Key Frame Extraction using Uniform Local Binary Pattern

B. Reddy Mounika
Dept. of Electronics and Communication Engineering
University of Allahabad
Allahabad, India
reddymounika.b@gmail.com

Om Prakash*
Computer Engineering Department,
Institue of Technology
Nirma University
Ahmedabad, India
au.omprakash@gmail.com

Ashish Khare
Dept. of Electronics and Communication Engineering
University of Allahabad
Allahabad, India
ashishkhare@hotmail.com

***Abstract*— Key frames are useful in a wide variety of applications like summarizing, storing, indexing and retrieving video clips etc. In this paper, we present an efficient key frame extraction technique. The present method detects key frames efficiently by extracting Uniform Local Binary Pattern of frames of the video. The distance between uniform local binary pattern of consecutive frames is calculated and compared with threshold to extract key frames. We tested the proposed method on categories of serials and news videos. Performance evaluation is carried out using the parameters- figure of merit, detection percentage, accuracy and missing factor. The proposed method offers greater detection percentage. The furnished comparative results analysis proven the efficiency of the proposed method over the state-of-art-methods.**

Keywords*— *key frame, local binary pattern, figure of merit, detection percentage, accuracy and missing factor

I. INTRODUCTION

The latest technology left a huge amount of multimedia data on the internet, personal and professional databases. Processing such a huge amount of multimedia data is a complex task and a big barrier to the real-time applications. Key frame extraction attracted researchers by representing a video clip in a concise manner and with the advantages of less storage space requirement, time and computational complexities. Key frame extraction plays a significant role in many applications such as video management/retrieval [1] and surveillance [2] etc. A frame which can represent the whole content present in a particular video clip is known as key frame. In general key frames can be extracted in sequential manner [3] or by clustering [4]. Although key frames obtained by these methods efficiently represent the video clip, these approaches have certain limitations such as need of proper threshold selection methods [5] and efficient shot detection algorithms [6]. So, efficient key frame extraction is still an open issue.

In this paper, we introduce an efficient framework towards key frame extraction with the following properties:

- Proposed method doesn't need any prior knowledge about the video content and there is no learning phase.
- Proposed method is so simple, fast, less complex and more accurate.

The proposed algorithm uses Uniform Local Binary Pattern (uniform LBP) as a feature. Difference of Uniform LBP between the current and next frames of video is calculated and if this value is greater than the threshold then current frame is selected as key frame. . We tested the proposed method on categories of serials and news videos. The results are compared with the state-of-the-art methods and with this analysis we observed that the proposed method shown sound performance over other methods.

Remainder of this paper is organized as follows: In Section 2. we address literature review; Section 3 outlines briefly about Uniform Local Binary Pattern (Uniform LBP) and the proposed method. Section.4 presents experimental analysis and conclusion is give Section. 5.

II. LITERATURE REVIEW

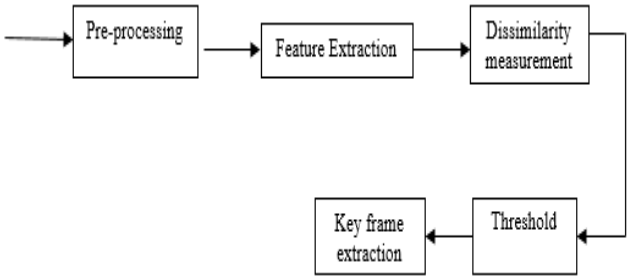
Video browsing has became a routine part in day to day life of human being. Key frame extraction plays a critical role in browsing video clips. Dang et al. [7] proposed a novel key frame extraction based on robust principal component analysis in which the input data is decomposed into low rank components and set of sparse components. An adaptive key frame selection method for summarizing video clips have been introduced by Chakraborty et al. [8]. Poonam et al. [9] used higher order color moments of block based histogram for shot dection and then from each shot the frames of highest mean and standard deviation have been taken as key frames. Sheena et al. [10] compared variation between histogram of consecutive video frames with a threshold to extract key frames. Hannane et al. [11] detected shot transitions by utilizing SIFT point distribution histogram and then extracted key frames using singular value decomposition. A wavelet transformed blocks have been used by Thakre et al. [12] in establishing shots and then chosen the frames of local minima and maxima of each shot as key frames. Guoliang Lu et al. [13] represented each frame as a two tuple feature which is the combination of color quantity information, occupancy histogram. Abnormality is calculated as the difference of actual and estimated two tuple features at time t. From these

*Corresponding Author

abnormalities key frames are detected by employing a Randomized Power Martingale (RPM) statistical test.

