Video Pause Detection using Wavelets

Shiva Zaboli and David A. Clausi
Vision and Image Processing (VIP) Lab, Systems Design Engineering
University of Waterloo, Waterloo, Canada N2L 3G1
sh.zaboli@gmail.com, dclausi@uwaterloo.ca

Abstract

As the volume of digital video captured and stored continues to increase, research efforts have focused on content management systems for video indexing and retrieval applications. A first step in generic video processing is shot boundary detection. This paper addresses a novel algorithm for abrupt shot (cut/pause) detection - especially on frames with similar statistics - based on the wavelet transform and content entropy. The algorithm has been successfully tested on some video categories including sport and live videos. Its quantitative performance has been compared to other known methods including pixel, histogram, frequency domain and statistics difference. In each test, the proposed wavelet method outperforms the others.

1. Introduction

Recent increased interest in multimedia research has drawn upon the development of video indexing and retrieval. A first step towards the semantic based video indexing and retrieval is detection of elementary video structures. Research on video content processing can be based on automatically detecting the boundaries between camera shots. A shot is a temporally continuous sequence of frames from one camera. There are a number of different types of transitions or boundaries between shots such as cut, fade, dissolve, wipe and so on. A cut is an abrupt change that occurs between two consecutive frames. A fade is a slow change in brightness usually resulting in or starting with a solid black frame. A dissolve occurs when the images of the first shot get dimmer and the images of the second shot get brighter, with frames within the transition showing one image superimposed on the other. A wipe occurs when pixels from the second shot replace those of the first shot in a regular pattern such as in a line from the left edge of the frames. Of course, other types of gradual transition are possible [2]. Cuts/pauses may separate frames with similar statistical information and have no significant differences in histograms, objects, colors or edges. In such videos, we can benefit from unique features of wavelet transform -such as multi-resolution decomposition- and content entropy.

In this paper, we propose a method of shot boundary detection on cut videos using wavelet transform and entropy features. In Section 2, existing methods are reviewed. The proposed pause detection algorithm is explained in Section 3 and experimental results are shown in Section 4.

2. Existing Techniques

Many algorithms have been proposed for detecting video shot boundaries including pixel difference [2], statistical differences, color histograms [6], compression differences, edge tracking [5], motion vectors, block matching and transform coefficients methods [3]. Pixel differences techniques are the easiest way to detect significant differences between two frames, but this method is very sensitive to camera and object motion and generates many false positives. Statistical methods expand on pixel differences by breaking images into regions and comparing statistical measures of the pixels in those regions. These techniques are tolerant of noise but are slow and also generate many false positives. Histograms are the most popular methods which are used to detect shot boundaries [2]. Histogram methods offer a good trade off between algorithm accuracy and speed [2].

3. Pause Detection Scheme

The proposed pause detection scheme (PDS) utilizes the discrete wavelet transform (DWT) and entropy features for locating cuts in a video stream.

3.1. Use of wavelets and entropy

Wavelet and entropy concepts capture unique features that can be used for content analysis of frame sequences in video streams. The DWT has a number of advantages over

978-0-7695-3651-4/09 $25.00 © 2009 IEEE
DOI 10.1109/CRV.2009.20

other transforms. The DWT breaks down a signal in the same manner as the human visual system (HVS). The DWT generally can be implemented with lower computational costs. The DWT also offers efficient energy compaction [4]. The multi-resolution characteristics of the transform is suitable for video applications.

Since Shannon's work in 1948 [8], entropy is used as a major tool in information theory. Interesting approaches involving direct use of entropy for signal processing applications can be found in areas such source seperation, blind de-convolution, source coding, image alignment and detection of abrupt changes. In many applications, a measure of complexity of underlying probability density functions, or a measure of dependence between components or signals, allows the design of an optimal processing scheme, possibly in non-stationary contexts [8]. Thus, entropy-based approaches might be useful for such problems [7], because a pause in a video is such an abrupt change.

