International Journal of Electronics and Communication Engineering (IJECE)
ISSN(P): 2278-9901; ISSN(E): 2278-991X
Vol. 2, Issue 5, Nov 2013, 225-230

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

VIDEO SUMMARISATION USING CONTOURLET TRANSFORM AND ECLUDIAN DISTANCE

PAMARTHY CHENNA RAO[1] & M. RAMESH PATNAIK[2]

[1]Research Scholar, Department of Instrument Technology, Andhra University, Visakhapatnam, Andhra Pradesh, India

[2]Assistant Professor, Department of Instrument Technology, A. U. College of Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India

ABSTRACT

Video summarization is representing the full video in terms of key frames and is the main key process in the video content management system to summarize the video and content searchers can easily search particular scene from given video using this technique and also users can easily find summary of the given video without seeing full video. Key frames provide a most desirable abstraction for video indexing, browsing, and retrieval. Key frame extraction process is to represent the full video in terms of short video clips and gives information about that video. In this work, video summarization in terms of key frames are extracted using feature vectors, which are obtained from features calculated for each sub-band in the contour let transform. Where the energy and standard deviation features of each sub - band are used to form a feature vector. The experimental results proved that this novel method had more accuracy rate and low error rate.

KEYWORDS: Contourlet, Key Frames, Energy, Euclidean Distance, Accuracy Rate, Error Rate

INTRODUCTION

Due to rapid increase in video data on the Internet, demands efficient techniques for management and storage of the video data and also users wasting lot of their valuable time by watching and browsing videos, even though they are not interested in the entire movie to know summary of the video. This situation demands a video abstraction and summarization techniques that produces video summaries highlighting only the relevant contents of the video while preserving the continuity of the video. The summarized video helps the user to evaluate whether it is interesting or not [1]. Video summarization finds wide applications in security, entertainment and military areas [2]. Video summarization techniques are considered within three broad categories: (i) internal (analyze information sourced directly from the video stream), (ii) external (analyze information not sourced directly from the video stream) and (iii) hybrid (analyze a combination of internal and external information) [3]. Shot boundary detection and key frame extraction interchangeable used in this paper.

A shot is defined as an unbroken sequence of frames recorded from a single camera [4]. This forms the building block of a video. The main purpose of shot boundary detection is to segment the video stream into multiple shots and then key frames can be extracted from each shot. Generally Key frame refers the starting frame of the shot, which can represent the salient content of the shot. Depending on the content complexity of the shot one or more key frames can be extracted from a single shot.

Key frames provide a most desirable abstraction for video indexing, browsing, and retrieval. Key frames allow users to quickly browse over the video by viewing only a few highlighted frames. Key frames also reduce the amount of data required in video indexing and also furnish a framework to dealing with video content [4].

Due to the importance of key frame extraction, much research effort has been given in key frame extraction [4],

and research progress has been made in this area, however the existing approaches either are computationally expensive or cannot effectively capture the major visual content and also having less accuracy rate and high error rates. Accuracy rate is the ratio of extracted number of key frames to actual number of key frames. Error rate is the ratio of number of non matching key frames extracted to the actual number of key frames. The proposed method uses contour let transform [5] for extraction energy and standard deviation features for each sub band and each sub band's feature values are cascaded to form a feature vector and Euclidean distance measure is used for measuring similarity between feature vectors and then key frame classification has been done based on threshold value. This work also proved that contour let transform is good enough to extract the key frames and also proved that the method is having high accuracy rate and low error rates with selected feature values and distance measure technique.

This paper is organized as follows; section II explains the overview of the proposed work. Section III describes about importance of contour let transform. Section IV explains, feature extraction, feature vector formation, similarity measure using Euclidean distance and key frame classification. Section V shows the experimental results and section VI concludes the present work.

OVERVIEW OF THE PROPOSED WORK

The block diagram of the key frame extraction using present invention is shown in figure 1. First process is framing the input video. Then apply the contour let transform on each an input image using the decomposition parameter [4 3 3]. This contour let transform results in33 sub bands, out of which 32 are high frequency sub bands and one is low frequency sub band. In the second step energy and standard deviation features are computer for each sub band and cascade them to form a feature vector. Here feature vector length is 66, two from each sub band. In the third step, Euclidean distance is calculated between two successive frames. Finally this distance is compared against the predetermined threshold in classification process. This threshold is fixed after testing on few training videos. According to threshold value if the distance is above threshold then the frame is key frame or the frame belongs to same shot.