III. Proposed Method

We briefly discussed the proposed method in this Section. The concept of proposed method lies on the fact that the difference between features of consecutive frames will be less for redundant frames. The proposed method consists of five stages: pre-processing, feature extraction, dissimilarity measurement, threshold computation and extraction of key frames.



Pre-processing- In pre-processing frames of the video are extracted and then the frames are converted to gray level and given as input to the feature extraction stage.

Feature Extraction- For each frame uniform local binary pattern (Ulbp) is extracted using equation (2).

Dissimilarity measurement- The difference between uniform local binary patterns of current and next frames belonging to the video clip is calculated and stored into an array named dissimilarity.

Threshold computation- The mean of the dissimilarity array is calculated and it is selected as threshold.

Key frame extraction- The dissimilarity array is now scanned to get the entries greater than the threshold value. The current frame at which dissimilarity greater than the threshold is selected as key frame.

A. Uniform Local Binary Pattern

Uniform Local Binary Pattern (uniform LBP) is a Local Binary Pattern with a maximum of two circular 0-1 or 1-0 transitions. Uniform Local Binary Pattern is an extra ordinary operator and it has proven its efficiency in variety of applications [14]. For example, patterns 11111111, 11011111, 00111000 and 00000000 are uniform patterns whereas 01010000, 01001110 or 10101100 are not uniform. With the use of Uniform Local Binary Pattern we can reduce length of feature vector and we can improve performance. The Uniform Local Binary Pattern has the advantage of invariance to different illumination conditions; it is more accurate, sparse and simple in calculation. Uniform LBP is given by

$$LBP^{riu2}{}_{P,R} = \begin{cases} \sum_{P=0}^{P-1} S(g_P - g_c) & if \quad u(LBP_{P,R}) \geq 0 \\ P+1 & otherwise \end{cases} \quad (1)$$

Where

$$u(LBP_{P,R}) = \left|S(g_P - g_c) - S(g_O - g_c)\right| + \sum_{P=1}^{P-1}\left|S(g_P - g_c) - S(g_{P-1} - g_c)\right|$$

g_c - gray level value of a center pixel in a local neighborhood

g_P - gray level values of P pixels lying on a circle at equi-distance forming circular symmetric neighbor set.

B. The Proposed Algorithm

The proposed algorithm of key frame selection:

Input: video clip

1. Extract the frames of video.
2. Convert each frame of video to gray level.
3. Then Uniform LBP of each frame is determined till the end of the video using equation (1) and are stored in a matrix named ulbp1.
4. The difference between each and every consecutive entries of ulbp1 is calculated using equation (2) and is stored into an array named dissimilarity.

$$D(i,i+1) = \sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n}\left|Ulbp_i(j,k) - Ulbp_{i+1}(j,k)\right|^2} \quad (2)$$

5. The mean of the dissimilarity array is calculated and is chosen as threshold to extract key frames.

   T = mean (D)

6. The dissimilarity array is now scanned to get the entries greater than the threshold value. The current frame at which dissimilarity value greater than threshold is chosen as key frame.
7. Now from the shots obtained we select the final frame of the shot as key frame.