3.2. The Proposed Pause Detection Scheme

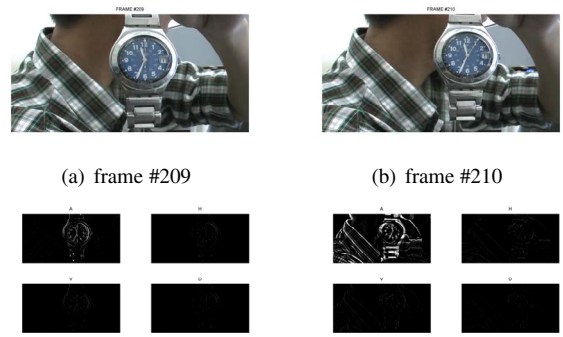
The Pause Detection Scheme (PDS) utilizes the DWT and entropy features for locating the cut frames in video stream. PDS is developed based on statistical information of frame's content. Different wavelet basis may be used. They are chosen base on the content type (sport, live, news and so on); 'haar' basis is one most popular basis that widely used in video processing. It is used for all presented results in this paper.

PDS is developed in MATLAB 7.7 for processing of AVI and WMV file formats. In this paper results for "watch" and "sport" videos are presented. "watch" is a video of running clock, in length of 1162 frames and 640×480 pixels frame size, with minimum difference in frame sequence (sometimes, just in clock hands). "sport" video is a basketball game with length of 902 frames with 640×480 pixels frame size. "watch" video is selected because of its static property, it has minimum variation and no significant changes before and after pause, and "sport" video has dynamic property such as live videos.

In this scheme, after loading the video file, for every frame in video stream; we apply the following procedure:

- Step 1: Extract the difference of current and previous frames in video stream.
- Step 2: Apply the wavelet filter to the difference frame and compute the sub-bands coefficients.
- Step 3: Compute the entropy in each sub-band as well as the mean of coefficients in each sub-band.
- Step 4: Find the pause detection index (PDI) values by the inner product of entropy and mean of each sub-band.

Figure 1- 3 show the procedure steps. Figure 1(a)- 1(b) show the frame before and after pause. Figure 1(c)- 1(d) show DWT coefficients of frame 209 and 210 (Step 2). These results are based on "watch" video in length of 1162 frames.



(a) frame #209 (b) frame #210

(c) Coefficients of frame #209 (d) Coefficients of frame #210

Figure 1. (a)-(b) last frame before pause (#209) and first frame after pause (#210) for "watch" video. (c)-(d) wavelet sub-band (A, H, V and D) coefficients of frame #209 & #210.

In Figure 2, E(A), E(H), E(V) and E(D) vectors are entropy measurements extracted from wavelet sub-band coefficients in approximation (A), horizontal (H), vertical (V) and diagonal (D) sub-bands (Step 3). For entropy measurement, we have applied equation 1 to each sub-band coefficients for every frame difference. W_{ij}^k is wavelet transform value corresponding to index (i,j) in each level that can be a negative value.

$$E(k) = -\sum_{i=1}^{N_k}\sum_{j=1}^{M_k}(W_{ij}^k)^2 \log(W_{ij}^k)^2$$
$$k \in \{A, H, V, D\} \quad (1)$$

Equation 2 is used to extract the mean value of wavelet coefficients in each sub-band. M_k and N_k are sub-band dimension of each frame (Step 3). Figure 2 shows the extracted entropy and mean value of wavelet sub-band coefficients.

$$M(k) = \frac{1}{N_k \times M_k}\sum_{i=1}^{N_k}\sum_{j=1}^{M_k} W_{ij}^k$$
$$k \in \{A, H, V, D\} \quad (2)$$

$$ME = \sum_{k=A,H,V,D} M(k) \cdot \sum_{k=A,H,V,D} E(k) \quad (3)$$

Equation 3 shows the product (ME) of summation of entropy and summation of mean value in all sub-bands that

can be used as PDI for each frame. Results are shown in Figure 3 (Step 4). Indeed ME represents rate of frame sequence changes both in time and frequency domains.