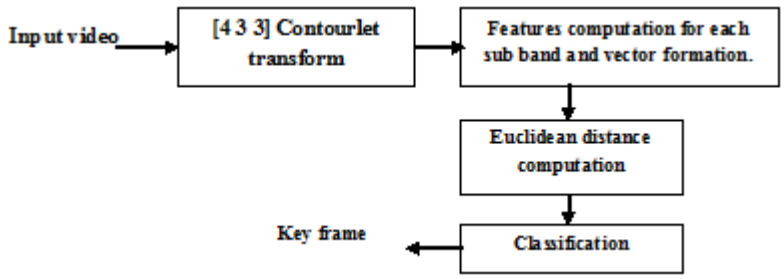

Figure 1: Block Diagram of Key Frame Extraction Process

The above step is carried out on entire video sequence to summarize the given video.

CONTOURLET TRANSFORM

The primary goal of the contour let construction was to obtain a sparse expansion for typical images that are piecewise smooth away from *smooth contours*. Two dimensional wavelets are lack of directionality and are only good at catching *point* discontinuities, but do not capture the *geometrical smoothness* of the contours [6].

Due to inefficiency in wavelet transform, contour lets were developed as an improvement over wavelets. The resulting transform has the multi scale and time-frequency localization properties of wavelets, but also offers a high degree of directionality and anisotropy. Specifically, contour let transform involves basis functions that are oriented at any

power of two's number of directions with flexible aspect ratios [6]. The general block diagram of contour let transform is shown in the figure 2.

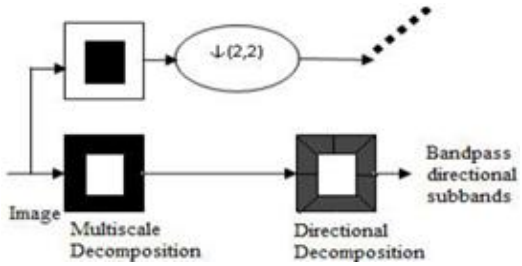

Figure 2: Contour Let Filter Bank

Maintaining the Integrity of the Specifications

The contour let transform is implemented using two dimensional Filter bank that decomposes an image into several directional sub bands at multiple scales. This is accomplished by combining the Laplacian pyramid with a directional filter bank at each scale. Due to this cascade structure, multiscale and directional decomposition stages in the contour let transform are independent of each other. Contour lets is a unique transform that can achieve a high level of flexibility in decomposition while being close to critically sampled. Other multiscale directional transforms have either a fixed number of directions, or are significantly over complete. Figure 2 shows an example frequency partition of the contour let transform where the four scales are divided into four, four, eight, and eight directional sub bands from coarse to fine scales, respectively.

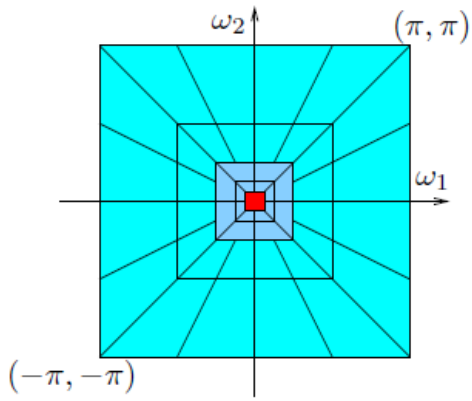

Figure 3: An Example Frequency Partition by the Contour Let Transforms

For further details about contour let transform, please read [6] [7]

FEATURE VECTOR COMPUTATION AND CLASSIFICATION

Feature Calculation and Vector Formation

So many methods have been used to construct feature vectors, here we use the energy value and stand deviation of each contour let domain directional sub-band.

$$E(s,k) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} | W_{s,k}(m,n) | \tag{1}$$

For a sub-band in contour let domain, we can use formula (1) to calculate its energy, where *E(s, k)* denotes the average energy of the band which is indexed by scale *s* and direction *k*, and *M,N* stand for the row and column number of the sub-band coefficients [8].

The stand deviation used here is defined as formula (2) where *M, N, s, k* have the same meaning as in

formula (2), σ (*s*, *k*) means the stand deviation of a certain sub-band coefficients, μ *s*, *k* denotes the average value of the sub-band coefficients.

$$\sigma(s,k) = [\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} | W_{s,k}(m,n) - \mu_{s,k} |^2]^{1/2} \quad (2)$$

Feature vector of an image is constructed using the feature values of energy and standard deviation of each sub band in the contour let image. Every image in the given video is represented with feature vector.

Determination of Similarity Measure Using Euclidean Distance

The similarity measure is used to calculate the distance between different feature vectors. we directly choose Euclidean distance as distance measure. This Euclidean distance measure shows greater performance than they are: Manhattan (L1), Weighted-Mean–Standard deviation (WMV), Euclidean (L2), Chebychev (L), Mahalanobis, Canberra, Bray-Curtis, Squared Chord, Squared Chi-Squared and Kull-back Leibler. Kokare compared the nine measures except Kull-back distance (KLD).

