Time Stamp based Sampled Subspace Clustering for Video Key Frame Extraction

Jeyapandi Marimuthu[1] and Vanniappan Balamurugan[2]

[1,2] *Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Abishekapatti,* Tirunelveli, India, 627 012
athishmarimuthu@gmail.com
bala_vm@msuniv.ac.in

Abstract

Video key frames are the non-reduntant contents of video clips that reflect the characteristics of the whole video. Key frame representations are useful in video summarization, video content analysis, video indexing, etc. Among the existing key frame extraction techniques, the performance of the cluster based key frame extraction techniques is found to be better due to its computational simplicity. However, it exhibits poor performance while handling high dimensional data since it is difficult to maintain inter cluster separation and intra cluster similarity due to non-preservation of time order of frames. Also, it fails to preserve the temporal order. To address these issues this paper proposes a time stamp based subspace clustering technique for key frame extraction. The proposed technique is tested on the WEB benchmark dataset and the accuracy is found to be 91%.

Keywords: *Subspace, Time stamp, Linear dependency, Dimensionality reduction.*

DOI: 10.53555/ecb/2022.11.6.135

1. INTRODUCTION

In recent years, the technological advancements and the availability of inexpensive video capturing system resulted in widespread usage of video camera in all domains. A video is a sequence of two dimensional images projected from a dynamic three dimensional scene onto the image plane of a video camera [1]. The video sources include documentaries, movies, animation, surveillance videos, sports, news telecasts, etc. The video sequence comprises of several frames that are captured with a frame rate of at least 25 per second.

These voluminous videos need to be analyzed in order to extract the useful hidden knowledge. The video analysis comprises of several activities viz. scene analysis, shots detection, key frames extraction, abnormal event detection, object tracking, and video summarization. A shot is a continuous video recording that is captured by a single camera where as a scene is a combination of shots that are captured by one or more cameras at a particular location. Key frames are the succinct and semantic representation of a video. These extracts can further be used in various applications such as video summarization, abnormal event detection, video indexing, etc.

Since the voluminous contents of a video make video analysis difficult, key frames are considered for analysis. There are several key frame extraction techniques that are based on the feature contents or the methods applied. Among these techniques, the performance of the cluster based key frames extraction techniques is better due to its computational simplicity. Though several works have been carried out using the cluster based key frame extraction, the accuracy achieved so far is 89% [2]. This limitation is overcome by means of SubSpace Clustering (SSC) [3][4] techniques.

SSC maps high dimensional data onto a low dimensional subspace so as to retain only the relevant common features. The representative frame of the resultant subspaces are extracted as key frames. As the temporal order not preserved in the SSC there is a limitation on accuracy while extracting the key frames. Therefore the inclusion of time order of frames during the SSC will enhance the accuracy of key frame extraction.

The main challenges in the SSC based key frame extraction techniques are as follows:

- Non preservation of time order of frames.
- Poor clustering accuracy.
- High computational complexity due to voluminous and redundant contents.

To overcome the above challenges, this paper proposes a Time Stamp based SSC (TSSSC) technique to improve the clustering accuracy. To reduce the volume and redundancy of the input color video, the frame sequence is sampled at the ratio of 10:1. The RGB color channel of the sampled input video frames are divided into equal-sized blocks and the feature vectors are extracted from the respective color histograms of each block. These feature vectors are represented against its frames as a matrix. The irrelevant and redundant feature vectors are eliminated using the measure of dispersion among the feature vectors. The frame vectors of the dimensionality reduced matrix are grouped by analyzing their dependencies and as a result several clusters are formed. The frames of each cluster are ordered as sequence based on their time stamps and the median frame of each subsequence is extracted as key frames.

The proposed TSSSC technique is tested on WEB benchmark data set and the result shows that the TSSSC technique outperforms the existing SSC based key frame extraction techniques in terms of accuracy.

The major contributions of the proposed work are:

1) The temporal order of the extracted key frames is preserved using their time stamps.
2) Prior knowledge on the number of clusters is not needed.
3) The spatial information is taken into consideration while extracting the features.
4) The similarity based on linear dependency reflects the reality while clustering.

The rest of the paper is organized as follows: Section 2 reviews the important research works which are relevant to the current work, section 3 deals with the proposed TSSSC technique, section 4 narrates the experimentation and results, and finally section 5, summarizes the TSSSC technique based key frame extraction.

