6th International Conference on Emerging Trends in Engineering, Technology, Science and Management
(IETE,Bengaluru) Institution of Electronics and Telecommunication Engineers, Ganganagar, Bengaluru, Karnataka, India (ICETETSM-17)
07th October 2017, www.conferenceworld.in ISBN: 978-93-86171-70-2

SPATIO TEMPORAL FEATURE EXTRACTION WITH VISUAL CONTENT CLUSTER AND VP-TREE INDEXING FOR VIDEO RETRIEVAL

Renuka Devi.S[1], Dr.S.Murugappan[2], S.Kokila[3]

[1,3]*Assistant professor , Department of Computer Science, KSIT, Bengaluru, (India)*

[2]*Associate professor, Department of Computer Science,*

Tamil Nadu Open University, Tamil Nadu, (India)

ABSTRACT

Video clustering and indexing is significant in order to improve performance of video retrieval. Recently, few research works has been designed for video indexing and retrieval. In order to overcome such limitation, Spatio Temporal Feature Extraction Based VP-Tree Indexing (STFE-VI) technique is proposed. The main objective of STFE-VI technique is to achieve higher video retrieval rate with minimum time. Initially, STFE-VI technique performs spatio temporal feature extraction in which visual content of video features i.e. colour, texture; shape and motion video data set. After that, clustering process is accomplished in order to group the video clips visual contents. Next, STFE-VI technique employed VP tree for indexing the clustered video clips based features. Finally, STFE-VI technique carried outs video retrieval process that efficiently extracts more similarly videos based on user query. The performance of STFE-VI technique is measured in terms of parameters such as clustering time, clustering accuracy, true positive rate of video retrieval and video retrieval time. The experimental results demonstrate that STFE-VI technique is able to enhance the true positive rate of video retrieval process and also reduces the video retrieval time when compared to state-of-the-art works.

Keywords— Spatio Temporal Features, videos, visual contents, VP tree, user query, indexing, clustering

I. INTRODUCTION

Discovering and retrieving similar videos from video collections is a significant problem is to solved. Video retrieval is more essential when videos are generated at increasing rate. Motivated by this challenge, many research works have been designed for video indexing and retrieving. However, performance of existing indexing and retrieving techniques was not sufficient to achieve higher true positive video retrieval rate.

A new automatic keyframe extraction method was developed in [1] for video indexing and retrieval. But, video retrieval rate was not sufficient. Histogram clustering technique was designed in [2] for improving performance of video content retrieval in which the similarity parameter employed to cluster the video objects into different groups. However, video retrieval time was higher.

Novel video retrieval architecture was presented in [3] in which the image query is given as input to videos database for retrieving video and improving retrieval scalability. However, performance of video retrieval was

 ISBN: 978-93-86171-70-2

not efficient. Content based video retrieval was performed in [4] by using Hadamard matrix and discrete wavelet transform and sparse representation to improve the accuracy of video retrieval rate. But, the true positive rate of video retrieval is not at required level.

A sample-based hierarchical adaptive K-means clustering method was intended in [5] for performing large-scale video retrieval. This clustering method improves video retrieval rate and reduce the time complexity. However, video retrieval time was more. Visual Semantic Based 3D Video Retrieval technique was introduced in [6] Using Hadoop Distributed File System for clutering the video contents and to reduce time for video retrieval process. Though, clustering accuracy was remained unsolved.

A novel framework was designed in [7] for multimodal video indexing and retrieval through shrinkage optimized directed information assessment (SODA). However, this framework does not provide sufficient information for accurate indexing which lacks video retrieval rate and increases the retrieval time. An active learning approach was developed in [8] for improving the overall video indexing performance.

Rebuilding Visual Vocabulary through Spatial-temporal Context Similarity was presented in [9] for Video Retrieval and to solve the OverQuantize problem. A novel framework was intended in [10] in which the video was first indexed according to temporal, textual, and visual features and followed by implicit user feedback analysis was realized with help of a graph-based methodology. But, precision and recall of video retrieval process was poor.

In order to overcome the above mentioned existing limitations, Spatio Temporal Feature Extraction Based VP-Tree Indexing (STFE-VI) technique is proposed. The STFE-VI technique is designed for improving video indexing and video retrieval process.

The research objective of STFE-VI technique is formulated as follows,

- To efficiently extract the different video features in data set, Spatio temporal feature extraction is performed in STFE-VI technique.
- To improve the clustering accuracy with lower time, visual content clustering is used in STFE-VI technique.
- To enhance the performance of video retrieval with minimum time, VP-tree indexing is carried out in STFE-VI technique.

The rest of the paper structured is as follows. In Section 2, the proposed STFE-VI is described with the help of neat architecture diagram. In Section 3, simulation environment is presented with exhaustive analysis of results explained in Section 4. In Section 5, summary of different video retrieval techniques are explained. The Section 6 concludes the remarks are presented.

1.1 SPATIO TEMPORAL FEATURE EXTRACTION BASED VP-TREE INDEXING TECHNIQUE

The STFE-VI technique is designed to perform spatio temporal feature extraction based video indexing and video retrieval. Video Retrieval is a process of retrieving relevant videos based on user query. For an input user query , the probability of query being related with a given video data set is determined by using $P(Q)|V)\forall V \in$. Besides, video indexing is a process of storing videos in a sorted order of their features to provide easy way for searching video clips. The STFE-VI technique is used VP tree structure for indexing and retrieving the video clips. The overall architecture diagram of Spatio Temporal Feature Extraction Based VP-Tree Indexing technique is shown in below,