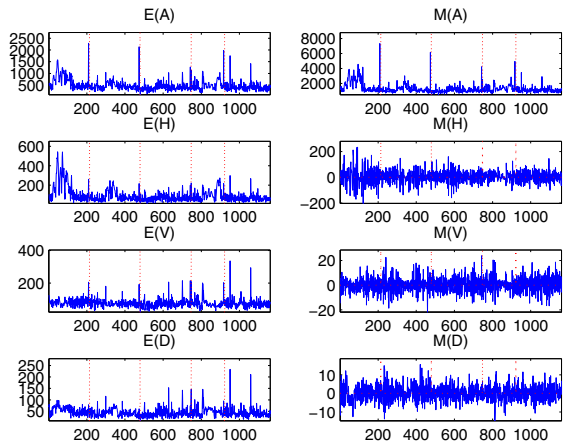


Figure 2. Extracted entropy and mean features for all video frames sequence in each sub-band [Step 3].

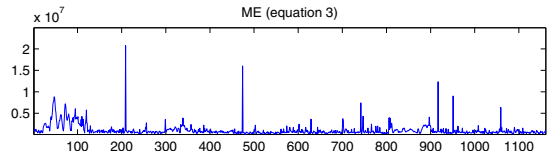


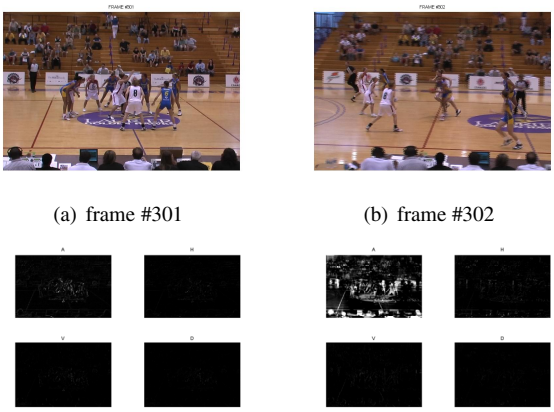
Figure 3. Product of summation of entropy and mean in all sub-bands (PDI) for "watch" video [Step 4].

The "watch" video has been paused at frames 209, 474, 742 and 917. We observe some other peaks comparable to the PDI values around paused locations in Figure 3. These peaks occur at the frames with rapid moving or vibration in frame objects. In our sample "watch" video, for four seconds at the beginning, there is distortion in the frame's scenes. A post processing stage is used to improve the detection performance and remove the undesirable peaks.

3.3. Enhancing performance

To enhance the performance of system, a post-processing stage is added to improve the accuracy of shot detections. The following figures show the PDS system outputs for "sport" video. Figure 5(a) shows the PDI plot for "sport" video and corresponding index values.

A post-processing Gaussian filter of length n+1 is applied to ME. Figure 5(b) shows the fixed PDI plot after filtering. The result is improved and unexpected peaks have been removed. In this case, three pauses are detected via suitable selected thresholds. In the "sport" video, the paused frames are located at frames number 105, 301, and 602; these are indicated with stars in the Figure 5(b).



(a) frame #301 (b) frame #302

(c) Coefficients of frame #301 (d) Coefficients of frame #302

Figure 4. (a)-(b) last frame before and first frame after pause for "sport" video. (c)-(d) wavelet sub-band (A, H, V and D) coefficients of frame #301 and #302.

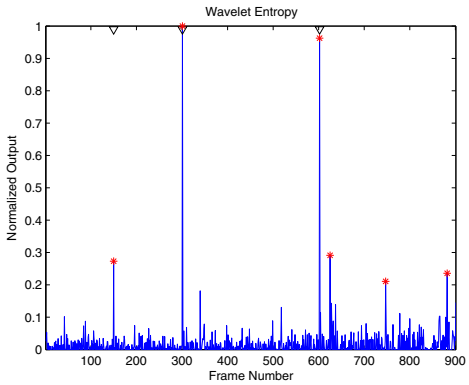
4. Testing

In this section, we compare our method with five well-known methods; then evaluate the results of these methods.

