7th International Conference on Communication, Computing and Virtualization 2016

Novel Video Content Summarization using Thepade’s Sorted n-ary Block Truncation coding

Shalakha R. Badre, Dr. Sudeep D. Thepade

*Student, Department of information technology, Pimpri chinchwad college of engineering, Pune, 411044, India
bProfessor, Department of information technology, Pimpri chinchwad college of engineering, Pune, 411044, India

Abstract

With the quick growth of multimedia technology, a huge amount of videos is available across the world. Length of these videos may be large; user may require only viewing the summary of video. In such cases, video content summarization is used. A video comprise of no. of frames. Video content summarization gives the concise form of the video. Key frame represent the main content of video. For video content summarization key frame extraction is mainly considered. This paper proposes novel method to extract key frames from video using Thepade’s sorted n-ary block truncation coding. Here total five variations of TSBTC’s has done. Out of these, Thepade’s sorted pentnary (TSPBTC), performs better in all similarity measures. Canberra distance give’s effective performance in all TSBTC’s & L1 distance family delivers highest performance as compared to other family.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Organizing Committee of ICCCV 2016

Keywords: Video content summarization; Key frame; similarity measure; TSNBTC.

1. Introduction

The extensive appearance of video cameras together with the lot of social networking and video sharing website has resulted in the increase in the no of videos in the world which has never seen before. These video’s are not in proper format and well structured. Length of these video’s are not properly defined, they can be of any length. Video summarization helps to summarize such type of video. For video summarization, key frame extraction is done. Key frame is frame which is having major difference as compared to previous frame in the series of frames. Entire video can be converted into small no of frames which is having major content of video. Video summarization can have application such as video skimming in which important audio and video are extracted which create skim

* Corresponding author.
E-mail address: badre.shalakha@gmail.com
video which represent the concise form of video [5]. Like that video summarization is used in video indexing and filmmaking. Key frame extraction helps out to find the most important frames from video. While extracting key frame, if we found that the two consecutive frames are having major difference then that frames are termed as key frame [6]. If we are working with whole video content then it will take large processing time but if key frame extraction is considered it reduces maximum time. Until now, large amount of work has been done in video processing such as video indexing [5], shot boundary detection, video classification and content based video retrieval [4].

1.1. Block Truncation Coding (BTC)

Block truncation coding was developed in 1979 and initially used for gray-scale images. Block truncation coding is widely used for video content summarization. Initially, block truncation coding was used for only gray scale images [1]. In block truncation coding, image is divided into no of blocks. Individual formulation is done on each block. Block truncation coding proves to be better in color feature extraction from video and threshold is considered here for formulation [10]. Here we get two values for each color component such as two values for red, green and blue. Less computational complexity is there if we are working with BTC. With the help of block truncation coding features are extracted from image. These features are properties of image. From that feature key frame can be extracted from video.

1.2. Thepade’s Sorted Ternary Block Truncation Coding

In video, there are set of intensity values that are related to respective color component of the frame. In Thepade’s sorted ternary block truncation coding intensity values of frames are extracted from the frame with respect to that color component.

For ex. if we are considering RGB color space, then intensity values of R, G and B are extracted from the frame. Then these values are arranged in one feature vector with respect to color. Then these feature vectors are sorted in ascending order and divided into three parts. Average of each part is taken. That means for each color component, we are getting three values and they are such as lower, middle and upper red and similarly for green and blue component. The nine values that can be obtained from this are the feature vector of the respective frame. After that these feature vectors are considered for key frame extraction purpose. Various similarity measures can be applied on these feature vector to determine the key frames from video. These key frame comprise summarize form of video.

In the proposed methodology, five variations of TSNBTC has tried and feature vector that we are getting from this on that various similarity measure has been applied on that to extract key frames from video.

2. Thepade’s sorted N-ary block truncation coding

In Thepade’s sorted n-ary block truncation coding, there are five variation that are introduced in this paper. These are such as:

- Thepade’s sorted quaternary block truncation coding. (TSQBTC)
- Thepade’s sorted pentenary block truncation coding. (TSPBTC)
- Thepade’s sorted hexenary block truncation coding. (TSHBTC)
- Thepade’s sorted septenary block truncation coding. (TSSBTC)
- Thepade’s sorted octanary block truncation coding. (TSOBTC)