ISBN: 978-93-86171-70-2

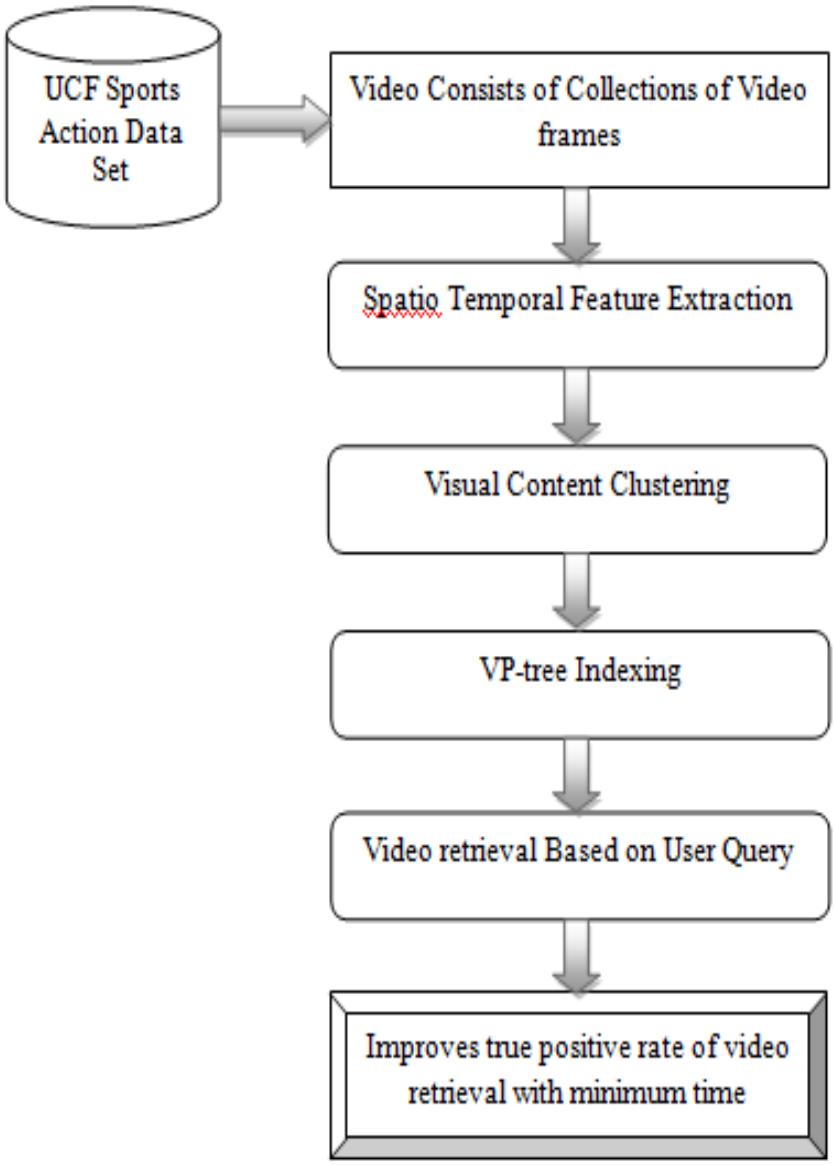

Fig.1 Architecture Diagram of Spatio Temporal Feature Extraction Based VP-Tree Indexing Technique

As shown in Figure 1, STFE-VI technique initially takes UCF Sports Action Data Set as input. The video comprises many video frames. Next, Spatio Temporal Feature Extraction is carried out that extracts the visual features (i.e. color, texture; shape and motion) in video frames by using spatio temporal texture map. After that, video clustering process is carried out depends on their visual contents of video frames in given data set. This helps for improving the clustering accuracy and reducing clustering time in an effective manner. Then, STFE-VI technique is used VP-tree structure for indexing the clustered videos with respect to their visual contents of spatio temporal features. Finally, video retrieval process mines more similar videos related to the input user queries. Therefore, STFE-VI technique increases the true positive rate of video retrieval with lower time. The detailed explanation of STFE-VI technique is described in subsequent sections,

1.1.1 Spatio Temporal Feature Extraction based Visual Content Clustering

 ISBN: 978-93-86171-70-2

Feature extraction process is significant to improve performance of video clustering and retrieval. The STFE-VI technique used spatio temporal texture map to extract spatial and temporal features of video frames with low computational complexity. Different interest events in videos are differentiated by motion variations of image structures over time.

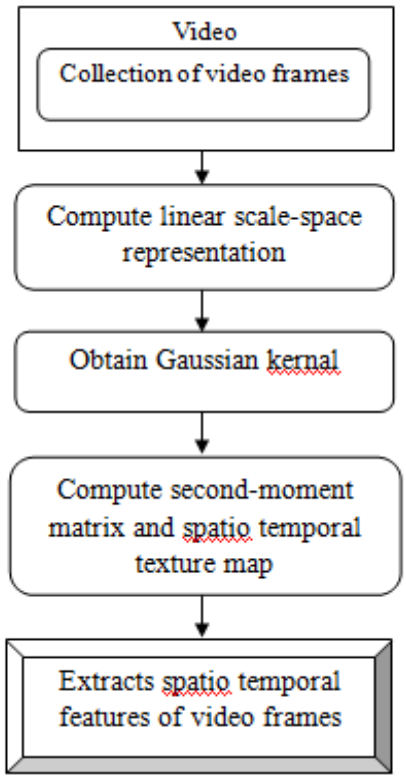

Fig.2. Spatio Temporal Feature Extraction

Figure 2 shows the spatio temporal feature extraction process in videos. As shown in Figure 2, for an input video sequence V, linear scale-space representation L is first constructed through the convolution of V with a 3D Gaussian kernel which is formulated as,

$$L(\cdot\,;\sigma_l^2,\tau_l^2) = g(\cdot\,;\sigma_l^2,\tau_l^2) * V(\cdot) \quad (1)$$

From the equation (1), σ_l^2 and τ_l^2 represents the spatio variance and temporal variance of Gaussian kernel g respectively. The spatio temporal Gaussian kernel is mathematically defined as below,

$$g(x,y,t;\sigma_l^2,\tau_l^2) = \frac{\exp\left(-(x^2+y^2)/2\sigma_l^2 - t^2/2\tau_l^2\right)}{\sqrt{(2\pi)^3\sigma_l^4\tau_l^2}} \quad (2)$$

From the equation (2), x and y indicate x and y axes of frame form the input video V (i.e. spatio domain) and t refers the time axis (i.e. temporal domain). The distinct point is identified through considering a Gaussian window in the image and discovering the location in which Video V has significant changes of image intensity in space and time domains by shifting the window by a small amount in various directions. Such points are identified by convolution of a spatiotemporal second-moment matrix with a Gaussian weighting function $g(\cdot\,;\sigma_l^2,\tau_l^2)$. The second-moment matrix is a 3 × 3 matrix includes of first-order spatial and temporal derivatives which is mathematically represented as,

