I J C T A, 9(7), 2016, pp. 3231-3238
© International Science Press

ISSN: 0974-5572

Non Subsampled Contourlet Transform Based Shot Boundary Detection in Videos

Sasithradevi A.*, S. Md. Mansoor Roomi and R. Raja*****

ABSTRACT

Increasing popularity and capacity of data storage devices lead to the subsequent development of large video data repository. With the explosive growth of multimedia data in internet, there is a tremendous need for researches in video indexing, retrieval, summarization and analysis. Video Shot Boundary Detection (VSBD) is the primary step for most of the prevailing video processing algorithms and content based video applications. A video shot is a sequence of continuous frames captured on a single camera. VSBD under varying lighting condition is one of the main challenging issues, specifically in entertainment videos. Hence, a robust technique that can detect shot boundary with minimal error rate is proposed. In this paper, Non Subsampled Contourlet Transform (NSCT) based illumination invariant cut detection method is addressed to detect the shot boundaries under varying lighting effects. The combination of normalized color model and NSCT guarantees the illumination invariant VSBD technique. The experimentation on test video using the proposed illumination invariant shot boundary detection methodology shows promising results.

Keywords: Shot Boundary Detection, Non Subsampled Contourlet Transform, Illumination effect, Threshold.

1. INTRODUCTION

The ever-growing video databases on the internet to watch TV shows and movie provoked researches in video database management, video indexing, video retrieval, etc. The primary task for any semantic video analysis algorithm is the video shot boundary detection. A Video shot is a sequence of frames, continuous in time and space from the perspective of a single camera. Transition on video shots can be broadly categorized into abrupt cut (hard cut) and gradual cut (Dissolve, fade in, fade out, wipe).An abrupt cut exists between the frame belonging to one shot and the subsequent shot as in Fig. 1. No transition frames occur between two frames in abrupt transition as its name implies. If transition frames occur between two shots, the transition is a gradual change. Fades can be further classified into fade in and fade out. A gradual transition between a scene and a constant image is termed as 'fade in' and 'fade out' is the inverse of fade in. 'Dissolve' refers to a whole picture fades away one frame to another frame. 'Wipe' is a gradual transition from one side of the frame to the other side as line moves across the frame (with shapes like clock, star, heart, etc).

2. RELATED WORK

Several works have been found in literature for shot boundary detection [1-5]. The common steps involved in shot boundary detection can be listed as:

1. Extracting visual content features from each frame.
2. Computing similarity/dissimilarity measures between the extracted features.
3. Detecting shot boundaries using these measures.

* Research Scholar, *Email: devisasithra@gmail.com*

** Assistant Professor, Thiagarajar College of Engineering, Madurai, *Email: smmroomi@tce.edu*

*** Professor, Pandian Saraswathi Yadav Engineering College, Sivagangai, *Email: raja.raju@rediffmail.com*

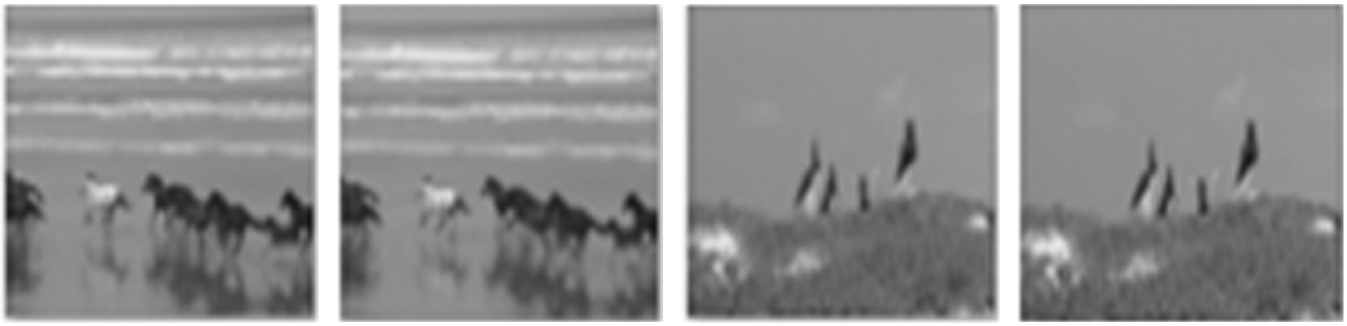


Figure 1: Sample frames for abrupt transition

Various feature representation techniques reported earlier are based on color histogram/block histogram and pixel intensities. But these techniques are vulnerable to lighting effects, object motion and camera motion. Even though histogram based shot boundary detection is robust to object motion [2], it fails under varying lighting effects and also when successive frames contain dissimilar content which results in similar histogram. Compared to histogram based methods, edge feature based methods [3, 4] are invariant to illumination and slightly variant to camera and object motion.

