



Journal of
**Software
Engineering**

ISSN 1819-4311



Academic
Journals Inc.

www.academicjournals.com

Shot Boundary Detection Based on Non-negative Matrix Factorization in DT-CWT Domain

Jia-Xing Wang, Lie-jun Wang and Ting Yan

School of Information Science and Engineering, Xinjiang University, Urumqi, 830046, China

Corresponding Author: Wang Liejun, Information Science and Engineering, Xinjiang University, Urumqi, 830046, China

ABSTRACT

An efficient algorithm for shot boundary detection is proposed in this study. In order to eliminate the disturbances caused by illumination and camera motion, Dual Tree Complex Wavelet Transform (DTCWT) is used in the process of raw feature generation. Then, the enhanced texture features were extracted from the six detailed subbands by calculate the mean and standard deviation of the high-frequency coefficients. Non-negative Matrix Factorization (NMF) is used for extracting structure feature from the low-frequency sub-bands image and dynamic threshold is used for shot boundaries detection. The experiment result conducted on TRECVID 2001 test data and other test videos shows that the proposed scheme can not only overcome illumination and motion effects efficiently, but also achieve a high detection speed.

Key words: Shot boundary detection, DTCWT, texture features, NMF, dynamic threshold

INTRODUCTION

Due to the advancement of video technology, the volume of digital video data has increased significantly. To address this problem, the video summary based on key frame and video information retrieval has become an active area of research in recent years. Shot boundary detection is the basic and critical step of it.

According to the principles of video-editing, shots are joined together with either cuts or gradual transition such as dissolves, wipes and fades. Cut shots complete the scene transformation between two adjacent frames, gradual shots need continuous change between the multiple frames.

Many effective algorithm have been proposed to detect video shot boundaries (Boreczky and Rowe, 1996). In early studies (Joyce and Liu, 2006; Zhang and Smoliar, 1994; Zong-Ping *et al.*, 2006) pixel comparison and histogram comparison were used for cut detection. But they ignore the influence of illumination variation especially in the news reports, then lead to a wrong shot boundary detection. In Zong's study (Zong-Ping *et al.*, 2006), motion compensated pixel difference was used to improve the tolerance of camera and video object motion. It can achieve good performance on cut shot detection.

Detecting gradual transitions is the difficult point in video shot detection. The main reason is that it is difficult to capture the visual discontinuities and the gradual transitions detection is more likely to be confused with object and camera motions. Various algorithms (Zabih *et al.*, 1999; Yoo *et al.*, 2006; Gargi *et al.*, 2000) detect gradual transition shot by tracking the contour change of every frame to judge whether it is shot boundary. For those videos with simple scene, it may

have a good effect, especially in the detection of gradual change. But the most of the videos may have very complex background, a lot of contour informations will disturb the judgment and cause false retrieval.

From the above studies on shot boundary detection, the methods have achieved a better results, but there is no robust method can be applied in a variety of video sequences.

Detecting the gradual transition and eliminating the disturbances are the major challenges to the current shot boundary detection methods. The disturbances of illumination change and object or camera motion are often mistaken as shot boundaries. In Warhade's study (Warhade *et al.*, 2011), a method is proposed to eliminate the effect of illumination and motion, but some gradual transition shots will be missed due to characters cannot be effectively extracted. They ignored the rich texture information getting from the six detailed subbands.

Then, our approach is to, using (Warhade *et al.*, 2011) as a basis, develop a new scheme. We use Dual-Tree Complex Wavelet Transform (DT-CWT) instead of traditional metrics to extract raw features for shot boundary detection. As a result, the disturbances caused by illumination change and object motion can be eliminated due to its shift invariance and directional selectivity properties. Also, rich texture features and structure features can be extracted from the raw features. Then, an enhanced method for texture feature extraction is proposed in my study. These texture features are important for reducing the missing rate. Non-negative Matrix Factorization (NMF) is used for extracting structure feature in our scheme, owing to large amounts of data in the video, NMF can significantly decreased the computing load and the time delay.

RELATED WORK

Dual-tree complex wavelet transform: The DT-CWT have been developed by Kingsbury (1998), which allows perfect reconstruction in addition to shift invariance and directional selectivity.

It have good directional selectivity in 2 dimensions and rich phase informations. Using short linear-phase filters effectively guarantee the property of perfect reconstruction.

In dual-tree complex wavelet transform, two parallel Discrete Wavelet Transforms (DWT) are used to calculate the real component and imaginary part by providing lowpass and highpass filters in different scales as shown in the Fig. 1.