Output: Key frames

IV. Experimental Results and Analysis

We conducted experiments on videos of serials and news category containing 1,60,000 frames. To evaluate the performance of the proposed technique figure of merit, detection percentage, accuracy, missing factor were calculated. The obtained results are compared with the other state-of-the-art methods Poonnam et al. [9], Sheena et al. [10], Rachida Hannane et al. [11], Thakre et al. [12], Gualiang et al. [13] and we furnished these comparative results in Table. I and Table. II. A tool named Virtual Dub [15] which is available freely was used to built ground truth. A detailed discussion of performance measures is below-

A. Recall, Precision, Figure of Merit

According to Debabrata Dutta et al. [16] Figure of merit (F) is calculated as

$$F = \frac{2*(P)*(R)}{R+P} \quad (3)$$

Where precision P, Recall R is given as

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

Where the true positives (TP) is the number of key frames correctly detected by the proposed method. The false negative (FN) is the number of key frames in ground truth for which doesn't present in result generated by proposed approach, FP is the number of key frames detected by the proposed approach, which does not present in ground truth. The range of figure of merit (F) is 0 to 1. A good method offers higher values of figure of merit. The results of the proposed method and the methods [9-13] are shown in Table I and II. From Table I and II we can say that proposed method's figure of merit is superior over the other methods [9-13].

TABLE I. PERFORMANCE ANALYSIS OF THE PROPOSED AND THE OTHER METHODS [9-13] FOR SERIALS CATEGORY EACH VIDEO OF LENGTH 20,000 FRAMES

Video	Method	Shots detected	Key frames detected	Key frames in ground truth	F	DP	Accuracy	mf
1	Poonam et al. [9]	1836	1837	2895	0.1328	11.57	0.3113	9.3763
	Sheena et al.[10]	1308	1390		0.1416	9.64	0.264	7.6418
	SIFT- PDH [11]	897	897		0.3472	11.5717	0.3113	7.6418
	Thakre et al.[12]	3920	7840		0.145	17.0639	0.4021	4.8603
	Un super-vised[13]	4222	4222		0.344	11.5371	0.3957	7.6677
	Proposed	**7334**	**7334**		**0.6870**	**22.9016**	**0.4517**	**3.3665**
2	Poonam et al. [9]	1183	1184	3066	0.1206	8.68	0.2414	12.566
	Sheena et al.[10]	681	782		0.1252	7.37	0.1712	10.5263
	SIFT- PDH [11]	612	612		0.2891	9.6373	0.2641	9.3763
	Thakre et al. [12]	4147	8294		0.1397	16.4384	0.4005	5.0833
	Un super-vised[13]	4350	4350		0.3513	11.7091	0.3938	7.5404
	Proposed	**8711**	**8711**		**0.7798**	**25.9948**	**0.4610**	**2.8469**
3	Poonam et al. [9]	1053	1054	2670	0.1088	8.09	0.2424	14.1704
	Sheena et al.[10]	565	890		0.116	6.59	0.1638	11.3611
	SIFT- PDH [11]	612	612		0.3461	11.5371	0.3957	7.6677
	Thakre et al. [12]	3593	7186		0.1389	16.2921	0.3998	5.1379
	Un super-vised[13]	3740	3740		0.3517	11.7228	0.3927	7.5304
	Proposed	**8338**	**8338**		**0.8157**	**27.191**	**0.4660**	**2.6777**
4	Poonam et al. [9]	1905	1906	3226	0.1327	10.85	0.3007	11.8015
	Sheena et al.[10]	573	774		0.1364	7.81	0.1481	8.2171
	SIFT- PDH [11]	433	433		0.6787	22.6244	0.45	3.42
	Thakre et al. [12]	3490	6980		0.1498	15.5921	0.3793	5.4135
	Un super-vised[13]	3976	3976		0.3106	10.3534	0.3783	8.6587
	Proposed	**8294**	**8294**		**0.7142**	**23.8066**	**0.4538**	**3.2005**

TABLE II. PERFORMANCE ANALYSIS OF THE PROPOSED AND THE OTHER METHODS [9-13] FOR NEWSS CATEGORY EACH VIDEO OF LENGTH 20,000 FRAMES