2. Related Work

The clustering based key frame extraction applies the similarity analysis on the feature space to cluster the frames and the frame on the cluster centre is selected as key frame. Since the computational complexity of the clustering based technique was high, SSC for key frame extraction was introduced by Agrawal et. al [5] .

Unlike the traditional clustering techniques, SSC seeks to find clusters in various subspaces instead of the whole feature space. Thus the dimensionality of the feature set is reduced by discarding the irrelevant and redundant dimensions.

There are several work available in literature on SSC techniques viz. iterative algorithms [6-8], statistical SSC [9-11] and algebraic SSC [11-13].

In iterative subspace clustering, the feature space is divided into subspaces using dimensionality reduction technique and these subspaces are iteratively refined using optimization techniques. Though the iterative SSC handles the noise and missing values effectively, its convergence to global optimum depends on the selection of initial partitions and also the use of right similarity measure.

Statistical SSC techniques consider the probability distribution of data or noise to partition the feature space into subspaces. Since these techniques are not iterative the computational complexity is improved. However, the number and dimensions of the subspaces need to be known beforehand and the algorithm is not scalable.

The algebraic SSC [18] applies the algebraic properties such as factorization, decomposition, polynomial fitting, and linear dependency to cluster the dataset. The factorization based segmentation techniques [11-13] deal with motion features in which the number of independent motions of subspaces are estimated based on singular value decomposition. The motion and shape of moving objects are recovered using the invariant shape interaction matrix. The maximum likelihood based statistical approach was used to estimate noise. The factorization techniques [15-18] describes the reduction and null space algorithm in which the base vectors of null spaces are derived in order to remove irrelevant dimensions. Vidal et. al [14] elaborated a generalized PCA technique which uses the derivatives of the homogeneous polynomials to extract the base vectors of each subspaces. Algebraic SSC handles data that are lying in linear, affine and dependent subspaces. As these techniques fail to consider the spatial information there is degradation in accuracy.

To improve the accuracy this paper proposes an algebraic approach which applies stratified sampling on the input frames to reduce the dimensionality and considers the time stamp of the frames to cluster them.

3. Key Frame Extraction using TSSSC

This section defines a problem of key frame extraction and proposes a novel TSSSC technique for extracting the key frames of a given input color video. The technique comprises of four major phases viz. feature extraction, dimensionality reduction, clustering and key frame extraction.

Proposed TSSSC Technique

Consider an input color video sequence of N frames with $r \times s$ dimensions and let F be the set of frames sampled at regular interval T, i.e., $F = \{f_i\}$ where i is the frame sequence number which is also considered as time stamp. The value of i ranges from 1 to N / T *and* $F \subset F_N$. Let F_R, F_G, and F_B be the decomposed frame sequences corresponding to red, green, blue channels. Now the problem is to identify those f_i's, which can be identified as key frames where the key frames are distinct frames that do not share common color features.

The variations in similarity distance D between two consecutive frames of a frame sequence determine the key frames. Let F_k be the set of key frames where $F_k \subset F$. Let U be the set of m blocks in a frame with equal size, i.e., $U = \{u_{ij}\}$ where $i = 1..3$ represents each color channel R, G, and B and $j=1..m$, represents the number of blocks. Let H be the set of histograms of N / T frames, where $H = \{h_{ijl}\}$, derived from U by representing each u_{ij} *with* l bins. These values of h_{ijl} are represented as a matrix A with $d \times N / T$ dimensional data space where d is the number of feature vectors represented in rows.

The feature vectors in A, which do not contribute significantly in extracting the key frames leads to poor accuracy. To overcome this, the SSC technique is applied in which a single feature vector of A is chosen initially and the frames are clustered. The cluster quality is determined using the cluster cohesiveness which is determined by estimating the average inter cluster distance δ_1 and intra cluster distance δ_2. Based on the clustering quality the feature vector is either retained or discarded. Once the feature vector is retained, the consecutive feature vector is added along with the retained feature vectors. This process is repeated for the entire matrix. Finally, the data space A is transformed into a subspace B whose dimensions are $r \times N / T$ where r is the resultant feature vectors .

To cluster the frames, the transpose of B *i.e.* B^T is derived and represented in a hyper plane. The frames are grouped into various clusters based on their linear dependency. The frames in B^T are clustered by generating subspace matrices S_x where d denotes the frame clusters to be evolved. The creation of subspace matrix has two phases. In the first phase, the first two frame vectors are included in the S_1 and its rank is estimated. The frames are retained in S_1 in case if the rank is one and the consecutive frame vector is included in S_1 iteratively till the rank is one. Else, in the second phase an additional subspace matrix S_2 is created with the independent row vector as a base vector. The process is repeated by adding the consecutive frame vectors till all the frame feature vectors are clustered.