Each scale can generate two low frequency sub-bands and 6 different directional ($\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$) high frequency subbands. It is insensitive to illumination variations and camera motion (Selesnick *et al.*, 2005). The property of limited redundancy and computationally efficient make it widely applied in characteristics extraction area. The detail explanation for 2D dual-tree complex wavelet transform has been given in (Selesnick *et al.*, 2005).

Non-negative matrix factorization: Various feature extraction methods have been used for shot boundary detection, due to there are large amounts of data in the video, one effective method for dimensional reduction can improve the efficiency and the rate obviously. Linear subspace projection has been widely used for feature extraction such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA).

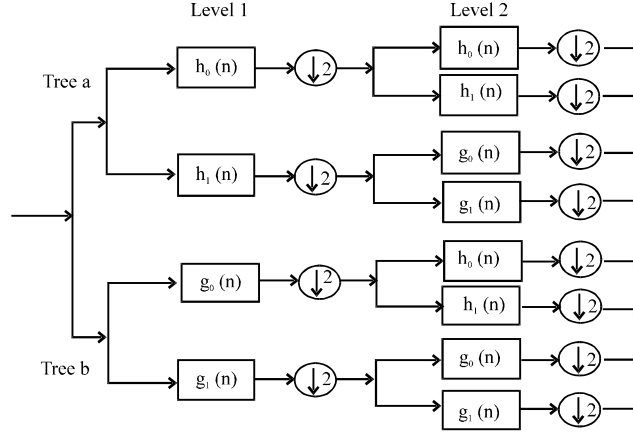


Fig. 1: Dual-tree complex wavelet transform

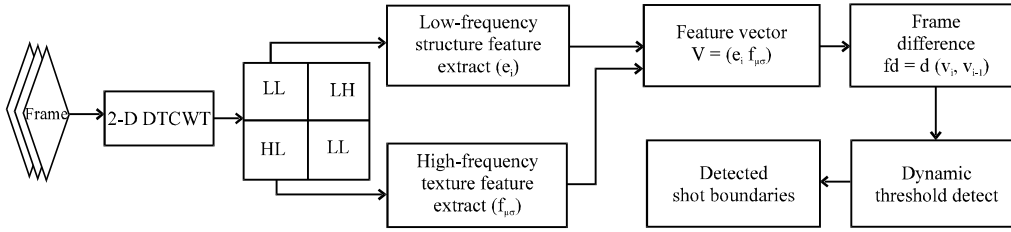


Fig. 2: Framework of the study method to detect the shot boundaries

But on decomposition process, negative values are existed in this algorithms, negative values are meaningless in reality, the positive and negative coefficients will offset each other, it will weaken the effects of features extraction. To address this problem, Lee and Seung proposed Non-negative Matrix Factorization (NMF) in 1999 (Lee and Seung, 1999). NMF is a new method of data analysis and processing, tend to local feature extraction. It possesses the characteristics of small storage capacity, simple iteration and fast convergent rate.

NEW SHOT DETECTION METHOD

The new method has the following major steps: (1) Raw feature generation from DT-CWT, (2) The texture feature extracted from high frequency coefficients based on mean and variance, (3) The low frequency sub-bands structure feature extraction based on NMF and (4) Dynamic threshold for shot boundaries detection.

The framework of the paper method to detect the shot boundaries is shown in Fig. 2.

Raw feature generation from dual-tree complex wavelet transform: Discrete Wavelet Transform (DWT) lacks shift invariance and has poor directional selectivity. We use DT-CWT to detect shot boundaries in video because of its shift invariance and directional selectivity property.

The base of 2-D discrete wavelet transform is shown in Fig. 3, there are three directions (horizontal, vertical, diagonal direction). The base of 2-D dual-tree complex wavelet