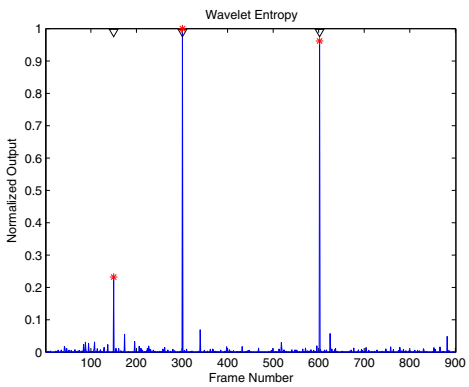
4.1. Comparison Methods

For performance evaluation of proposed scheme and comparison with other cut detection methods, we have implemented five well-known methods, including pixel, color histogram, DCT, statistical mean and standard deviation difference methods. As the source codes of these algorithms are not available, they were implemented in the same environment by ourselves. Figure 6 and Figure 7 show the results for implemented methods on test videos; triangles on top x axis determine the exact pause indexes.

- **Pixel Differences:** This is the simplest method to determine cuts. The difference between corresponding pixels of two consecutive frames is computed. If the difference is greater than the threshold, a cut is assumed [2]. Figure 6(a) and Figure 7(a) show the results of this method for "watch" video and "sport" video respectively.



(a) before filtering



(b) after filtering

Figure 5. PDI for "sport" video with pause locations at 105, 301 and 602; In (a) some unexpected peaks have been shown and in (b) unexpected peaks have been removed.

- **Statistical Differences:** Statistical methods expand on the idea of pixel differences by breaking the images into regions and comparing statistical measures of the pixels in those regions [2]. Mean and standard deviation are two statistical measurements which are used in this method. Figure 6(b)- 6(c) and Figure 7(b)- 7(c) show the results of mean standard deviation difference methods for "watch" and "sport" videos respectively.

- **Histogram:** The histogram method computes 64 bin gray level histograms of the two images and Euclidean or Chi-square distance is used to find the histogram difference [2]. Figure 6(d) and 7(d) show the results of this method for "watch" and "sport" videos.

- **DCT differences:** This method uses differences in the discrete cosine transform coefficients of frames. The same 15 DCT coefficients from each block of frame is taken and concatenated to produce a vector. The difference is computed by subtracting the vectors of consecutive frames. If this difference exceeds the threshold, declare a possible cut [1]. Figure 6(e) and 7(e) show the results of this method for "watch" and "sport" video respectively.

4.2. Evaluation

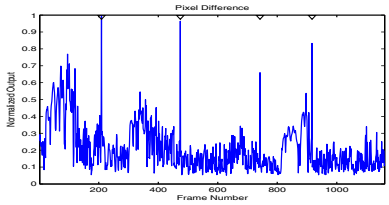
For evaluation, we have chosen recall and precision criteria [2]. Recall and precision criteria are commonly used in the field of information retrieval. Recall is defined as the percentage of desired shots that are retrieved among all. Precision is the percentage of retrieved shots that are desired shots. To make comparisons among algorithms just based on recall and precision criteria is difficult, so for each application a trade-off must be defined between recall and precision. Here, we have defined the product of recall and precision to combine the benefit of each measure. If recall and precision both have high values, their product will be increased, and if both of them are low, then the product will be decreased. This motivates to use of the product of recall and precision for evaluating the algorithms.

Figure 8 and 9 show the recall/precision product as a function of threshold. One hundred evenly distributed thresholds in steps of 0.01 from 0.01 to 1 are used. Then the product of recall and precision for each algorithm is calculated over all threshold levels. So, these figures illustrate the robustness of the threshold for each method. Some methods have narrow ranges of the threshold that are overstated and some other methods decay rapidly out of the overstated ranges. Thresholds that generate peak results should be consistent from video to video, but this is only apparent in the wavelet result.