Though numerous researches are available in this field, shot boundary detection on entertainment videos is vital since these videos get more average views in youtube. Several algorithms proposed in literature can detect shot boundaries on benchmark dataset, but their performance is not fair in entertainment videos, owing to its complex characteristics like illumination, camera zoom, object motion, camera motion and extremely varying lighting conditions. In order to improve the performance of video shot boundary detection in such complicated entertainment videos, a method based on normalized color domain and Non Subsampling Contourlet Transform (NSCT) which extracts illumination invariant geometric features from video frames is proposed. The main contributions in this work are listed as:

- A normalized domain for illumination change in individual channels of the frame namely red, green and blue channels is presented.
- A multidirectional, multiscale, shift invariant and over complete transform referred as Non Subsampled Contourlet Transform (NSCT) is used for efficient representation of smooth contours and geometrical structures in the illumination normalized frames.
- A threshold approach for shot boundary detection is presented.

3. BACKGROUND

In this section, the construction of NSCT [6] is discussed briefly. Contourlet transform, the extension of wavelet transform provides efficient representation of images in diverse orientations. The Contourlet transform uses Laplacian pyramid and directional filter banks (DFB) for multiscale and directional decomposition respectively. Images can be represented more flexibly and completely using Non Subsampled Contourlet Transform (NSCT) by allowing redundancy. NSCT achieves the shift invariant property using non sub sampled pyramids (NSP) and non sub sampled DFB (NSDFB).

3.1. Non Sumsampled Pyramids

The subband decomposition obtained from the NSP filtering structure is similar to the Laplacian pyramid. The NSCT decomposition using NSP with decomposition stages N = 3 is depicted in fig. 2. NSP decomposition employs no up sampling or down sampling which guarantees the shift in-variant nature of the non sub sampled filter banks. The perfect reconstruction condition is given as,

$$\sum_{i=0,1} \Lambda_i(\omega)\tilde{\Lambda}_i(\omega) \equiv 1 \tag{1}$$

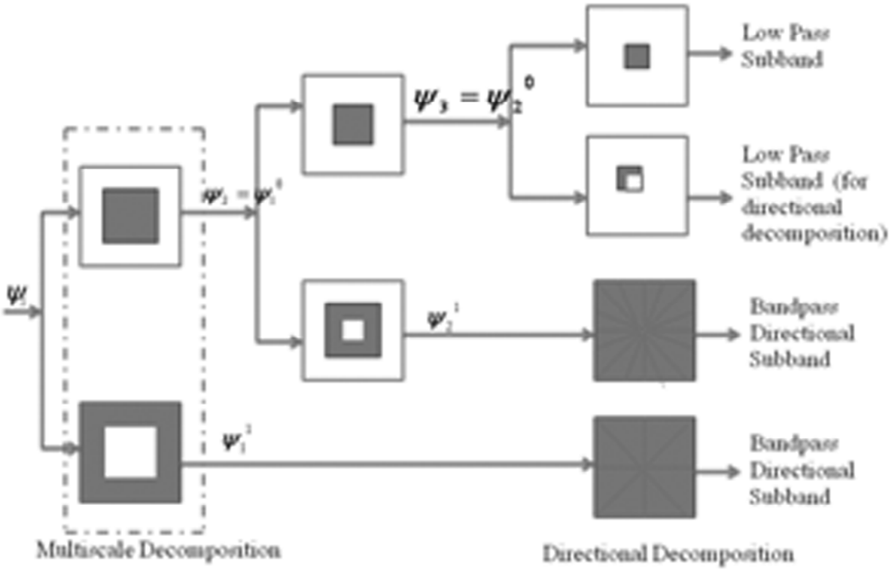


Figure 2: NSCT decomposition

Where $\Lambda_i(\omega)$ is the set of analysis filter and $\tilde{\Lambda}_i(\omega)$ is the set of synthesis filter. As the filter banks do not employ critical sampling, satisfying the reconstruction condition (1) is much simpler. Let ψ_k be the input signal at the k^{th} level, $1 \leq k \leq N$. At the k-th level decomposition, the input signal ψ_k is decomposed into low pass subband ψ_k^0 and high pass subband ψ_k^1 using low pass filter h_0 and high pass filter h_1 respectively as follows,

$$\psi_k^i = h_i \otimes \psi_k \tag{2}$$

Where $\otimes$ is the convolution operator.