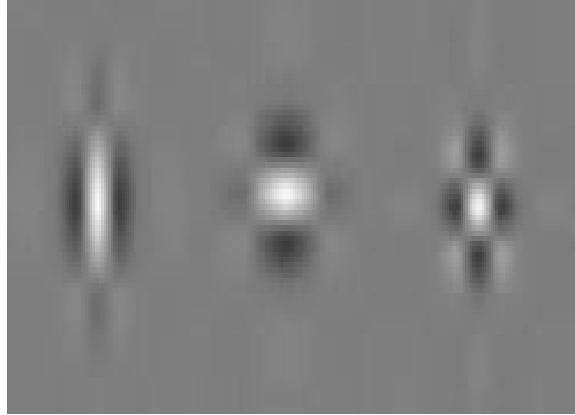


Fig. 3: 2-D discrete wavelet base

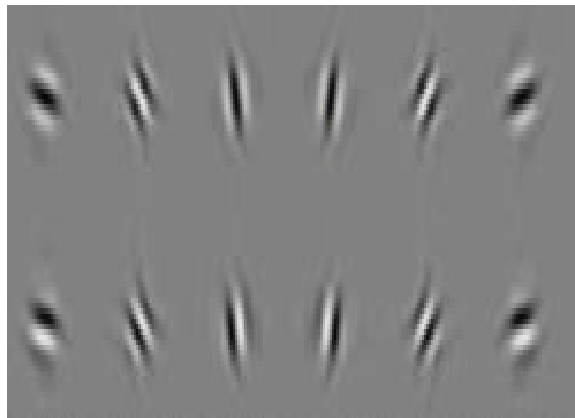


Fig. 4: 2-D dual-tree complex wavelet bases

transform is shown in Fig. 4, there are six directions (+15, +45 and +75°). The first line shows six direction of real component, the second line shows imaginary part.

From Fig. 3 and 4, to the property of directional selectivity, 2-D DT-CWT is better than 2-D DWT and the direction of the actual texture is very rich, so, 2-D DT-CWT can better than 2-D DWT depict texture features.

In our approach, every frame was decomposed into 12 sub-bands using 2D DT-CWT at first level of decomposition. Then, the real and imaginary coefficients of each sub-band was obtained. These 12 sub-bands gave detail and contour information of the frame strongly oriented at (+15, +45, +75, -15, -45 and -75°) directions for six real and six imaginary sub-bands.

Texture feature extracted: A variety of texture feature extraction methods have been proposed. The wavelet domain (Mallat, 1989) is quite appropriate for the purpose of texture description, because texture is the reflection of a local structural features of the each frame. It shows gray level or color changes in the neighborhood pixel.

The interrelationship between the pixels according to the human perception system (Sheikh and Bovik, 2006). Many methods use wavelet to extract the texture feature (Smith and Chang, 1994; Do and Vetterli, 2002). However, many texture patterns have different directional properties in similar space (Selesnick *et al.*, 2005). To address this problem, approximate shift invariance and good directional selectivity properties of DT-CWT make it particularly useful in texture analysis.

DT-CWT are used to extract the texture feature from the six detailed subbands ($\pm 15^\circ$, $\pm 45^\circ$ and $\pm 75^\circ$) at each level (Priya and Rajesh, 2010). The statistical measures such as mean and variance of the complex coefficients at each decomposition level are used to represent the frame description. μ_k and σ_k is the mean and variance of high frequency coefficients in the k th layer (Eq. 1-2):

$$\mu_k = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N |W_k(i, j)| \quad (1)$$

$$\sigma_k = \sqrt{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N |W_k(i, j) - \mu_k|^2} \quad (2)$$

where, $W_k(i, j)$ is the high frequency coefficients of each sub-band, $M \times N$ is the size of the wavelet sub-band, Thus, μ_k and σ_k can structure the feature vectors, for a n layers DTCWT, the feature vectors can be expressed as Eq. 3:

$$\vec{f}_i = [\mu_1 \mu_2 \dots \mu_n \sigma_1 \sigma_2 \dots \sigma_n] \quad (3)$$

In this study, we choose a video clip from a documentary to test the performance of the texture feature extraction, as show in Fig. 5. In this clip, there is a gradual transition shot boundary between frame 10 and frame 34, frame number 10 and 11 are in a same shot, frame number 33 and 34 are from another shot, gradual transition shot boundary are at the frame number 20 and 21. There is a cut shot boundary between frame 72 and frame 73.

A 4 layers DTCWT is used to extract the texture feature from the six detailed subbands ($\pm 15^\circ$, $\pm 45^\circ$ and $\pm 75^\circ$) at each level (Priya and Rajesh, 2010), so the feature vectors can be expressed as:

$$\vec{f}_i = [\mu_1 \mu_2 \mu_3 \mu_4 \sigma_1 \sigma_2 \sigma_3 \sigma_4]$$

The performance of the frame difference is shown in Fig. 6 and 7. From the Fig. 6 and 7, we can find that only the mean μ_4 and variance σ_4 of high frequency coefficients have a big difference in the 4th layer, it can better distinguish the frames from different shots. While other mean and variance of each layer, especially the mean μ_4 and variance σ_4 at the first layer are very similar. It is difficult to distinguish the frames, whether it's a gradual transition shot boundary or a cut shot boundary.