Video	Method	Shots detected	Key frames detected	Key frames in ground truth	F	DP	Accuracy	mf
1	Poonam et al. [9]	2837	2838	1965	0.2031	26.14	0.4467	5.7692
	Sheena et al.[10]	796	967		0.1742	14.77	0.2816	2.8261
	SIFT- PDH [1]	163	163		0.7969	26.5625	0.4152	2.7647
	Thakre et al. [12]	833	1666		0.1323	14.7727	0.3888	5.7692
	Un super-vised[13]	11069	11069		0.983	9.4318	0.5226	5.7
	Proposed	**459**	**459**		**0.9205**	**30.6818**	**0.4655**	**2.2593**
	Poonam et al. [9]	2837	2838	1842	0.158	20.20	0.4189	8.0294
	Sheena et al.[10]	163	796		0.159	11.07	0.2539	3.9516
	SIFT- PDH [11]	163	163		0.5928	19.7605	0.4413	4.0606
	Thakre et al. [12]	833	1666		0.1092	7.9262	0.259	11.6164

2	Un super-vised[13]	11069	11069		0.5603	5.20087	0.5017	9.228
	Proposed	**5278**	**5278**		**0.9805**	**32.6819**	**0.4715**	**2.0598**
3	Poonam et al. [9]	4249	4250	2688	0.1317	18.79	0.4176	9.752
	Sheena et al.[10]	261	1192		0.1456	9.3	0.2517	4.3228
	SIFT- PDH [11]	261	261		0.6364	21.2121	0.4441	3.7143
	Thakre et al. [12]	3147	6294		0.1364	14.8065	0.3844	5.7538
	Un super-vised[13]	2783	2783		0.2589	8.631	0.3573	10.5862
	Proposed	**4331**	**4331**		**0.5045**	**16.8155**	**0.4146**	**4.9649**
4	Poonam et al. [9]	4091	4092	2963	0.1878	20.62	0.4123	6.5586
	Sheena et al.[10]	332	1296		0.1732	13.23	0.2632	3.8494
	SIFT- PDH [11]	332	332		0.1818	3.03	0.5815	3.75
	Thakre et al. [12]	3297	6594		0.1815	19.1698	0.3915	4.2165
	Un super-vised[13]	2973	2973		0.2612	8.7074	0.3543	10.4845
	Proposed	**5334**	**5334**		**0.6490**	**21.6335**	**0.4319**	**3.6225**

Poornima and Kanchana [17] outlined computational equation of detection percentage (DP) as follows:

$$DP = 100 \times \frac{TP}{TP + FN} \tag{6}$$

An efficient method will high value of detection percentage and it lies in the range 1 to 100. From Table I and II we can say detection percentage of the proposed method is superior to the methods [9-13].

B. Accuracy

Accuracy in another known as percentage of corrected classification. Good segmentation will have higher values accuracy. Khare et al. [18] outlined computational equation of accuracy as:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{7}$$

From the results shown table.1 we can say that the values of accuracy for the proposed are superior to the values of the methods outlined above [9-13].

C. Missing Factor

Ibrahim et al. [19] defined missing factor (mf) as

$$mf = \frac{FN}{TP} \tag{8}$$

The lower the value of missing factor the better the performance of the method. From the results shown in Table I and II the value of missing factor for the proposed method is less compared with the values of the methods outlined above [9-13].

V. CONCLUSIONS

In present article, we proposed a novel method of key frame extraction based on Uniform Local Binary Pattern (uniform LBP). The better performance of the proposed over the other methods [9-13] is due to illumination invariance and simplicity of uniform LBP. Uniform LBP were chosen as features to extract key frames. The proposed method is experimented on different videos of serials and news category containing a total frames of 1, 60,000 in which each video is of length 20,000 frames. Performance comparison is done with the state-of-art methods [9-13] and from which we came to a conclusion that the proposed method given promising detection percentage with less missing factor.

REFERENCES

[1] J.Song, L .Gao, L .Liu, X .Zhu, N .Sebe, "Quantization-based hashing: a general framework for scalable image and video retrieval," Pattern Recognition, Vol.75, pp.175-187, 2018.