ISBN: 978-93-86171-70-2

$$\mu = g(\cdot;\sigma_l^2,\tau_l^2) * \begin{pmatrix} L_x^2 & L_xL_y & L_xL_t \\ L_xL_y & L_y^2 & L_yL_t \\ L_xL_t & L_yL_t & L_t^2 \end{pmatrix} \quad (3)$$

Thus, L_x, L_t and L_y denotes first order derivates which mathematically expressed as,

$$L_x(\cdot;\sigma_l^2,\tau_l^2) = \partial_x(g * V) \quad (4)$$

$$L_y(\cdot;\sigma_l^2,\tau_l^2) = \partial_y(g * V) \quad (5)$$

$$L_t(\cdot;\sigma_l^2,\tau_l^2) = \partial_t(g * V) \quad (6)$$

From the equation (6),$\sigma_i^2 = s\sigma_l^2$, $\tau_i^2 = s\tau_l^2$ and s is a constant. The large eigen values λ_1, λ_2 and $\lambda_1$of second-moment matrix point outs existence of distinct points in V. The variations of image intensity in spatiotemporal domain are then obtained by means of combining the determinant and the trace of μ to construct the extended Harris corner function for the spatiotemporal domain which is formulated as,

$$H = \det(\mu) - k\ trace^3(\mu) = \lambda_1,\lambda_2,\lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3) \quad (7)$$

From the equation (7), k is a constant. The H function is then normalized to reduce the effect of illumination variations of the images. By using H as a texture map, the spatio temporal features of video frames are efficiently extracted.

After extracting the features, video clustering process is carried out based on their visual contents (i.e. color, texture; shape and motion) of obtained spatio temporal features. The process of Spatio Temporal Feature Extraction based Visual Content Clustering is shown in below Figure 3. As demonstrated in Figure 3, the visual content clustering process initially takes UCF Sports Action Data Set and extracted spatio temporal features as input.

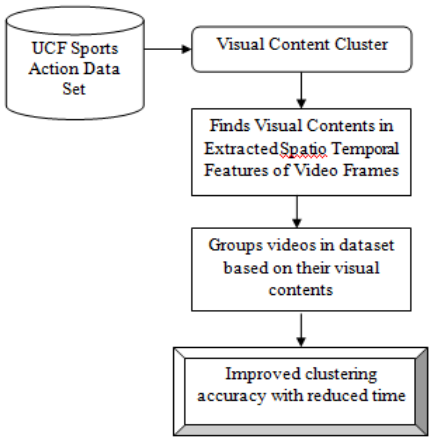

Fig.3.Process of Spatio Temporal Feature Extraction based Visual Content Clustering

Next, visual content cluster finds visual contents in extracted spatio temporal features of video frames and finally groups videos in dataset based on their visual contents. Therefore, STFE-VI technique improves the clustering accuracy and also reduces the clustering time of videos in a significant manner. The algorithmic process of Spatio Temporal Feature Extraction based visual content clustering (STFE-VCC) is shown in below,

// Spatio Temporal Feature Extraction based visual content clustering Algorithm

 ISBN: 978-93-86171-70-2

Input: Videos '$V_i = V_1, V_2, .. V_n$', set of frames $frame_i = frame_1, frame_2, .., frame_n$.

Output: Improved clustering accuracy with lower time

Step 1: Begin

Step 2: For each video V_i

Step 3: For each video frame $frame_i$

Step 4: Construct linear scale-space representation L of the input video V of frame using (1)

Step 5: Obtain Gaussian kernel using (2)

Step 6: Compute the second-moment matrix using (3)

Step 7: Measure the spatio temporal texture map H using (7)

Step 8: Extract the spatio temporal features

Step 9: Groups the videos in given data set based on their visual contents in extracted spatio temporal features

Step 10: Outputs clustered video clips

Step 11:End for

Step 12:End for

Step 13:End

Algorithm 1 Spatio Temporal Feature Extraction Based Visual Content Clustering

As shown in Algorithm 1, STFE-VCC algorithm initially extracts spatio temporal features for each video frames in given videos by using texture map H. After extracting the collection of features, clustering process is performed in order to group the videos in given data set based on the visual contents of spatio temporal features. This in turn assists for enhancing the clustering accuracy and reducing the clustering time.

1.1.2 VP-tree for Video Indexing

The STFE-VI technique used VP-tree for indexing the clustered videos based on their visual contents of spatio temporal features in video. In VP-tree, the storing of video object is represented by a circle. Each node of VP tree comprises an input point and a radius. All the left children of a given node are positioned inside the circle and all the right children of a given node are positioned in outside of the circle. The tree itself does not want to know any other information regarding what is being stored. All it requires is the distance function which fulfils the properties of the metric space. Let consider a circle with a radius. The left children are all placed within the circle and the right children are placed outside the circle as shown in Figure 4.

ISBN: 978-93-86171-70-2

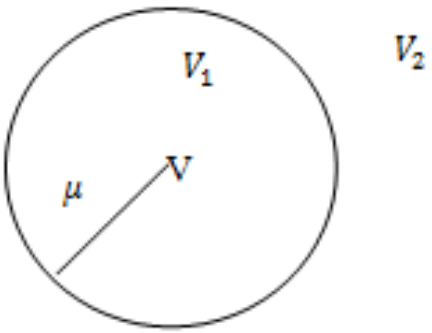

Fig.4. Way of Storing Video Objects In Vantage-Point Tree

Let consider a clustered data set V_i that consist of N video objects. For each node in the tree, a video object is elected to be the vantage point by using a Vantage Point Selection. Let considered video is selected for the root node is v and μ be the median of the distance values of all the other video objects in V_i with respect to v and V_i is divided into two subsets of approximately equal sizes as V_1 and V_2 which is mathematically formulated as follows,

$$V_1 = \{v \epsilon V | d(v, VP) < \mu\} \quad (8)$$

$$V_2 = \{v \epsilon V | d(v, VP) \geq \mu\} \quad (9)$$

From the equation (8) and (9), *d(v,VP)* epresents the distance among the video objects *v* and *VP*. The node partitioning concepts of Vantage Point Tree is shown in below Figure 5.