So, in our approach, we proposed a new method to extract the texture feature. We only choose the high frequency coefficients in the 4th layer to represent the frame description. Then, we can calculate the mean μ_3 and variance σ_3 of the six detailed subbands ($\pm 15^\circ$, $\pm 45^\circ$ and $\pm 75^\circ$) in the 4th layer to extract the texture feature as expressed in Eq. 4-5:

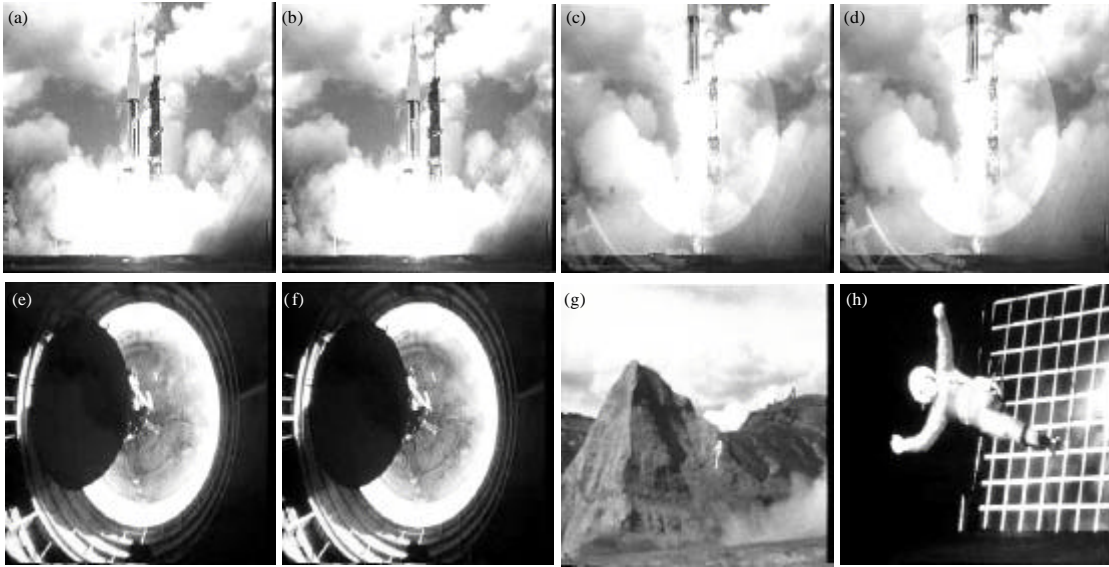


Fig. 5(a-h): A video clip from a documentary, (a) Frame 10, (b) Frame 11, (c) Frame 20, (d) Frame 21, (e) Frame 33, (f) Frame 34, (g) Frame 72 and (h) Frame 73

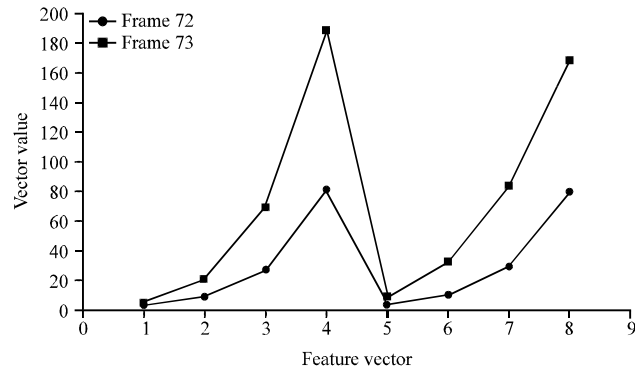


Fig. 6: Texture feature extraction at each layer for cut shot

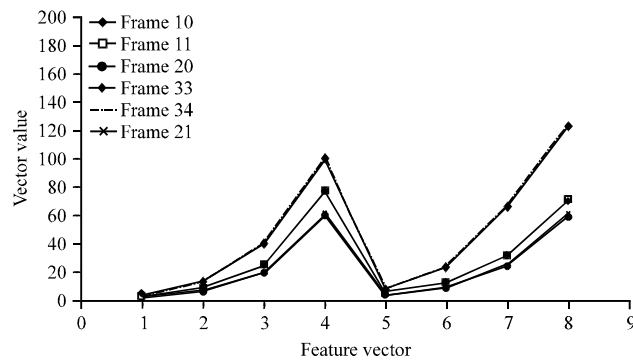


Fig. 7: Texture feature extraction at each layer for gradual transition shot

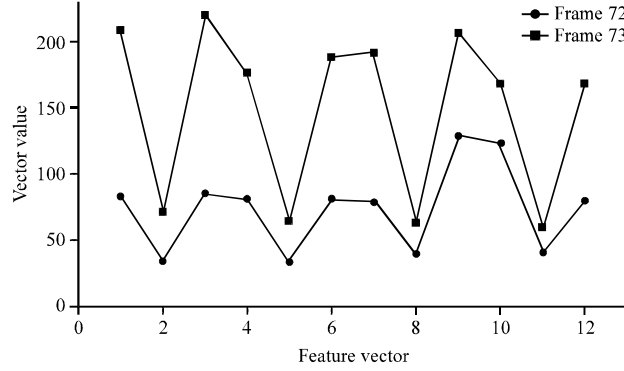


Fig. 8: Texture feature extraction in 4th layer for cut shot (our method)