Based on the time stamps, the frames in each cluster S_k are ordered as one or more subsequences and the median frames are identified as key frames. The incorporation of time stamps improves the accuracy of key frame extraction as it identifies two similar key frames in a cluster to be identified as two different key frames based on their time order.

The major workflow of TSSSC technique is illustrated in Figure 1.

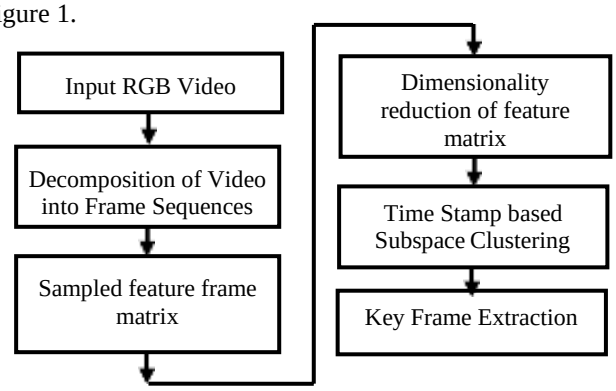

Figure 1. Block diagram of TSSSC technique

TSSSC Algorithm:

The algorithm for this TSSSC technique is illustrated in Algorithm 1.

Algorithm 1 TSSSC

Input: Input Color Video, Interval T, Bins k, blocks U
Output: Extracted Key Frames

1. BEGIN
2. Decompose Video into RGB channels
   /* Stratified Sampling */
3. FOR each input frame f_i
   Frame set F ← $\{f_i\}$ + T
4. ENDFOR
   /* Feature vector for each frame */
5. FOR each frame f_i in set F
   FOR each block U_{ij}
   Compute histogram h_{ijk} with k-bins
   Feature vector for a frame f_i using h_{ijk}
   ENDFOR
   ENDFOR
   /* Feature matrix A for the video */
6. Represent feature frame matrix A
   /* Dimensionality Reduction of A */
   /* BD & WD – between & with-in cluster ditance */
7. Dimensionality_reduction(A, δ_1, δ_2)
   FOR each dimension A[i]
   a. Compute B$D(r, s)$ and WD(r,s)
   b. IF B$D(r, s)$ > $\delta_1$and WD(r,s) < δ_2
   Include dimension A[i]
   ELSE
   Reduce dimension A[i]
   ENDIF
   ENDFOR
   Data space A transformed to B after reduction
   /* Rank based Linear dependency for clustering */
8. Rank_based_Cluster(B^T)
   Base frame ← $B^T[1]$
   FOR each row vector or frame $B'^T[j]$
   WHILE(! assigned)
   IF(Dependent = rank[Base frame;$B^T[j]$]==1)
   $B^T[j]$ is assigned to this cluster
   ELSEIF(Unchecked clusters)
   Base frame ← Base frame of next cluster
   $B^T[j]$ is assigned to different cluster
   ELSE
   Base frame ← $B^T[j]$
   ENDIF
   ENDWHILE
   ENDFOR
   Resultant clusters K
   /* Time stamp based key frame selection */
9. Time_stamp_Key_ frame(Clusters K)
   FOR each cluster C_i
   FOR each time ordered subsequence Q_i
   Key frame ← Median of Q_i
   ENDFOR
   ENDFOR
   END

4. Experimental Results

4.1 Experimental Setup

This section presents the experimentation details on the TSSSC using the WEB benchmark dataset [19]. The data sets contain video sequences comprising of cartoons, TV shows, news and sports events, and commercial activities. The frame rate of videos varies from 24 fps to 30 fps and the play time of videos varies from 5 to 20 seconds. The algorithm was implemented in MATLAB version 2012a.

The sampling rate T is set as 10 since it will be ideal to choose the sample around the median value of the frame rate. The number of equal sized block is set as 9. The value of the histogram bin k is set as 8. The value of δ_1 and δ_2 is set experimentally as 0.2 and 2.5 respectively. The details on the dataset used, total frames, number of extracted key frames and the percentage of dimensionality reduction are furnished in table 1. The key frames extracted using TSSSC technique for the Web dataset are shown in Figure 2. The extracted key frames match with the user summary in the aspect of frames with similar contents. The accuracy of this key frame extraction technique is given in table 1.