Classification

Euclidean distance between two feature vectors of two consecutive frames. This distance is compared against predetermined threshold. If the distance is above threshold, then the given frame is classified as a key frame, otherwise it is classified as a frame belongs to same shot.

EXPERIMENTAL RESULTS

This section introduces the implementation procedure of present work with an example and also the accuracy rate and error rates are calculated. In the transform, we chose the Laplacian pyramid and DFB filter are "pkva". This contourlet had chosen the decomposition parameter as [4 3 3] means that the numbers of directional sub-bands are 16, 8, 8. adding the low frequency sub-band, the number of sub-bands are 33, each sub-band needs two parameters to describe, so, for every image video, the dimension of feature vector is 66.

This work also tested on different videos i.e. sports video (foot ball video), cartoon video and ordinary video. The figure 4 shows distances calculated for a video of 36 frames.

This work extracted the 18 key frames out of total 36 frames of cartoon movies. Actual number of key frames using human calculation is also 18.

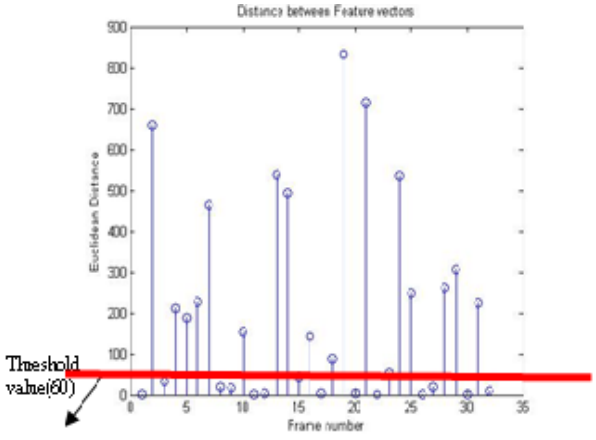

Figure 4: Distance verses Frames

The accuracy rate is defined as ratio of key frames retrieved to the actual number of key frames. And also error rate is defined as ratio of number of non key frames to the actual number of key frames. Accuracy rate varies from 0 to 1. If it is 1 then the algorithm is good enough. Error rate varies from 0 to maximum. If it is zero, then the algorithm is good.

The below figure shows the sample video sequence with original and its key frames as well.

Figure 5: Input Sample Video Sequence

Figure 6: Shows the Extracted Key Frames of the Given Sequence (The Abobe Video Sequence from Youtube. So Thanks to Youtube)

The following table gives the accuracy and error rates for the few video.

Table 1: Accuracy and Error Rates of the Present Work

Input Video	Total Frames	Key Frames	Extracted Key Frames	Extracted Non Key Frames	Accuracy Rate	Err Rate
Cartoon video	38	18	18	0	1	0
football	36	12	12	5	1	0.4
Casino movie	150	29	29	2	1	.06

CONCLUSIONS

In this present work, key frame extraction using contour let transform and energy and standard deviation as a feature values has implemented and experimentally proved and also this work had shown high accuracy rate and low error rate in key frame extraction.

REFERENCES

1. Daniel M. Russell, “A design pattern based video summarization technique: moving from low – level signals to high – level structure”, in IEEE Proc. of the 33rd Hawaii International Conference on system sciences-2000.
2. Zhu Li, Guido M. Schuster, Aggelos K. Katsaggelos and Bhavan Gandhi, “Rate – distortion optimal video summary generation”, IEEE Trans. On Image Processing, vol. 14, No. 10, Oct. 2005
3. Ying Li, Shih-Hung Lee, Chia-Hung Yeh, and C.-C. Jay Kuo, “Techniques for Movie Content Analysis and Skimming”, IEEE Signal Processing Magazine, March 2006.
4. Zhuang Y, Rui Y, Huang TS, Mehrotra S. *Adaptive key frame extraction using unsupervised clustering*. In: ICIP'98, *IEEE* Computer Society, 1998

5. Minh N. Do and Martin Vetterli, “The Contourlet Transform: An Efficient Directional Multiresolution Image Representation”, IEEE Trans. on Image Processing, vol. 14, no. 12, pp. 2091-2096, Dec 2005.

6. Duncan D. Po, Minh N. Do: Directional multiscale modeling of images using the contourlet transform. Statistical Signal Processing, 2003 IEEE Workshop on. 28 Sept- 1 Oct. 2003 pp.262-265.

7. M. N. Do and M. Vetterli, "The contour let transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process*. vol. 14, no. 12, pp. 2091-2106, Dec. 2005

8. X. Chen, G. Yu, J. Gong , “Contourlet-1.3 texture image retrieval system”, IEEE International Conf on wavelet analysis and pattern recognition (2010), pp. 49–54