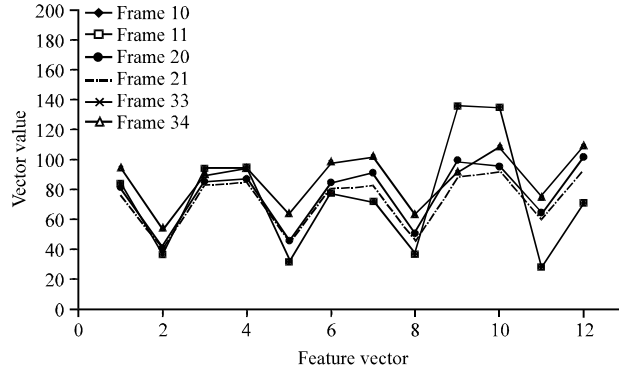


Fig. 9: Texture feature extraction in 4th layer for gradual transition shot (our method)

$$\mu_s = \frac{1}{P \times Q} \sum_{i=1}^P \sum_{j=1}^Q |I_s(i, j)| \quad (4)$$

$$\sigma_s = \sqrt{\frac{1}{P \times Q} \sum_{i=1}^P \sum_{j=1}^Q |I_s(i, j) - \mu_s|^2} \quad (5)$$

where, $I_s(i, j)$ is the high frequency coefficients of the six detailed subband, $P \times Q$ is the size of the wavelet sub-band in the 4th layer. The feature vector can be expressed as Eq. 6:

$$\vec{f}_{\mu\sigma} = [\mu_1 \mu_2 \dots \mu_6 \sigma_1 \sigma_2 \dots \sigma_6] \quad (6)$$

The performance of the frame difference is shown in Fig. 8 and 9. From the Fig. 8 and 9, we can find that the values of mean μ_s and variance σ_s are different in different shots. To the gradual transition shot boundary, the frame number 10 and 11 from a same shot almost have the same value, so as the frame 33 and 34. The frame 20 and 21 is the gradual transition shot boundary, so it have different values. To the cut shot boundary, the frame 72 and frame 73 have a difference on each feature point, the effect of cut shot boundary detection is very good. The new method of texture feature extraction in our study can better distinguish the frames from different shots.

Low frequency sub-bands structure feature extraction based on NMF: In NMF, as a result of part based representation and clear physical meanings, there are no negative values in the basic matrix and rebuild coefficients. Non-negative Matrix Factorization (NMF) (Lee and Seung, 1999) has been widely used in feature extraction (Chang *et al.*, 2010).

Giving a non-negative matrix V , in order to estimate the matrix V , NMF need to find two non-negative matrix W and H makes $V \approx WH$. Then, matrix V can approximately factorized into an $n \times r$ matrix W and a $r \times m$ matrix H . Where the range of r need to satisfy the condition: $(m+n) \times r < m \times n$. A columns of W is weighted by the components of H and the result of linear combination is the approximate value of each data vector in V . Therefore, W can be regarded as a basis that makes the V have a optimal estimation. Given a data matrix V , Lee and Seung (1999) found a technique for factorizing V to yield matrices W and H . They select the Euclidean distance as cost function defined as Eq. 7:

$$F = \sum_{i=1}^n \sum_{u=1}^m (V_{iu} - (WH)_{iu})^2 \quad (7)$$

The following iterative learning rules are used to find the linear decomposition as shown in Eq. 8-9:

$$H_{iu} \leftarrow H_{iu} \frac{(W^T V)_{iu}}{(W^T W H)_{iu}} \quad (8)$$

$$W_{iu} \leftarrow W_{iu} \frac{(V H^T)_{iu}}{(W H H^T)_{iu}} \quad (9)$$

After the 2-D dual-tree complex wavelet transform, we will get a low frequency sub-bands image with the size of $m \times n$. First, it is represented by a $m \times n$ -D column vector. Non-negative data matrix $V(m \times n, t)$ is composed of those vectors from a video which has total number of frames is t . We get decomposed matrices W and H which satisfy $V \approx WH$ by non-negative matrix factorization algorithm. The columns of the decomposed matrix W are regarded as the basis vectors.

Project the $m \times n$ -D column vector X_i to the subspace which is formed by these basis vectors $e_i = W^T X_i$. Thus, we obtain r -D e_i as feature vectors for each frame. Figure 10 is the framework of this step.