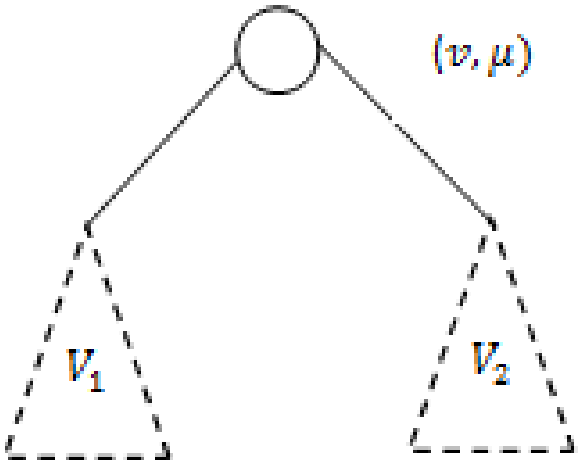

Fig.5. Structure of Node Partitioning for Vantage-Point Tree

This partitioning concept is then applied to nodes V_1 and V_2 recursively. Each subset such as V_1 and V_2 , corresponds to one node of the VP-tree. For each node, a distinct vantage point is selected to store the video object in the consequent subset. VP-tree stores numerous video objects at one leaf node. Finally, the whole clustered video objects are sorted out as a balanced tree as in other spatial index structures.

The structure of a VP-tree is very simple in which each node is of the form (VP, MD,R_{ptr},L_{ptr}). Here VP denotes the vantage point and MD refers to the median distance among all the video objects indexed below that node whereas R_{ptr} and L_{ptr} are pointers to the left and right branches. Left branch of the node indexes the

 ISBN: 978-93-86171-70-2

video objects whose distances from VP are less than or equal to MD. Subsequently, right branch of the node indexes the video objects whose distances from VP are greater than or equal to MD. In leaf nodes, instead of pointers to the left and right branches, references to the video objects are kept. The median distance between the vantage point VP and the video object V_i is determined by using following mathematical formula,

$$d(VP, V_i) = \sqrt{(VP - \sum_{i=1}^{N} V_i)^2} \qquad (10)$$

From the equation (10), median distance is measured. Given a data set of clustered N video objects $V_i = \{V_1, V_2, .., V_N\}$, and a median distance function $d\ (VP, V_i)$, a VP tree is constructed by using the following algorithmic process,

// VP tree based Video Indexing Algorithm

Input: Clustered N video objects $V = \{V_1, V_2, .., V_N\}$

Output: Create VP tree for Indexing of Video objects

Step 1: Begin

Step 2: if |V|=0, then construct a empty tree

Step 3: M = median of { $d(VP, V_i)) \mid V_i \in V$}

Step 3: For each clustered video object V_i

Step 4: Randomly select vantage point VP form a data set

Step 5: Measure distance from the vantage point $VP$to the video object V_i using (10)

Step 5: Calculate the mean and variance of these distances

Step 6: if $d(VP, V_i) \leq M$, then

Step 7: video object V_i is stored in left branch of the tree

Step 8: else if $is(VP, V_i) \geq M$, then

Step 9: video object V_i is stored in right branch of the tree

Step 10:end if

Step 11:end if

Step 12:End for

Step 13:End

Algorithm 2 VP tree based Video Indexing Algorithm

By using the above algorithmic process, clustered video clips are efficiently stored in VP tree structure based on their visual features. This in turn helps for enhancing the true positive rate of video retrieval with reduced retrieval time.

1.2 Video Retrieval

 ISBN: 978-93-86171-70-2

Video retrieval is a process of retrieving relevant videos from the video data base based on user query. For retrieving the videos, initially user query is given as input. After that, the user queried videos are searched and outputs to the corresponding user. The process of Video retrieval is shown in below Figure 6.

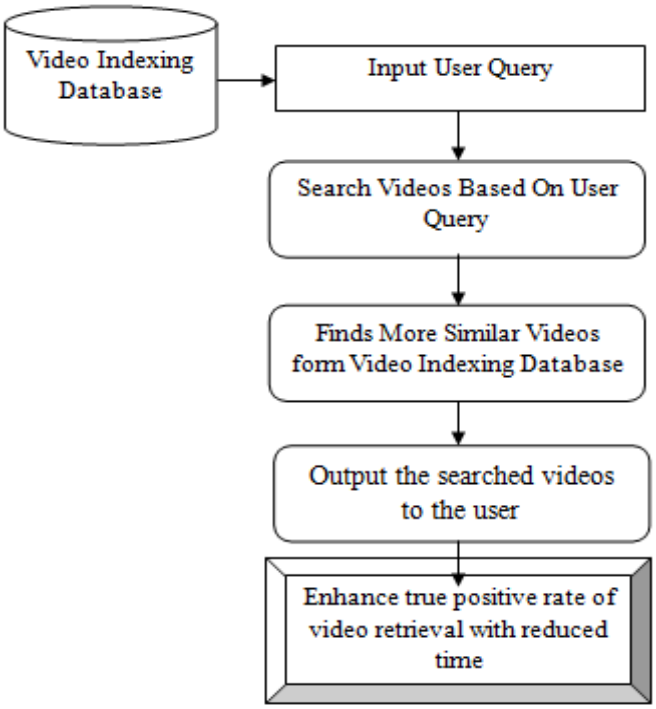

Fig.6.Video retrieval processes

Figure 6 demonstrates block diagram of video retrieval process. For a given user query Q, the set of video objects that are within distance r of Q are retrieved using the search algorithm given below,

// Video Retrieval Algorithm

Input: User query $Q = Q_i, Q_2, Q_3, \dots Q_N$, r is query range, VP is a vantage point of a node that visit during the search, and M is the median distance value for the same node

Output: Improved True Positive Rate of Video Retrieval with lower time

Step 1: Begin

Step 2: For each User query Q

Step 3: If $d(Q, VP) < r$, then the vantage point at the root

Step 4: If $d(Q, VP) + r \geq M$, **then**

Step 5: Search the right branch of tree

Step 6: Else if $d(Q, VP) - r \leq M$, **then**

Step 7: Search the left branch of tree

Step 8: End if

Step 9: End if

Step 10:If both search conditions are satisfied, **then**

Step 11:both branches of tree is searched for retrieving user queried video objects

Step 12:outputs the searched videos to the user

 ISBN: 978-93-86171-70-2

Step 12:End if

Step 13:End for

Algorithm 3 Video Retrieval Algorithm

By using the above algorithmic process, STFE-VI technique efficiently retrieves video clips from the VP tree indexing based on user query. Therefore, STFE-VI technique increases the true positive rate of video retrieval and reduces the video retrieve time in an effectual manner.