3.2. Non Sumsampled Directional Filter Banks

The high pass subband is decomposed into several directional subbands using Non Subsampled Directional Filter Banks (NSDFB). The next stage involves the same NSP and NSDFB decomposition of . NSDFB is a combination of two channel fan filters and parallelogram filters without downsamplers and upsamplers. The DFB can be efficiently implemented using l-level binary tree decomposition that leads to 2l subbands with wedge shaped frequency partitioning. In this work, DFB decomposition stage is composed of modulating the input image and using Quincunx filter banks with Diamond shaped filters. Here, 'maxflat' filters and 'dmaxflat7' are used for NSP and NSDFB respectively.

The availability of flexible number of directional subbands at each decomposition level makes NSCT as a unique multiscale, multi directional analysis transform. In frequency domain, DFB combines the neighborhood points for constructing contour using the obtained directional information at each level which results in detection of contours of an image. Hence, the use of NSP and NSDFB has made the NSCT as shift in-variant transform in addition to its multi directional, multi scale and over complete characteristics. Thus, NSCT provides better geometric representation for image compared to wavelet and Contourlet Transform. In this paper, illumination invariant geometrical feature representation for the frames in video for the purpose of shot transition detection is formulated

4. PROPOSED METHODOLOGY

Being an over complete, shift in-variant, multiscale and multi directional transform, NSCT can efficiently preserve the geometrical features of frames. Extracting these features in illumination normalized domain, the illumination effect can be reduced. Thus, the geometrical feature extracted from the illumination normalized domain is the illumination invariant feature for shot transit detection. Normalized Color Model.

A video can be represented as $V_j = (f_1, f_2 \cdots\ f_k \cdots f_K)$

where f_k is the k-th frame, $1 \leq k \leq K$. The normalized [7] RGB color model of the frame 'f_k' is given by,

$$\tilde{f}_k = \{\tilde{f}_{Rk}\ \tilde{f}_{Gk}\ \tilde{f}_{Bk}\} \tag{3}$$

Where,

$$\begin{pmatrix} \tilde{f}_{Rk} \\ \tilde{f}_{Gk} \\ \tilde{f}_{Bk} \end{pmatrix} = \begin{pmatrix} \dfrac{f_{Rk}}{f_{Rk} + f_{Gk} + f_{Bk}} \\ \dfrac{f_{Gk}}{f_{Rk} + f_{Gk} + f_{Bk}} \\ \dfrac{f_{Bk}}{f_{Rk} + f_{Gk} + f_{Bk}} \end{pmatrix}$$

In normalized RGB model, chromaticity components $\tilde{f}_{Rk}$ and $\tilde{f}_{Gk}$ describe the color information in the image. Owing to normalization, $\{\tilde{f}_{Rk}\ \ \tilde{f}_{Gk}\ \tilde{f}_{Bk}\}$ are robust to light intensity changes, shading and shadows.

5. NSCT PARMETER SELECTION AND FEATURE VECTOR COMPUTATION

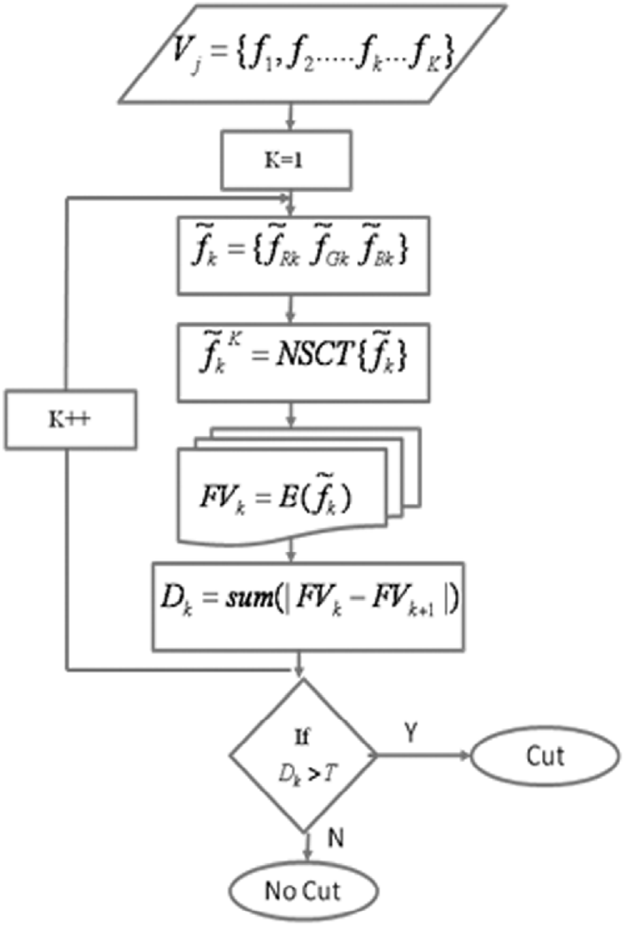
The illumination invariant and geometric feature for the framecan be computed as illustrated in fig. 3. The NSCT decomposition of, involves several parameters like number of scales, number of directions at each scale. In this work, the decomposition parameter chosen for NSCT is [4 3 3]; which provides the directional subbands of [2^4 2^3 2^3] = [16 8 8]. Along with the lower frequency subband, the total number of subbands for each frame is +16+8+8=33 sub bands. From each of these 33 sub bands in $\tilde{f}_k^{\ N}$, energy feature is extracted using,