[2] P.L. Venetianer, A.J. Lipton, AJ .Chosak, MF. Frazier, N. Haering, GW. Myers, W. Yin, Z. Zhang, R. Cutting, inventors; Avigilon Fortress Corp, assignee. Video surveillance system employing video primitives. United States patent US 9,892,606. 2018 Feb 13.

[3] G. Guan, Z. Wang, S. Lu, J. Da Deng, DD. Feng, "Keypoint-based keyframe selection," IEEE Transactions on circuits and systems for video technology, Vol. 23, issue. 4, pp.729-34, 2013 Apr.

[4] N. R. VáZquez-Martí, A. Bandera, "Spatio-temporal feature-based keyframe detection from video shots using spectral clustering," Pattern Recognition Letters, Vol. 34, issue. 7, pp. 770-779.

[5] J.Choi, C. Kim, "A framework for automatic static and dynamic video thumbnail extraction," Multimedia Tools and Applications, Vol. 75, issue. 23, pp. 15975-91, 2016 Dec.

[6] J. Majumdar, K. S. Kumar, G. M. Venkatesh, "Analysis of video shot detection using color layout descriptor and video summarization based on expectation-maximization clustering," In Cognitive Computing and Information Processing (CCIP), pp. 1-5, International Conference on 2015 Mar 3. IEEE.

[7] C. Dang, H. Radha, "RPCA-KFE: key frame extraction for video using robust principal component analysis," IEEE Transactions on Image Processing, Vol.24, issue.11, pp. 3742-3753, 2015 Nov.

[8] S. Chakraborty, O. Tickoo, R. Iyer, "Adaptive keyframe selection for video summarization," In Applications of Computer Vision (WACV), pp. 702-709, 2015 IEEE Winter Conference on 2015 Jan 5. IEEE.

[9] M.P. Jadhav, D.S. Jadhav, "Video Summarization Using Higher Order Color Moments (VSUHCM)," Procedia Computer Science, Vol. 45, pp. 275-281, 2015 Jan.

[10] C.V. Sheena, N. K. Narayana, "Key-frame extraction by analysis of histograms of video frames using statistical methods," Procedia Computer Science, Vol. 70, pp. 36-40, 2015 Jan.

[11] R. Hannane, A. Elboushaki, K. Afdel, P. Naghabhushan, M. Javed, "An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram," International Journal of Multimedia Information Retrieval, Vol.5, issue.2, pp. 89-104, 2016 Jun.

[12] K.S.Thakre, A.M.Rajurkar, R.R.Manthalkar, "Video Partitioning and Secured Keyframe Extraction of MPEG Video", In proc. Of

International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA, Procedia Computer Science, Vol. 45, pp. 275 – 281, 2015

[13] G. Lu, Y. Zhou, X. Li, P. Yan, "Unsupervised, efficient and scalable key-frame selection for automatic summarization of surveillance videos," Multimedia Tools and Applications, Vol. 76, issue. 5, pp. 6309-6331.

[14] S. Nigam, A. Khare, "Integration of moment invariants and uniform local binary patterns for human activity recognition in video sequences," Multimedia Tools and Applications. Vol. 75, issue. 24, pp. 17303-17332, 2016 Dec.

[15] Lee, VirtualDub home page, http://www.virtualdub.org/index.html

[16] K. Poornima, R. Kanchana, "A Method to Align Images Using Image Segmentation," International Journal of Soft Computing and Engineering, Vol. 2, Issue. 1, pp. 294-298, 2012.

[17] Manish Khare, Rajneesh Kumar Srivastasava, Ashish Khare: 'Moving Object Segmentation in Daubechies Complex Wavelet domain', Journal of Signal, Image and Video Processing, Vol. 9, Issue 3, pp. 635-650, 2015.

[18] Ibrahim F. Shaker, Amr Abd-Elrahman, Ahmed K. Abdel-Gawad, Mohamed A. Sherief, "Building Extraction from High Resolution Space Images in High Density Residential Areas in the great Cairo Region", Remote Sensing, Vol. 3, pp. 781-791, 2011.