The results show that the degree of dimensionality reduction varies depending on the nature of the content of the video sequences. Further, the nature of the content richness has an impact on the extracted key frames. More number of key frames is extracted against sports and cartoons videos as the variation in contents are high.

There is a possibility that the extracted key frames do not represent the whole nature of the video. To measure the accuracy rate of the key frame extraction, the measures viz. true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are computed and presented in the Table 1.

The accuracy rate of key frame extraction is computed using the eq. (1).

$$Accuracy\ Rate = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

It shows that the TSSSC complements the effective clustering using linear dependency based similarity measure and by preserving the temporal order.

4.2 Result Analysis

The comparative analysis of the performance of the TSSSC key frame extraction technique with the existing techniques such as Video SUMMarization (VSUMM) [20], Video Summarization Using Key Frame Extraction (VSUKFE) [21], Row Echelon based Spectral Clustering (RESC) [22], and Structured sparse Subspace with Grouping Effect With-in Cluster (SSGEWC) [23] is given in Figure 3.

The proposed TSSSC technique performs well against the input dataset affects the accuracy rate of the TSSSC technique.

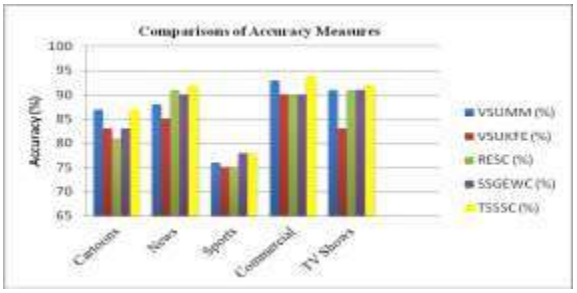

Figure 3. Comparisons of Accuracy Measures

Table 1. Experimental Results

Dataset / Measures	Total Frames	Dimension reduction in %	True Positive	True Negative	False Positive	False Negative	Extracted Key	Accuracy in %
Cartoons	1500	89.8	7	1487	3	3	10	87
News	2200	99.5	4	2194	1	1	5	92
Sports	4000	88.5	22	3966	6	6	28	78
Commercial	3000	99.3	14	2982	2	2	16	94
Tv-shows	2100	99.6	5	2091	2	2	7	92

news, commercial, and tv-shows videos with the accuracy of 92%, 94%, and 93% respectively. The performance measure of this technique outweighs other techniques against news and tv-shows videos even in the presence of redundant events. Further, it is observed that the frequent changes in contents of

a) Key Frames of a Sports Video – v76

b) Key Frames of a News Video – v91

c) Key Frames of a Cartoon Video – v17

Figure 2. Extracted Key Frames for the Web Dataset

Based on the rate of change of visual contents and the types of motion of the video, appropriate key frame extraction technique needs to be applied. It is expected that the content changes need not to be too high or too less as far as SSC algorithms are concerned. The selection of a technique can be decided using factors such as the resultant number of frames, quality of the extracted key frames, and shot reconstruction degree.

5. Conclusions and Future Directions

A time stamp based sampled SSC technique for the extraction of key frames has been introduced in this paper. As this technique considers the time stamps of the frames the clustering accuracy is improved. Further, the similarities between the frames are computed using their linear

dependency rather than using the geometric similarity distance. As the linear dependencies consider the spatial order of features the overall accuracy is improved to 91%. It can further be extended such that the SSC is tolerant to luminance effect and it maintains inter cluster separation for dependent and overlapping subspaces.