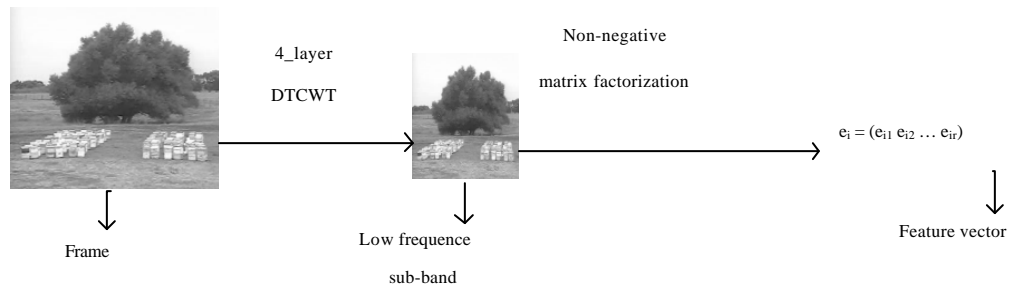


Fig. 10: Framework of low frequency sub-bands feature extraction

Dynamic threshold for shot boundaries detection: The key point to using frame differences to detect the shot boundary is the threshold selection. Many methods propose a fixed threshold to detect the shot boundary (Cernekova *et al.*, 2003) many shot boundaries will be missed due to the presence of some smaller or larger frame differences. The flexibility and adaptability of algorithm will be restricted. In our scheme, we use dynamic threshold to separate video frames into shots and detect the shot boundaries. Euclidean distance is used in our method as dissimilarity measure between two vectors \mathbf{v} and \mathbf{v}_{i-1} :

$$d(\bar{\mathbf{v}}_i, \bar{\mathbf{v}}_{i-1}) = \|\bar{\mathbf{v}}_i - \bar{\mathbf{v}}_{i-1}\|_2 \quad (10)$$

In fact, the range of frame differences change is very small in a shot and the frame differences is approximate to the average frame differences. Meanwhile, the frame differences in a shot boundary are significantly larger than the average frame differences in this shot. Therefore, if there has a cut shot, the current frame difference is larger than the average frame difference which is before this cut shot boundary and the next frame difference is far less than the current frame difference.

But to a gradual transition shot boundary, the frame difference of the first frame for the gradual transition shot boundary is larger than the average frame difference which is before the gradual transition shot boundary. The frame difference of the end frame for the gradual transition shot boundary is smaller than the average frame difference which is the gradual transition shot boundary. The frame difference in the gradual transition is usually approximately equal to others in the same gradual transition shot boundary.

In our algorithm, a temporal window of size s ($s = 20$) is used to further distinguish the hard cut shot boundary and gradual transition shot boundary. Using variate avg_bf denote the average frame difference which is before a shot boundary in a temporal window. Using variate avg_gd denote the average frame difference of gradual transition shot boundary. Our algorithm is as follows:

Proposed algorithm

Initialization: Get the first N ($N = 5$) vectors, calculate the frame differences, then get the mean of the frame differences, using the mean to initialize the variate avg_bf .

Dynamic threshold:

Step 1: Get a new frame feature vector $\mathbf{v}_i = (e_i, f_{\mu 0})$ and calculate the frame difference $\text{fd}_{(i,i-1)} = d(\bar{\mathbf{v}}_i, \bar{\mathbf{v}}_{i-1})$

Step 2: If $\text{fd}_{(i-1,i)} > \alpha \cdot \text{avg_fd}_{(i-1-s,i-1)}$ and there is not a frame of shot boundary have been declared, moving to the step 3, using the value of $\text{fd}_{(i-1,i)}$ to initialize the variate avg_gd , where is a predefined parameter to control the variation ranges of the dynamic threshold. Otherwise move the until this condition can be satisfied

Step 3: Get next frame and calculate the frame difference $\text{fd}_{(i,i+1)}$

If $\text{fd}_{(i,i+1)}$ is larger less than $\text{fd}_{(i+1,i)}$ a cut shot boundary is found

If $\text{fd}_{(i,i+1)} > \alpha \cdot \text{avg_fd}_{(i-1-s,i-1)}$ and $\text{fd}_{(i,i+1)}$ is approximately equal to avg_gd , the frame can be regarded as the gradient frame. Calculate the mean of the gradient frame differences to update the variate avg_gd

If $\text{fd}_{(i,i+1)} < \alpha \cdot \text{avg_gd}$, the current frame will be declared as the end frame of the gradual transition shot boundary and a gradual transition shot boundary is found

Step 4: Move the slide window and repeat steps 1-3

EXPERIMENTAL RESULTS

In the experiments, a variety of test videos are collected from the TRECVID 2001 test data, including news, moves and documentaries. The test video data contain many editing effects and camera motion. The experimental results are shown in Table 1. The performance is evaluated by recall, precision and F1 metrics that are defined as Eq. 11-13:

$$\text{Recall} = \frac{N_C}{N_C + N_M} \quad (11)$$

$$\text{Precision} = \frac{N_C}{N_C + N_F} \quad (12)$$

$$F_1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (13)$$

where, N_C is the number of shot boundaries that are detected correctly, N_M is the number of shot boundaries that are missed, N_F is the number of shot boundaries that are detected falsely and F_1 is a measure considering both recall and precision.