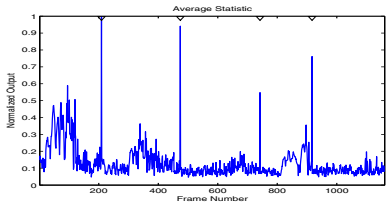
Some implemented methods are very sensitive to threshold value and their result significantly varies with different thresholds. It means that our approach is more robust than other methods to the threshold level. Figure 10 shows comparison chart on summation of recall/precision product for different methods. This chart shows that our proposed method is less sensitive to threshold value and the summation of recall and precision product over 100 threshold is greater than other methods for "sport" and "watch" videos.

5. Conclusion

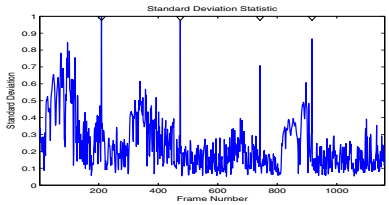
The proposed algorithm based on entropy and wavelet transform shows better results among compared cut detection methods especially for pause detection that scenes have no significant changes before and after pause such as "watch" video. In such scenes, most of cut detection methods such as histogram, statistical and transform-based methods can not detect the cut precisely, but we can detect such cuts with unique feature of wavelet transform and entropy.



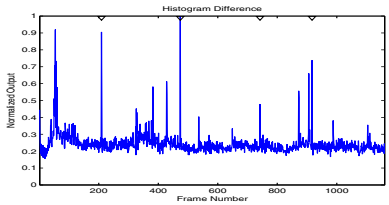
(a) Pixel Difference



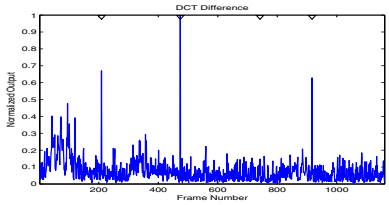
(b) Mean Statistic Difference



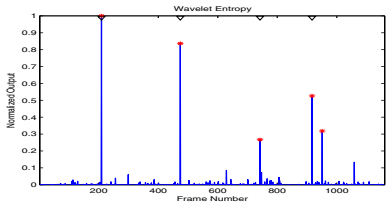
(c) Standard Deviation Difference



(d) Histogram Difference

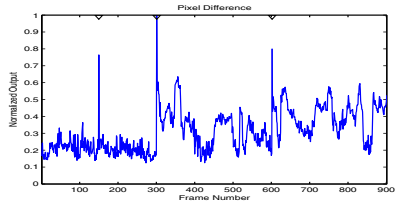


(e) DCT Difference

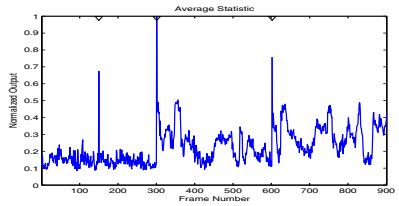


(f) Wavelet-Entropy Difference

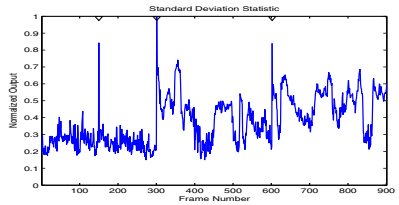
Figure 6. Experimental results (PDI) for "watch" video. The X axis shows the frame numbers and the Y axis shows normalized output of each method. (f) has better results because there are fewer unexpected peaks.