II. EXPERIMENTAL SETTINGS

The Spatio Temporal Feature Extraction Based VP-Tree Indexing (STFE-VI) technique is implemented in Java Language with aid of UCF Sports Action Data Set. The UCF Sports dataset includes of a set of actions collected from a different of sports. The UCF Sports Action dataset comprises 150 sequences with the resolution of 720 x 480. The UCF Sports Action Data Set contains 10 diverse actions of videos. The effectiveness of STFE-VI technique is compared against with the existing two methods namely Automatic Shot based Keyframe Extraction [1] and Histogram clustering technique [2] respectively. The performance of STFE-VI technique is measured in terms of clustering accuracy, clustering time, true positive rate of video retrieval and video retrieval time.

III. RESULTS AND DISCUSSIONS

The performance of STFE-VI technique is compared against with exiting two methods namely Automatic Shot based Key frame Extraction [1] and Histogram clustering technique [2]. The performance of STFE-VI technique is evaluated along with the following metrics with the help of tables and graphs.

3.1 Measurement of Clustering Accuracy

In STFE-VI technique, clustering accuracy (CA) is measures the ratio of number of correctly clustered videos based on their visual contents of spatio temporal features to the total number of videos. The clustering accuracy is measured in terms of percentages (%) and mathematically represented as follows,

$$CA = \frac{number\ of\ correctly\ clustered\ videos\ based\ on\ their\ visual\ contents}{total\ number\ of\ videos} * 100 \qquad (11)$$

From the equation (11), the clustering accuracy of videos is measured. While the clustering accuracy of videos is higher, the method is said to be more efficient.

The clustering accuracy result is obtained based on dissimilar number of videos taken in the range of 10-100 is presented in Table 1. The STFE-VI technique considers the framework with dissimilar number of videos for performing experimental works using Java Language.

ISBN: 978-93-86171-70-2

Table1. Tabulation for Clustering Accuracy

Number of videos	*Clustering Accuracy (%)*		
	Automatic Shot based Key frame Extraction	*Histogram clustering technique*	*STFE-VI technique*
10	65.02	73.26	84.56
20	70.14	75.26	85.92
30	72.05	75.98	87.39
40	73.69	77.65	88.14
50	75.18	78.67	88.98
60	79.58	83.16	89.52
70	86.12	89.21	90.14
80	87.16	89.93	91.26
90	87.97	90.13	92.06
100	88.23	90.95	94.13

The clustering accuracy result is obtained based on dissimilar number of videos taken in the range of 10-100 is presented in Table 1. The STFE-VI technique considers the framework with dissimilar number of videos for performing experimental works using Java Language. From the table value, it is expressive that the clustering accuracy using proposed STFE-VI technique is higher as compared to other existing methods [1], [2].

Figure 7 depicts the impact of clustering accuracy with respect to different number of videos using three methods. As exposed in figure, the proposed STFE-VI technique provides better clustering accuracy for efficient video retrieval as compared to existing Automatic Shot based Key frame Extraction [1] and Histogram clustering technique [2]. Besides, while increasing the number of videos, the clustering accuracy is also gets increased by using three methods. But, comparatively the clustering accuracy using STFE-VI technique is higher.

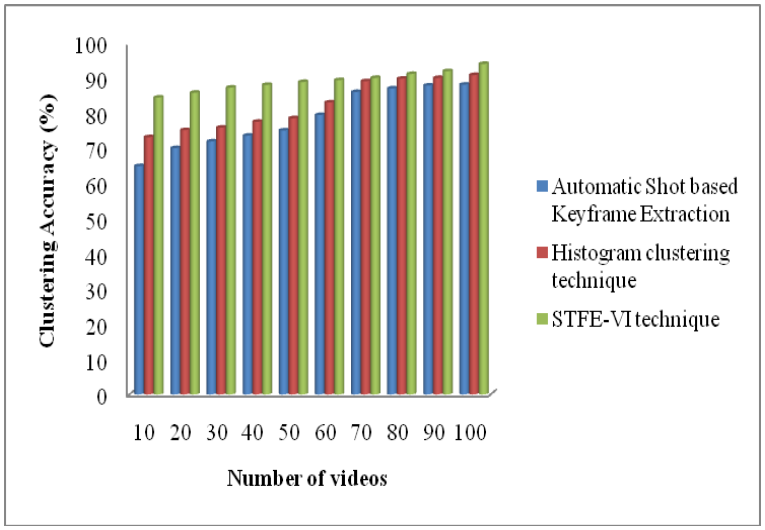

Fig.7.Clustering Accuracy Vs Number of Videos

 ISBN: 978-93-86171-70-2

This is because of the application of spatio temporal feature extraction based visual content clustering algorithm that efficiently groups the videos in given data set with respect to their visual contents of spatio temporal features in video frames. This assists for improving the clustering accuracy in a significant manner. Therefore, STFE-VI technique improves the clustering accuracy by 15% when compared to Automatic Shot based Keyframe Extraction [1] and 9% when compared to Histogram clustering technique [2] respectively.

3.1.1. Measurement of Clustering Time

In STFE-VI technique, clustering time (CT) measures the amount of time taken for clustering the videos based on their visual features. The clustering time is measured in terms of milliseconds (ms) and mathematically represented as follows,

$$CT = n * time(clusering\ the\ one\ video) \quad (12)$$

From the equation (12), clustering time is measured in which n denotes the number of videos taken. While the clustering time of videos is lower, the method is said to be more efficient.