Let us consider, Thepade’s sorted quaternary block truncation coding. In this, intensity values of color component are extracted from each frame of video. Then these intensity values are arranged in one feature vector according to the color component. These feature vectors are sorted in ascending order. Then these feature vectors are divided into four parts. That means for each color component we can get four values. Total intensity of R component of m×n can be presented in form of a single dimensional array ‘SDR’ having elements with indices 1 to m×n [1]. Red component four values can be formulated as follows which are shown in equations 1,2,3 &4.
\[ lR = \left( \frac{4}{m \times n} \right) \times \sum_{i=1}^{\frac{m \times n}{4}} \text{sortedSDR}(i) \]  
(1)

\[ muR = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{4}+1}^{\frac{m \times n}{3}} \text{sortedSDR}(i) \]  
(2)

\[ mlR = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{2}+1}^{\frac{m \times n}{3}} \text{sortedSDR}(i) \]  
(3)

\[ uR = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{3}+1}^{\frac{m \times n}{4}} \text{sortedSDR}(i) \]  
(4)

\[ lG = \left( \frac{4}{m \times n} \right) \times \sum_{i=1}^{\frac{m \times n}{4}} \text{sortedSDG}(i) \]  
(5)

\[ muG = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{4}+1}^{\frac{m \times n}{3}} \text{sortedSDG}(i) \]  
(6)

\[ mlG = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{2}+1}^{\frac{m \times n}{3}} \text{sortedSDG}(i) \]  
(7)

\[ uG = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{3}+1}^{\frac{m \times n}{4}} \text{sortedSDG}(i) \]  
(8)

\[ lB = \left( \frac{4}{m \times n} \right) \times \sum_{i=1}^{\frac{m \times n}{4}} \text{sortedSDB}(i) \]  
(9)

\[ muB = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{4}+1}^{\frac{m \times n}{3}} \text{sortedSDB}(i) \]  
(10)

\[ mlB = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{2}+1}^{\frac{m \times n}{3}} \text{sortedSDB}(i) \]  
(11)

\[ uB = \left( \frac{4}{m \times n} \right) \times \sum_{i=\frac{(m \times n)}{3}+1}^{\frac{m \times n}{4}} \text{sortedSDB}(i) \]  
(12)
The values that we are getting after solving this equations are feature vector of frame. Therefore feature vector will be as \([IR, muR, mlR, uR, IG, muG, mlG, uG, IB, muB, mlB \text{ and } uB]\). The equations which can be formed for TSQBTC, same equations can be made for TSPBTC, TSHBTC, TSSBTC, TSOBTC. For TSPBTC we can get five values for each color component. Like that for TSHBTC, TSSBTC and TSOBTC six, seven, eight values can be generated for each color component respectively. In this paper, similarity measure is used for key frame extraction.

3. Proposed video content summarization method

Fig. 1. Block diagram of proposed algorithm

Here, the video is taken as input from any source such as social networking site. The video can be of ‘N’ no of frame. Then the frames are extracted from video. Then on each frame TSQBTC, TSPBTC, TSHBTC, TSSBTC and TSOBTC has to apply to extract feature from frame. Then for every adjacent frame feature vector similarity measure can be used to calculate the difference(diff) between two frames. The similarity measures are Square-chord distance, Fidelity Distance, Mean Square Error, Chebyshev Distance, Sorensen Distance, Canberra Distance, Wavehedge Distance, Intersection Distance. With the help of eight similarity measures mean is calculated of all frames. After that standard deviation and threshold is calculated of frames. Threshold can be calculated by addition of mean and standard deviation [1]. Followed by this, similarity is calculated between frames by comparing threshold with difference between two frames. The mean, standard deviation can be calculated as follows as shown in equation 13, 14 and 15.

\[
Mean(M) = \frac{\sum_{i=1}^{N} \text{diff} (i)}{N-1} \quad (13)
\]

\[
S\text{tandardDeviat}(s) = \sqrt{\frac{\sum_{i=1}^{N-1} (\text{diff} (i) - M)^2}{N-1}} \quad (14)
\]

\[
Threshold(T) = M + a \times S \quad (15)
\]

Here ‘a’ is constant
If we want to get correct key frames
If (diff (i) > threshold)
Output of nth frame set i+1 set as Key frame [3]

4. Different similarity measures used

For each consecutive frame feature vector TSBTC is applied. On that feature vector similarity measure is
applied to calculate the difference between two frames. Similarity measures are as follows:

4.1. Square-chord Distance

\[ SC = \sum_{i=n}^{i=n} (\sqrt{R_i} - \sqrt{S_i})^2 \]  \hspace{1cm} (16)

Where R_i and S_i are consecutive frames. Here square of difference between frames has done to calculate
similarity between two frames.

4.2. Fidelity Distance

\[ FD = \sum_{i=n}^{i=n} \sqrt{R_i S_i} \]  \hspace{1cm} (17)

Square root of product of R_i and S_i is taken to calculate the distance between source image and destination
image.

4.3. Mean Square-error

\[ MSE = \frac{1}{N} \sum_{i=1}^{i=n} (R_i - S_i) \]  \hspace{1cm} (18)

Here there is fraction of difference between reference image & target is taken and no of frames. The equation
is given as above in equation 18.

4.4. Chebyshev Distance

\[ CHD = \max_i |R_i - S_i| \]  \hspace{1cm} (19)

Chebyshev distance is given as above in equation no 19. This is also called as maximum metric which was
invented by Pafnuty Chebyshev.