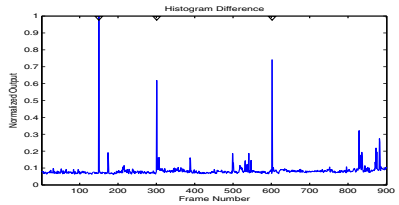
(a) Pixel Difference



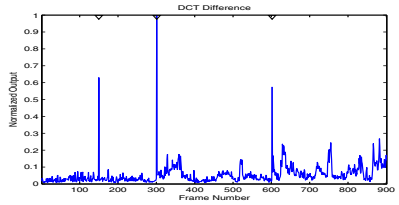
(b) Mean Statistic Difference



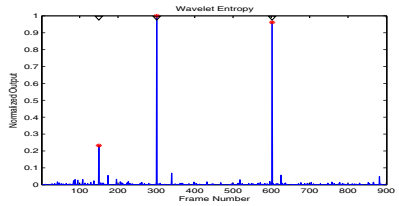
(c) Standard Deviation Difference



(d) Histogram Difference

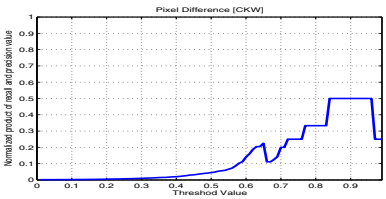


(e) DCT Difference

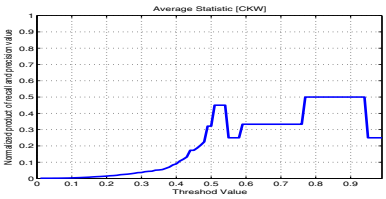


(f) Wavelet-Entropy Difference

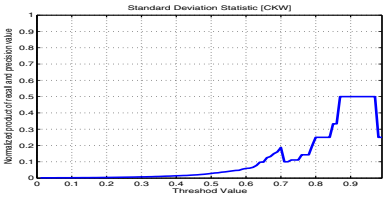
Figure 7. Experimental results (PDI) for "sport" video. The X axis shows the frame numbers and the Y axis shows normalized output of each method. (f) has better results because there are fewer unexpected peaks.



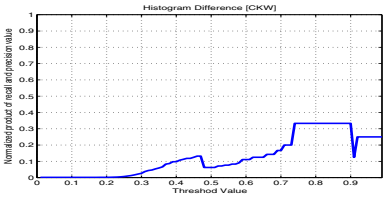
(a) Pixel Difference



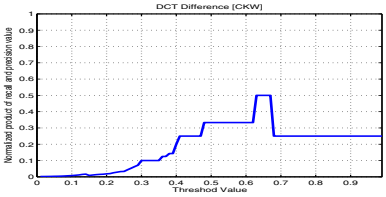
(b) Mean Statistic Difference



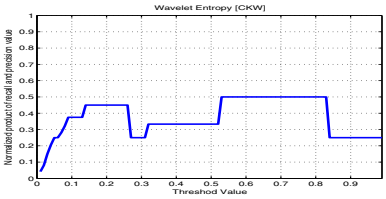
(c) Standard Deviation Difference



(d) Histogram Difference

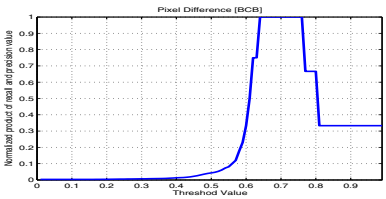


(e) DCT Difference

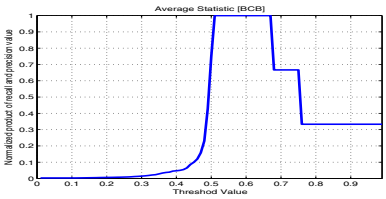


(f) Wavelet-Entropy Difference

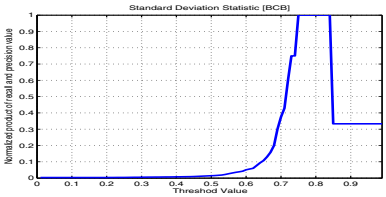
Figure 8. Product of recall and precision results of implemented algorithms (a)-(f) for "watch" video; the X axis shows threshold value between 0 and 1, and the Y axis shows normalized product of recall and precision within 100 threshold levels. (f) is less sensitive to thresholds.