Table.2.Tabulation for Clustering Time

Number of videos	*Clustering Time (ms)*		
	Automatic Shot based Keyframe Extraction	*Histogram clustering technique*	*STFE-VI technique*
10	23.1	15.8	9.15
20	39.6	27.6	11.8
30	52.8	35.2	15.2
40	66.1	41.5	19.7
50	79.3	49.6	25.3
60	92.7	55.3	28.9
70	99.8	58.2	35.6
80	101.5	67.1	39.5
90	105.2	72.8	43.8
100	112.8	78.6	48.1

Table 2 demonstrates the result analysis of clustering time with respect to different number of videos taken in the range of 10-100 using three methods. From the table value, it is clear that the clustering time using proposed STFE-VI technique is lower as compared to other existing methods [1], [2].

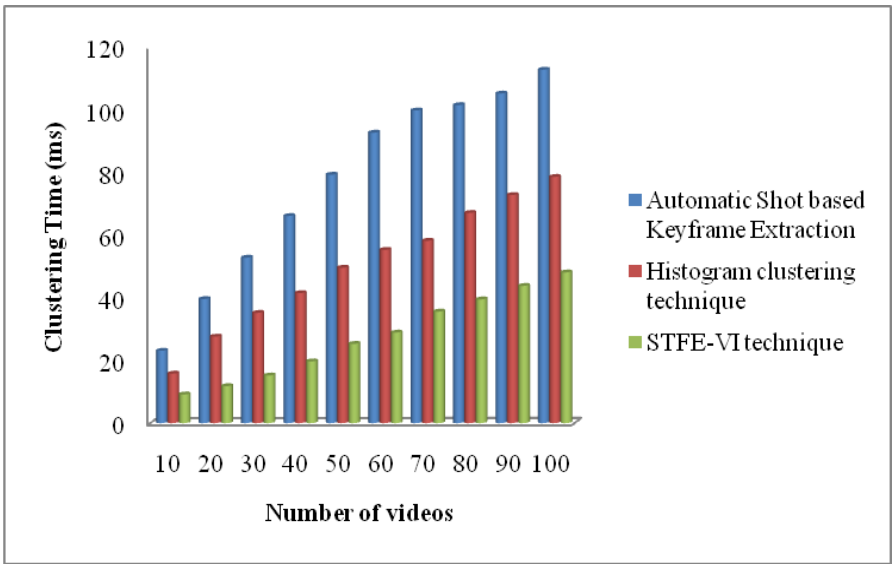

Fig.8. Clustering Time Vs Number of Videos

 ISBN: 978-93-86171-70-2

Figure 8 illustrates the impact of clustering time based on dissimilar number of videos using three methods. As exposed in figure, the proposed STFE-VI technique provides better clustering time for video retrieval as compared to existing Automatic Shot based Keyframe Extraction [1] and Histogram clustering technique [2]. In addition, while increasing the number of videos, the clustering time is also gets increased by using all three methods. But, comparatively the clustering time using STFE-VI technique is lower. This is owing to the application of spatio temporal feature extraction based visual content clustering algorithm in STFE-VI technique where it classifies the videos in given data set based on their visual contents of spatio temporal features in video frames with minimum time. This in turn supports for reducing the clustering time in an effective manner. Thus, STFE-VI technique reduces the clustering time by 65% when compared to Automatic Shot based Keyframe Extraction [1] and 46% when compared to Histogram clustering technique [2] respectively.

3.2. Measurement of True Positive Rate of Video Retrieval

In STFE-VI technique, true positive rate (TPR) of video retrieval is defined as the ratio of number of correctly retrieved videos based on user query to the total number of videos. The true positive rate of video retrieval is evaluated in terms of percentages (%) and expressed as,

$$TPR = \frac{number\ of\ correctly\ retrieved\ videos\ based\ on\ user\ query}{total\ number\ of\ videos} * 100 \qquad (13)$$

From the equation (13), true positive rate is measured. While the true positive rate of video retrieval is higher, the method is said to be more efficient.

Table.3.Tabulation for True Positive Rate Of Video Retrieval

Number of videos	*True Positive Rate of Video Retrieval (%)*		
	Automatic Shot based Keyframe Extraction	*Histogram clustering technique*	*STFE-VI technique*
10	69.26	79.46	87.46
20	71.69	82.65	88.15
30	73.26	84.19	88.91
40	76.91	85.67	90.52
50	78.69	87.16	90.98
60	80.66	89.05	91.25
70	85.69	90.35	92.83
80	88.16	90.98	93.46
90	88.55	91.23	93.90
100	89.37	92.85	94.62

Table 3 portrays the true positive rate of video retrieval is obtained using three methods versus different number of videos taken in the range of 10-100. From the table value, it is expressive that the true positive rate of video retrieval using proposed STFE-VI technique is higher as compared to other existing methods [1], [2].

