports video but classified

Table 4
Accuracy Analysis of SVM, Decision Tree and K-NN

	K-NN	*DECISION TREE*	*SVM*
VIDEO1	0.65	0.45	0.75
VIDEO2	0.64	0.68	0.73
VIDEO3	0.64	0.78	0.77
VIDEO4	0.54	0.77	0.72
VIDEO5	0.72	0.73	0.87
VIDEO6	0.67	0.75	0.94

Accuracy calculation is displayed by stating that the values in SVM are higher. The percentage of correct positives, correct negatives and overall instances appear in the training set as shown in figure.

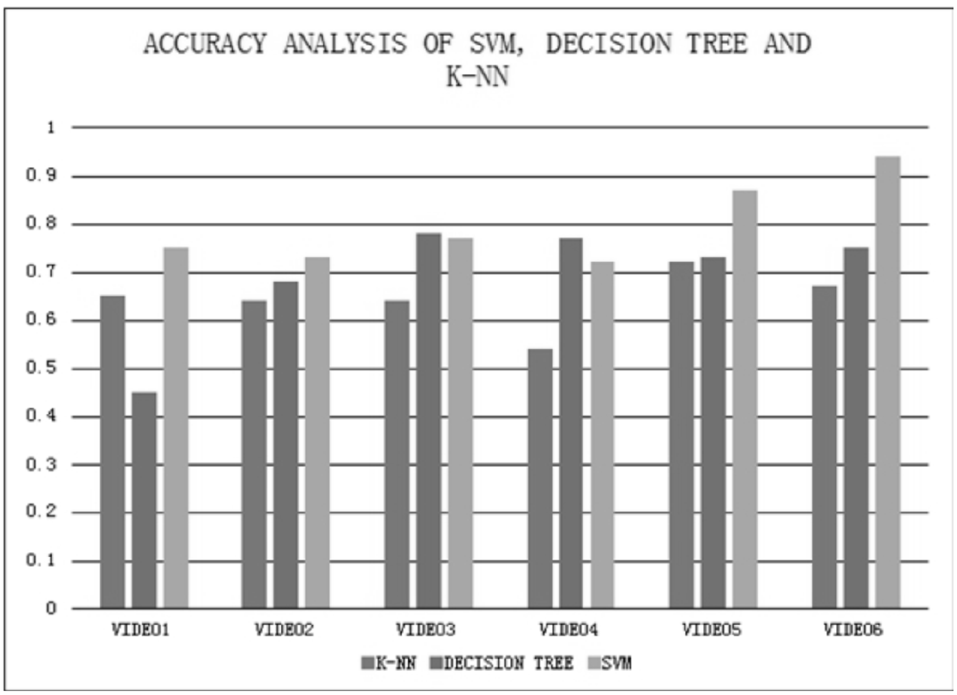

Figure 10: Accuracy Analysis of SVM, Decision Tree and K-NN

CONCLUSION

There are a number of methods used for key frame extraction but very less use the motion feature, the difference between consecutive frames to extract motion feature from key frame as used in video classification system. Most of the systems use only low level features such as colour, texture but these features provide incorrect results. Therefore there is a need to extract more features and integrate them for robust classification. High level features provide high accuracy and efficiency than the low level features. SVM efficiently classifies data sets and produces exact results. SVM works in any circumstances and it has ability to handle large data sets. It provides high performance, if there is salient object motion existing in the frame. SVM is more advantageous than the existing algorithms as the experimental results stand by it as a proof. Based on the results, PMES produces high performance of 89.4% than colour histogram and edge detection. The paper can be further enhanced by implementing multiple videos without restriction. Further feature extraction can be extended by including other features.

REFERENCES

[1] Usman Niaz, support vector machines content-based video retrieval based on motion information ,MultimediaTools And Applications, Kluwer Academic Publishers, pp 179-202, IEEE, 2014.

[2] Ashwini S.Mane, P.M.Kamde, video classification using svm, IJRTE, 2013.

[3] R. Datta, J. Li and J.Z. Wang, Content-based image retrieval: approaches and trends of the new age, Proceedings of the 7th International Workshop on Multimedia Information Retrieval, in conjunction with ACM International Conference on Multimedia, pp. 253-262, 2013.

[4] R. Datta, J. Li and J.Z. Wang, "Content-based image retrieval: approaches and trends of the new age," Proceedings of the 7th International Workshop on Multimedia Information Retrieval, in conjunction with ACM International Conference on Multimedia, pp. 253-262, 2005.

[5] Y. Rui, T.S. Huang, and S.F. Chang, "Image retrieval: current techniques, promising directions and open issues," Journal of Visual Communication and Image Representation, pp. 39-62, April 1999.

[6] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 1349-1380, Dec 2000.

[7] S. Z. Xiang, and T.S. Huang, "Relevance feedback in image retrieval: a comprehensive review," Multimedia Systems, pp. 536-544, April 2003.

[8] I. Koprinska, and S. Carrato, "Temporal video segmentation: a survey," Signal Processing: Image Communication, Elsevier, pp. 477- 500, January 2001.

[9] J. Calic, N. Campbell, S. Dasiopoulou, and Y. Kompatsiaris, "A survey on multimodal video representation for semantic retrieval," Computer as a tool (Eurocon 2005), the Third International Conference on, IEEE, pp. 135-138, November 2005

[10] C.G.M. Snoek, M. Worring, "Multimodal video indexing, a review of the state-of-the-art," Multimedia Tools and Applications, Springer Netherlands, pp. 5-35, January 2005.

[11] P. Aigrain, H. Zhang, and D. Petkovic, "Content-based representation and retrieval of visual media: a state-of-the-art review," Multimedia Tools and Applications, Kluwer Academic Publishers, pp 179-202, November 1996.

[12] F. Arman, R. Depommier, A. Hsu, and M.Y. Chiu, "Content-based browsing of video sequences," Multimedia, Proceedings of the Second ACM International Conference on, pp. 97-103, 1994.

[13] A.J.T. Lee, R.W. Hong M.F. Chang, "An approach to content-based video retrieval," Multimedia and Expo, IEEE International Conference on, pp. 273-276, Vol 1, 2004.