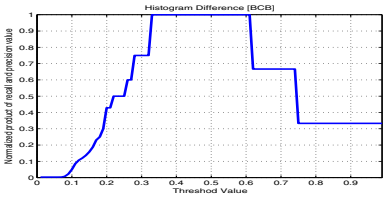
(a) Pixel Difference



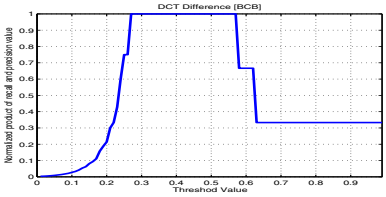
(b) Mean Statistic Difference



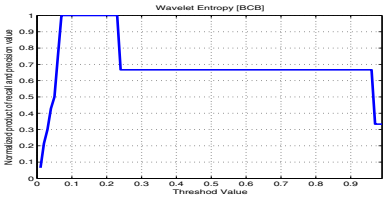
(c) Standard Deviation Difference



(d) Histogram Difference



(e) DCT Difference



(f) Wavelet-Entropy Difference

Figure 9. Product of recall and precision results of implemented algorithms (a)-(f) for "sport" video; the X axis shows threshold levels between 0 and 1, and the Y axis shows normalized product of recall and precision within 100 threshold levels. (f) is less sensitive to thresholds.

On the other hand, threshold adjustment is more robust in this method with a post-processing modification. We may improve this method by using different wavelet basis and filters. This method is suitable for off-line processing and not recommended for noisy videos or videos with disorderly scenes like traffic observation camera. We can extract different features from the estimated wavelet and entropy vectors to cover other shot detection types such as fade, dissolve and wipe.

Acknowledgements

Funding has been provided by NSERCs (Natural Sciences and Engineering Research Council of Canada) NCE (Network of Centres of Excellence) program for GEOIDE (Geomatics for Informed Decisions).

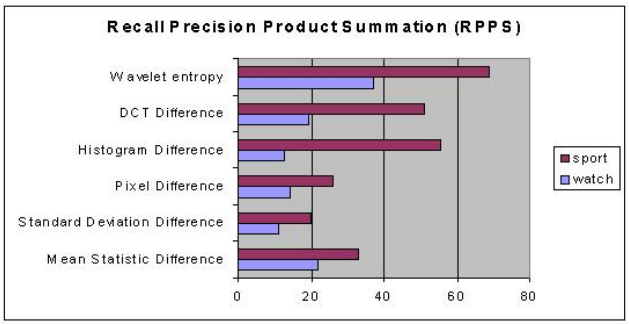


Figure 10. Recall Precision Product Summation (RPPS) of implemented methods. wavelet entropy (proposed) method has bigger results, so it's more robust under 100 thresholds.

References

[1] F. Arman, A. Hsu and M. Chiu. Image processing on encoded video sequences. *Multimedia Systems*, 1(5):211–219, 1994.

[2] J. S. Boreczky and L. A. Rowe. Comparison of video shot boundary detection techniques. *Journal of Electronic Imaging*, 5(2):122–128, 1996.

[3] E. Bruno and D. Pellerin. Video shot detection based on linear prediction of motion. *IEEE International Conference on Multimedia and Expo (ICME)*, 1:289–292, 2002.

[4] G. M. Davis and A. Nosratinia. Wavelet-based image coding: An overview. *Applied and Computational Control, Signals and Circuits*, 1(1), 1998.

[5] W. Heng and K. Ngan. An object-based shot boundary detection using edge tracing and tracking. *Journal of Visual Communication and Image Representation*, 12:217–239, 2001.

[6] J. Mas and G. Fernandez. Video shot boundary detection based on color histogram. *TREC Video Retrieval Evaluation (TRECVID'03)*, October 2003.

[7] R. Safabakhsh, S. Zaboli and A. Tabibiazar. Digital watermarking on still images using wavelet transform. *IEEE International Conference on Information Technology: Coding and Computing (ITCC'04)*, 1:671–675, April 2004.

[8] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. Journal*, 27:379–423;623–656, July/October 1948.