ISBN: 978-93-86171-70-2

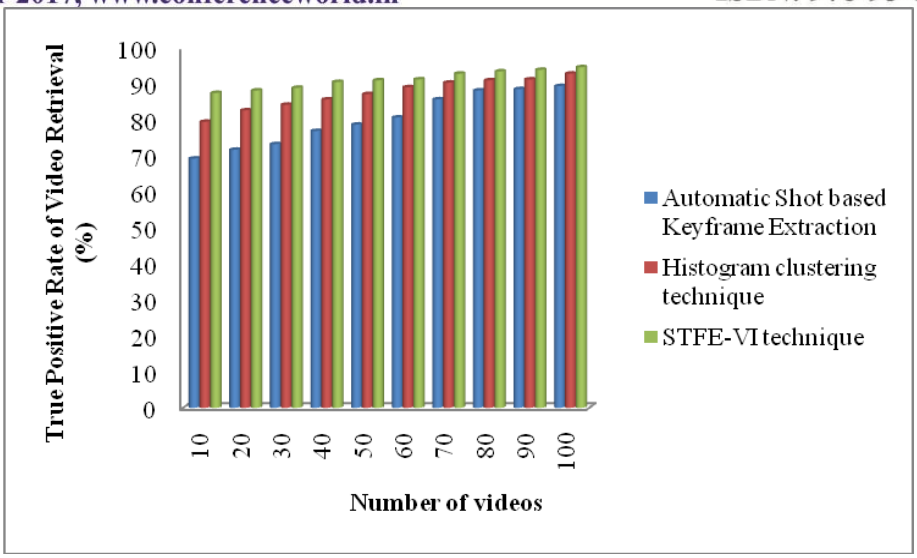

Fig.9. True Positive Rate of Video Retrieval Vs Number of Videos

The true positive rate of video retrieval is obtained with respect to dissimilar of videos using three methods is shown in Figure 9. As illustrated in figure, the proposed STFE-VI technique provides better true positive rate of video retrieval when compared to existing Automatic Shot based Keyframe Extraction [1] and Histogram clustering technique [2]. Further, while increasing the number of videos, the true positive rate of video retrieval is also gets increased by using all three methods. But, comparatively the true positive rate of video retrieval using STFE-VI technique is higher. This is owing to the usage of VP-tree indexing in STFE-VI technique. B-tree indexing efficiently stores the clustered video according to their extracted features of video contents. This helps for enhancing the true positive rate of video retrieval in an efficient manner. Therefore, STFE-VI technique improves the true positive rate of video retrieval by 14% when compared to Automatic Shot based Keyframe Extraction [1] and 5% when compared to Histogram clustering technique [2] respectively.

3.3.Measurement of Video Retrieval Time

In STFE-VI technique, video retrieval time (VRT) refers the amount of time taken for retrieving the videos based on user query. The video retrieval time is evaluated in terms of milliseconds (ms). While the video retrieval time is lower, the method is said to be more efficient.

Table.4.Tabulation for Video Retrieval Time

Size of video (MB)	*Video Retrieval Time (ms)*		
	Automatic Shot based Keyframe Extraction	*Histogram clustering technique*	*STFE-VI technique*
113.6	16.2	12.5	7.1
323.7	18.6	14.2	11.5
349.5	22.9	17.8	13.8
454.5	27.5	19.3	17.6
635.2	34.6	21.5	20.3
905.3	37.9	25.7	22.6
936.2	41.3	27.6	23.7
970.6	45.8	31.4	26.9

ISBN: 978-93-86171-70-2

1000.1	49.6	35.9	29.2
1040.7	55.3	42.8	31.5

Table 4 shows the video retrieval time for proposed STFE-VI technique and existing methods [1], [2]. From the table value, it is clear that the video retrieval time using proposed STFE-VI technique is lower as compared to other existing methods [1], [2].

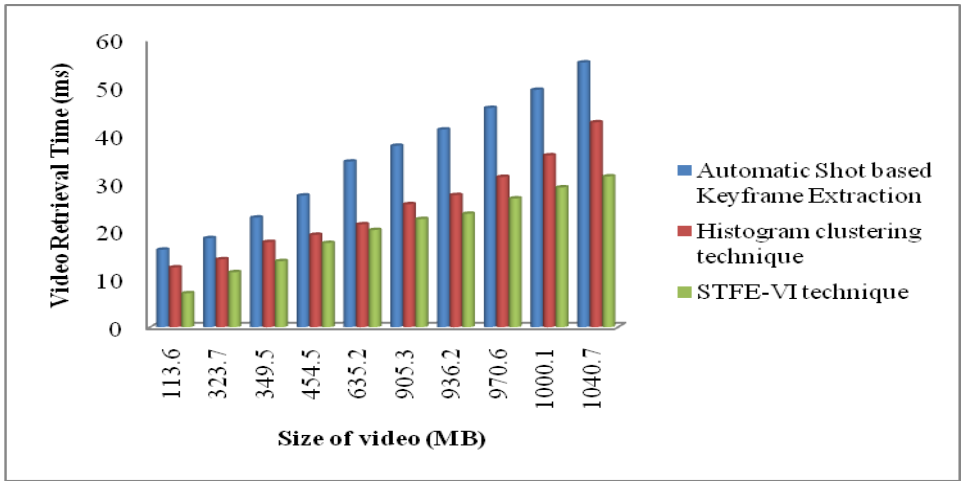

Fig.10. Video Retrieval time Vs Number of Videos

Figure 10 describes video retrieval time is obtained with respect to different number of video sizes using three methods. As revealed in figure, the proposed STFE-VI technique provides better video retrieval time as compared to existing Automatic Shot based Keyframe Extraction [1] and Histogram clustering technique [2]. Moreover, while increasing the number of videos, the video retrieval time is also gets increased by using all three methods. But, comparatively the video retrieval time using STFE-VI technique is lower. This is because of the application of VP-tree indexing in STFE-VI technique. With the assist of VP-tree indexing, STFE-VI technique retrieves similar videos based on user query with lower time. This in turn decreases the video retrieval time. Hence, STFE-VI technique reduces the video retrieval time by 42% when compared to Automatic Shot based Keyframe Extraction [1] and 19% when compared to Histogram clustering technique [2] respectively.

IV. RELATED WORKS

An Efficient Genre-Specific Semantic Video Indexing was carried out in [11] to enhance the video retrieval performance and to identify visual-based genre-specific concepts in a more effective manner. But, video retrieval performance was poor. A novel and efficient method jointing audio features was developed in [12] for video retrieval query by example in which the visual features are employed to refine retrieval.

A review of different techniques designed for content based video retrieval systems and their trends, challenges was analyzed in [13]. A novel algorithm was introduced in [14] for performing content-based video indexing and retrieval using key-frames texture, edge, and motion features. This algorithm improves average retrieval rate of videos. But, clustering accuracy and clustering time was remained unaddressed.

 ISBN: 978-93-86171-70-2

Video Retrieval performance was enhanced in [15] where the classification of Video Database was accomplished using multiple frames based on texture information. However, spatio temporal feature extraction was not considered. Content based video retrieval system was designed in [16] with help of entropy based shot detection method to perform video indexing and retrieval. But, video retrieval performance was not effectual.

A Matrix Based Sequential Indexing Technique was presented in [17] for Video Mining which retrieves more number of relevant videos based on the user input query. But, retrieval time was more. A review of different techniques designed for video indexing was analyzed in [18]. An information extraction techniques was implemented in [19] for performing automatic semantic video indexing to improve the video retrieval performance. A tree based classifier was used in [20] to discover the genre of the query video and to mine the similar genre videos from the video database.