To test the robustness of the texture feature extraction which used in our method, a comparative experiment was designed. Firstly, we only extracted the low frequency sub-bands features from the test videos (Warhade *et al.*, 2011), the performance is shown in Fig. 11.

Table 1: Test videos for shot boundary detection

Video	Transition			
	Frames	Total	Cut	Gradual
News 1	3720	23	21	2
News 2	3360	18	16	2
Movie	3938	20	6	14
Satellite launching	1812	12	6	6

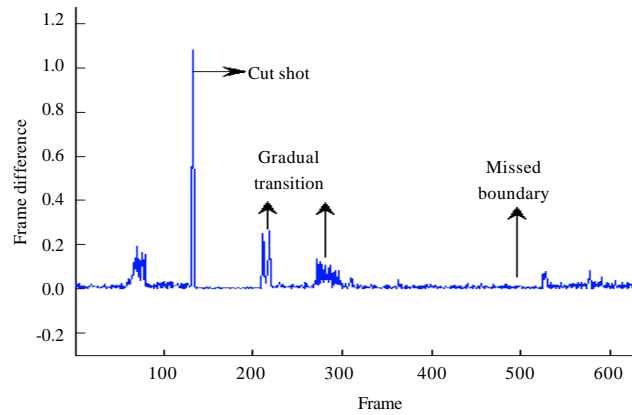


Fig. 11: Video shot boundaries are detected using low frequency sub-bands

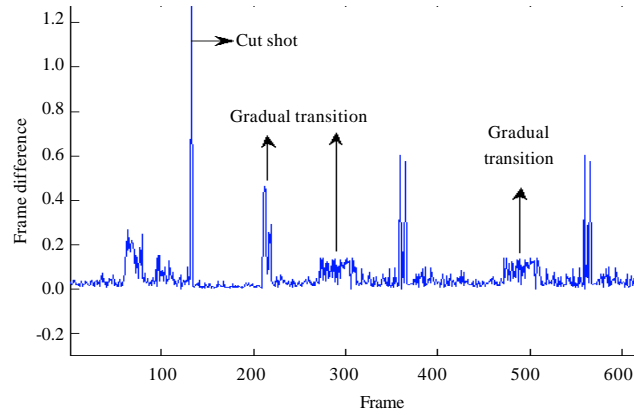


Fig. 12: Video shot boundaries are detected using low frequency sub-bands, features and enhanced texture features (our method)

Table 2: Results of shot boundary detection

Video	Li <i>et al.</i> (2009)			Zong-ping <i>et al.</i> (2006)			Proposed method		
	Recall	Pre	F1	Recall	Pre	F1	Recall	Pre	F1
News 1	95.2	90.9	93.0	90.0	85.7	87.7	95.4	95.4	95.4
News 2	94.1	94.1	94.1	93.3	82.3	87.4	88.9	100.0	94.1
Movie	88.2	83.3	85.6	88.2	83.3	85.6	88.9	88.9	88.9
Satellite launching	77.7	70.0	73.6	87.5	63.6	73.6	81.8	90.0	85.7

Secondly, we use our method to extract the low frequency sub-bands features and enhanced texture features of the six detailed subbands. The high frequency information preserve more detail information and it provides rich texture feature to the edge feature extraction of images. The performance is shown in Fig. 12. From the figures, we can find that only extracted the low frequency sub-bands features will be missing some gradual transition shot boundaries due to it can't extracted the frame difference effectively. Experiment result shows that our scheme effectively guaranteed the gradual transition detection. The tiny frames variations between a gradual transition shot can be captured because of the enhanced texture features extraction.

In the experiments, we also choose many video clips which have the disturbances caused by fast object motion to compare our method with the other methods (Li *et al.*, 2009). Table 2 lists the performance of the proposed algorithm compared with Li algorithm (Li *et al.*, 2009), which used the Discrete Wavelet Transform (DWT) for shot boundary detection.