References

1. Yao Wang, Jorn Ostermann, and Ya-Qin Zhang, "Video Processing and Communications," Prentice Hall, 1st Edition,2002, ISBN 0-13-017547-1.
2. Gao and Wei, "Improved Ant Colony Clustering Algorithm and Its Performance Study," Computational Intelligence and Neuroscience, pp. 1-14, 2016.
3. Yang, Xu, Zhang, Cao, and Huang, "Split Multiplicative Multi-View Subspace Clustering," IEEE Transactions on Image Processing, Vol. 28, No. 10, pp. 5147-5160, 2019.
4. Vidal, "Subspace Clustering," IEEE Signal Processing Magazine, Vol. 28, Issue: 2, pp. 52-68, 2011.
5. Agrawal, Gehrke, Gunopu, and Raghavan, "Automatic Subspace Clustering of High Dimensional Data," Journal on Data Mining and Knowledge Discovery, Vol. 11, No. 1, pp. 5-33, 2005.
6. Chen, Ye, Xu, and Huang, "A Feature Group Weighting Method for Subspace Clustering of High-Dimensional Data," Pattern Recognition, Vol. 45, Issue:1, pp. 434-446, 2012.
7. Li, Sheng, Kang Li, and Yun Fu, "Temporal Subspace Clustering for Human Motion Segmentation," Proceedings of the IEEE Int. Conference on Computer Vision, Santiago, Chile, 2015.
8. Kang, Lin, Zhu, and Xu, "Structured Graph Learning for Scalable Subspace Clustering: From Single View To Multiview," IEEE Transactions on Cybernetics, Vol. 52, No. 9, pp. 1-11, 2021.
9. Vijendra Singh, and Sahoo Laxman, "Subspace Clustering of High-Dimensional Data: An Evolutionary Approach," Applied Computational Intelligence and Soft Computing, pp. 1-12, 2013.
10. Menon, Vishnu, Gokularam Muthukrishnan, and Sheetal Kalyani, "Subspace Clustering Without Knowing the Number of Clusters: A Parameter Free Approach," IEEE Transactions on Signal Processing, Vol. 68, pp. 5047-5062, 2020.
11. Boult, E. Terrance, and L. Gottesfeld Brown, "Factorization-Based Segmentation of Motions," Proceedings of the IEEE Workshop on Visual Motion, IEEE Computer Society, Princeton, USA, 1991.
12. Gear, C. William, "Multibody Grouping From Motion Images," International Journal of Computer Vision, Vol. 29, Issue: 2, pp. 133-150, 1998.
13. Costeira, Joao Paulo, and Takeo Kanade, "A Multibody Factorization Method for Independently Moving Objects," International Journal of Computer Vision, Vol. 29, Issue: 3, pp. 159-179, 1998.
14. Vidal Rene, Yi Ma, and Shankar Sastry, "Generalized Principal Component Analysis (GPCA)," IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 27, Issue: 12, pp. 1945-1959, 2005.
15. Aldroubi, Akram, and Ali Sekmen, "Reduction and Null Space Algorithms for the Subspace Clustering Problem," arXiv.org, 2010.
16. Aldroubi, Akram, and Ali Sekmen, "Reduced Row Echelon Form and Non-Linear Approximation forSubspace Segmentation and High-Dimensional Data Clustering," Applied and Computational Harmonic Analysis, Vol. 37, Issue: 2, pp. 271-287, 2014.
17. Lu, Gui-Fu, Yong Wang, and Jian Zou, "Low-Rank Matrix Factorization With Adaptive Graph Regularizer," IEEE Transactions on Image Processing, Vol. 25, Issue: 5, pp. 2196-2205, 2016.
18. Aldroubi A, Sekmen A, Koku A. B, and Cakmak A. F, "Similarity Matrix Framework for Data From Union of Subspaces," Applied and Computational Harmonic Analysis, Vol. 45, Issue: 2, pp. 425-435, 2018.
19. WEB Dataset. [Online]. Available: http:/ /homepages. inf. ed. ac.uk/rbf/CVonline/ Imagedbase.html
20. De Avila S.E.F, Lopes A.P.B, da Luz Jr A. and de Albuquerque A, "VSUMM: A Mechanism Designed to Produce Static Video Summaries and A Novel Evaluation Method," Pattern Recognition Letters, Vol. 32, Issue: 1, pp. 56-68, 2011.
21. Ejaz Naveed, Tayyab Bin Tariq, and Sung Wook Baik, "Adaptive Key Frame Extraction for Video Summarization Using an Aggregation Mechanism," Journal of Visual Communication and Image Representation, Vol. 23, Issue: 7, pp. 1031-1040, 2012.
22. Marimuthu Jeyapandi, Vanniappan Balamurugan, and Sankaran Ramesh Kumar, "Row Echelon based Spectral Clustering Framework for Key Frame Extraction," 2021 5th Int. Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, Tamilnadu, India, 2021.
23. Chen Huazhu, Weiwei Wang, and Xiangchu Feng, "Structured Sparse Subspace Clustering With Within-Cluster Grouping," Pattern Recognition, Vol. 83, pp. 107-118, 2018.