V. CONCLUSION

An efficient Spatio Temporal Feature Extraction Based VP-Tree Indexing (STFE-VI) technique is developed in order to enhance the video retrieval performance with higher true positive rate. The key objective of STFE-VI technique is to attain higher video retrieval rate and video retrieval time. At first, STFE-VI technique extracts spatio temporal features from the collection of video frames in video data set. Next, visual content clustering is performed to cluster the video clips in data set based on extracted features of visual contents i.e. color, texture; shape and motion. Subsequently, STFE-VI technique used VP tree for storing the clustered video clips based on their features. At last, video retrieval process is performed that extracts more similarly videos based on user query which resulting in increased true positive rate of video retrieval. The effectiveness of STFE-VI technique is evaluated in terms of clustering time, clustering accuracy, true positive rate of video retrieval and video retrieval time and compared with two exiting methods. With the experiments carried outs for STFE-VI technique, it is observed that the true positive rate of video retrieval rate affords more precise results when compared to state-of-the-art works. The experimental results show that STFE-VI technique provides better performance with an enhancement of true positive rate of video retrieval rate and reduced the video retrieval time as compared to state-of-the-art works.

REFERENCES

[1] G.G. Lakshmi Priya, S. Domnic, “Shot based keyframe extraction for ecological video indexing and retrieval”, Ecological Informatics, Elsevier, Volume 23, Pages 107–117, September 2014

[2] D.Saravanan, Vaithyasubramanian, K.N. Jothi Vengatesh, “Video Content Reterival Using Historgram Clustering Technique”, Procedia Computer Science, Elsevier, Volume 50, Pages 560 – 565, 2015

[3] Andre Araujo and Bernd Girod, “Large-Scale Video Retrieval Using Image Queries”, IEEE Transactions On Circuits And Systems For Video Technology, Volume PP, Issue 99, Pages 1-14, 2017

[4] Sajad Mohamadzadeh and Hassan Farsi, “Content Based Video Retrieval Based On HDWT and Sparse Representation”, Image Analysis and Strereology, Volume 35, Issue 2, Pages 67-80, 2016

[5] Kaiyang Liao, Guizhong Liu, Li Xiao, Chaoteng Liu, “A sample-based hierarchical adaptive K-means clustering method for large-scale video retrieval”, Knowledge-Based Systems, Elsevier, Volume 49, Pages 123–133, 2013

 ISBN: 978-93-86171-70-2

[6] C.Ranjith Kumar, S. Suguna, "Visual Semantic Based 3D Video Retrieval System Using HDFS", Data Mining Knowledge Discovery, Volume 10, Issue 8, Pages 3806–3825, August 2016

[7] Xu Chen, Alfred O. Hero, and Silvio Savarese, "Multimodal Video Indexing and Retrieval Using Directed Information", IEEE Transactions on Multimedia, Volume 14, Issue 1, Pages 3 – 16, February 2012

[8] Zheng-Jun Zha, Meng Wang, Yan-Tao Zheng, Yi Yang, Richang Hong, Tat-Seng Chua, "Interactive Video Indexing With Statistical Active Learning", IEEE Transactions on Multimedia, Volume 14, Issue 1, Pages 17 – 27, 2012

[9] Lei Wang, Eyad Elyan, Dawei Song, "Rebuilding Visual Vocabulary via Spatial-temporal Context Similarity for Video Retrieval", International Conference on Multimedia Modeling, Springer, Pages 74-85, 2014

[10] Stefanos Vrochidis, Ioannis Kompatsiaris, and Ioannis Patras, "Utilizing Implicit User Feedback to Improve Interactive Video Retrieval", Hindawi Publishing Corporation, Advances in Multimedia, Volume 2011, Article ID 310762, Pages 1-18, 2011

[11] Jun Wu and Marcel Worring, "Efficient Genre-Specific Semantic Video Indexing", IEEE transactions on multimedia, Volume 14, Issue 2, Pages 291 – 302, April 2012

[12] Sujuan Hou and Shangbo Zhou, "Audio-Visual-Based Query by Example Video Retrieval", Hindawi Publishing Corporation, Mathematical Problems in Engineering, Volume 2013, Article ID 972438, Pages 1-8, 2013

[13] Aasif Ansari, Muzammil H Mohammed, "Content based Video Retrieval Systems - Methods, Techniques, Trends and Challenges", International Journal of Computer Applications, Volume 112, Issue 7, Pages 13-22, February 2015

[14] M.Ravinder and T.Venugopal, "Content-Based Video Indexing and Retrieval using Key frames Texture, Edge and Motion Features", International Journal of Current Engineering and Technology, Volume 6, Issue 2, Pages 672-676, April 2016

[15] Mohd. Aasif Ansari, Hemlata Vasishtha, "Enhanced Video Retrieval and Classification of Video Database Using Multiple Frames Based on Texture Information", International Journal of Computer Science and Information Technologies, Volume 6, Issue 2, Pages 1740-1745, 2015

[16] P. M. Kamde, Sankirti Shiravale, S. P. Algur, "Entropy Supported Video Indexing for Content based Video Retrieval", International Journal of Computer Applications, Volume 62, Issue 17, Pages 1-6, January 2013

[17] D.Saravanan, V.Somasundaram, "Matrix Based Sequential Indexing Technique for Video Data Mining", Journal of Theoretical and Applied Information Technology, Volume 67, Issue 3, Pages 725-731, 2014

[18] Muhammad Nabeel Asghar, Fiaz Hussain, Rob Manton, "Video Indexing: A Survey", International Journal of Computer and Information Technology, Volume 3, Issue 1, Pages 148-169, January 2014

[19] Dilek Kucuk, Adnan Yazıc, "Exploiting information extraction techniques for automatic semantic video indexing with an application to Turkish news videos", Knowledge-Based Systems, Elsevier, Volume 24, Pages 844–857, 2011

[20] Amit Fegade, Prof. Vipul Dalal, "Content Based Video Retrieval by Genre Recognition Using Tree Pruning Technique", International Journal of Computer Science and Information Technologies, Volume 5, Issue 4, Pages 5263-5267, 2014