From the experimental results, we can see that the performance of our method are better than Li algorithm and Warhade algorithm. Li algorithm, used the comparison of the color difference and edge difference as the feature, it didn't detect the shot boundaries effectively due to the disturbances caused by fast object and camera motion. Dual-tree complex wavelet transform is used in our algorithm to eliminate the disturbances and attain the raw features, because of its shift invariance and directional selectivity property and rich texture features, the problem of missing shot boundaries are avoided effectively, our algorithm achieves better overall performance.

CONCLUSION

In this study, a robust algorithm has been proposed for shot boundary detection, we use the DT-CWT to eliminate the disturbances due to illumination and fast camera motion. Texture features were extracted from the six detailed subbands and the structure features were extracted based on NMF which is the effective method for dimensional reduction, it can improve the efficiency and the rate obviously. We select many video clips which have the disturbances caused by fast object motion to compare our method with the other methods. It has been observed that the proposed scheme achieves better performance. It is robust to illumination and camera motion. Our future work is to extend the method to extract key frames.

ACKNOWLEDGMENT

Thanks for this study is supported by the National Natural Science Foundation of China under the project "Research on methods of mobile-based face recognition in Xin jiang Uyghur people" with grant No. 61261036.

REFERENCES

- Boreczky, J.S. and L.A. Rowe, 1996. Comparison of video shot boundary detection techniques. *J. Electron. Imaging*, 5: 122-128.
- Cerneкова, Z., C. Kotropoulos and I. Pitas, 2003. Video shot segmentation using singular value decomposition. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, April 6-10, 2003, Hong Kong, China, pp: 181-184.
- Chang, L., H. Kun, J. Zhou and J. Zhang, 2010. Generalized discriminant orthogonal non-negative matrix factorization. *J. Comput. Inf. Syst.*, 6: 1743-1750.
- Do, M.N. and M. Vetterli, 2002. Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models. *IEEE Trans. Multimedia*, 4: 517-527.
- Gargi, U., R. Kasturi and S.H. Strayer, 2000. Performance characterization of video-shot-change detection methods. *IEEE Trans. Circuits Syst. Video Technol.*, 10: 1-13.
- Joyce, R.A. and B. Liu, 2006. Temporal segmentation of video using frame and histogram space. *IEEE Trans. Multimedia*, 8: 130-140.
- Kingsbury, N.G., 1998. The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters. *Proceedings of the 8th IEEE DSP Workshop*, August 9-12, 1998, Utah, pp: 319-322.
- Lee, D.D. and H.S. Seung, 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401: 788-791.
- Li, J., Y. Ding, Y. Shi and Q. Zeng, 2009. Dwt-based shot boundary detection using support vector machine. *Proceedings of the 5th International Conference on Information Assurance and Security*, August 18-20, 2009, Xi'an, China, pp: 435-438.
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11: 674-693.
- Priya, K.J. and R.S. Rajesh, 2010. Local fusion of complex dual-tree wavelet coefficients based face recognition for single sample problem. *Procedia Comput. Sci.*, 2: 94-100.
- Selesnick, I.W., R.G. Baraniuk and N.G. Kingsbury, 2005. The dual-tree complex wavelet transform. *IEEE Signal Process. Mag.*, 22: 123-151.
- Sheikh, H.R. and A.C. Bovik, 2006. Image information and visual quality. *IEEE Trans. Image Proc.*, 15: 430-444.

- Smith, J.R. and S.F. Chang, 1994. Transform features for texture classification and discrimination in large image databases. Proceedings of the IEEE International Conference Image Processing, November 13-16, 1994, Austin, TX., pp: 407-411.
- Warhade, K.K., S.N. Merchant and U.B. Desai, 2011. Performance evaluation of shot boundary detection metrics in the presence of object and camera motion. IETE J. Res., 57: 461-466.
- Yoo, H.W., H.J. Ryoo and D.S. Jang, 2006. Gradual shot boundary detection using localized edge blocks. Multimedia Tools Appl., 28: 283-300.
- Zabih, R., J. Miller and K. Mai, 1999. A feature-based algorithm for detecting and classifying production effects. Multimedia Syst., 7: 119-128.
- Zhang, H.J. and S.W. Smoliar, 1994. Developing power tools for video indexing and retrieval. Proceedings of the International Symposium on Electronic Imaging: Science and Technology, International Society for Optics and Photonics, November 13-16, 1994, San Diego, CA., pp: 140-149.
- Zong-Ping, Z., L. Kun and P. Ji-Hu, 2006. Shot boundary detection based on histogram of mismatching-pixel count of FMB. Proceedings of the 1st IEEE Conference on Industrial Electronics and Applications, May 24-26, 2006, Singapore, pp: 1-5.