4.5. Sorencen Distance

\[ SD = \sum_{i=1}^{i=n} \frac{|R_i - S_i|}{|R_i + S_i|} \]  \hspace{1cm} (20)
It is used to calculate similarity between two samples.

### 4.6. Canberra Distance

\[
CD = \sum_{i=1}^{n} \left| \frac{R_i - S_i}{R_i + S_i} \right|
\]  

(21)

Canberra distance was invented by Lance and Williams in 1967. The equation is given as above.[1]

### 4.7. Wahehedge Distance

\[
WD = \sum_{i=1}^{n} \frac{|R_i - S_i|}{\max(R_i, S_i)}
\]  

(22)

Equation 22 returns the wavehedge distance between \(R_i\) and \(S_i\).

### 4.8. Intersection Distance

\[
ID = \sum_{i=1}^{n} \min(R_i, S_i)
\]  

(23)

Equation 23 returns the intersection distance between two frames. Here, it takes the minimum between two frames.

### 5. Experimentation environment

#### 5.1. Test bed used

In implementation test bed of 25 videos are used. These videos dataset are shown in figure 2. These videos are taken from standard dataset that is VSUMM. In this dataset, they have provided key frames from the video from five different users.

#### 5.2. Platform used

For this proposed work MATLAB 2012a is used. The basic system is of Intel core 2 duo(2.93GHz) with 4GB RAM.

#### 5.3. Performance Comparison

For performance comparison, the average accuracy is used.

\[
\text{Percentage Accuracy} = \frac{\text{Actual Correct Extracted Frames}}{\text{Total Expected Extraction Of Frames}}
\]  

(24)

Actual correct extracted frames means key frames extracted from video using proposed algorithm. Total expected extraction means key frames extracted manually [1].
6. Result and discussion

Figure 3 shows that percentage accuracy of various TSBTC with sorensen similarity measure. In this Thepade’s sorted pencylary block truncation coding is giving highest performance for sorensen similarity measure that is 79.69%.
Here in this figure various similarity measures are compared over TSBTC. In this we can see that TSPBTC is performing best in all similarity measures followed by TSHBTC.

Figure 5 shows that Canberra distance is giving better performance in all TSBTC. It is giving 83.65 for TSPBTC followed by TSHBTC and TSQBTC. Here in this proposed model, similarity measure are taken from specific family. These families are Minkiowski, L1, Intersection, Fidelity. Table 1 shows that percentage accuracy according to different family over various TSBTC. In this L1 family give highest performance which include Canberra distance and Sorensen distance. After that Minkiowski family is performing good which include MSE and Chebyshev.

<table>
<thead>
<tr>
<th></th>
<th>TSQBTC</th>
<th>TSPBTC</th>
<th>TSHBTC</th>
<th>TSSBTC</th>
<th>TSOBTC</th>
<th>Average of TSBTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minkiowski</td>
<td>69.58</td>
<td>73.26</td>
<td>69.07</td>
<td>70.42</td>
<td>71.1</td>
<td>70.68</td>
</tr>
<tr>
<td>L1</td>
<td>76.94</td>
<td>81.67</td>
<td>77.34</td>
<td>74.61</td>
<td>74.04</td>
<td>76.92</td>
</tr>
<tr>
<td>Intersection</td>
<td>63.74</td>
<td>69.64</td>
<td>66.82</td>
<td>64.62</td>
<td>62.85</td>
<td>65.53</td>
</tr>
<tr>
<td>Fidelity</td>
<td>62.8</td>
<td>69.24</td>
<td>67.49</td>
<td>66.2</td>
<td>63.65</td>
<td>65.87</td>
</tr>
<tr>
<td>Average of family</td>
<td>62.26</td>
<td>73.45</td>
<td>70.18</td>
<td>68.96</td>
<td>67.91</td>
<td></td>
</tr>
</tbody>
</table>
7. Conclusion

Key frame extraction plays a major role in video summarization. With help of key frames, processing of long video’s become very easy and it saves a lot of extra work. In the proposed work Thepade’s sorted n-ary block truncation coding is introduced. In which, five variation of TSBTC are tried such as TSQBTC, TSPBTC, TSHBTC, TSSBTC and TSOTBC. This is performed with eight different similarity measure from different family. Out of these, Thepade’s sorted pentnary block truncation coding gives better performance followed by Thepade’s sorted hexnary block truncation coding. Canberra distance performs better as compared to other similarity measure which is followed by Sorensen and wavehedge. Out of all these, L1 family gives highest performance.

References

1. Dr. Sudeep D. Thepade, Pritam H. Patil “Novel visual content summarization in videos using keyframe extraction with Thepade’s sorted ternary block truncation coding and assorted similarity measures” International conference on communication, information & computing technology(ICCICT), Jan 16-17 2015, India.
5. Huayong Liu, Huiwen Hao “Keyframe extraction based on improved hierarchical clustering algorithm” International conference on fuzzy system and knowledge discovery, 2014.