$$E_k^{\ l}(\tilde{f}_k^{\ N}) = \frac{1}{M * N} \sum_{i,j}^{M,N} |\ \tilde{f}_k^{\ N}(i, j)\,| \tag{5}$$

Where $E_k^{\ l}$ is the energy of lth subband in kth frame, M is the number of rows and N is the number of columns in the kth frame.

Hence the feature vector obtained from the NSCT can be written as,

$$\begin{pmatrix} FV_1 \\ FV_2 \\ \cdots \\ \cdots\cdot \\ FV_K \end{pmatrix} = \begin{pmatrix} E_1^{\ 1}\ E_1^{\ 2} \ldots\ldots\ldots\ldots E_1^{\ 33} \\ E_1^{\ 1}\ E_2^{\ 2} \ldots\ldots\ldots\ldots E_2^{\ 33} \\ \cdots\ \ \cdots\cdot\ \ \ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ E_k^{\ 1}\ E_k^{\ 2} \ldots\ldots\ldots\ldots E_k^{\ 33} \end{pmatrix} \tag{6}$$


$V_j = \{f_1, f_2 \ldots f_k \ldots f_K\}$
K=1
$\tilde{f}_k = \{\tilde{f}_{Rk}\ \tilde{f}_{Gk}\ \tilde{f}_{Bk}\}$
$\tilde{f}_k^{\,K} = NSCT\{\tilde{f}_k\}$
K++
$FV_k = E(\tilde{f}_k)$
$D_k = sum(|FV_k - FV_{k+1}|)$
If
$D_k > T$
Y
Cut
N
No Cut

6. VISUAL CONTENT DIFFRENCE ESTIMATION AND BOUNDARY DETECTION

The next step in video shot boundary detection is computation of similarity/dissimilarity measure between consecutive frames in video. The dissimilarity value between the consecutive frames is computed using sum of absolute difference between the features of the frames as given below,

$$D(k,k+1) = \sum_k |FV_k - FV_{k+1}| \tag{7}$$

7. THRESHOLD SELECTION

Selecting suitable threshold is a vital issue in shot boundary detection algorithm. A global threshold approach is used and it depends on mean and standard deviation of the Dissimilarity values. The threshold value is given by,

$$T = \mu + \alpha \times \sigma \tag{8}$$

Where μ is the mean of the dissimilarity values.

σ is the standard deviation of the dissimilarity values.

α is the weighting constant.

8. EXPERIMENTAL ANALYSIS

This section demonstrates the NSCT based shot boundary detection on videos. A famous Tamil movie song ‘Ninukori varanam’ having highly varying lighting effects from the feature film ‘Akni Natchathram’ is chosen as test data. The test data video runs for 258 seconds and comprises of 7736 frames of size 320 × 480 pixels. This video has drastically varying in lighting conditions to match the fast rhythm of the music and few of the frames are shown in fig. 4.

Most of the existing shot boundary detection algorithms identify the illumination variation in the frame as cut transition in the test video data. The result of one such algorithm, shot boundary detection using orthogonal vectors [8] is shown in fig. 5 which illustrates more wrong detection of varying lighting effects as shots. To eliminate this drawback, the proposed algorithm is designed based on NSCT features for better abrupt cut detection. For experimentation, the test video data is divided into five segments for analyzing the shot boundary. The ground truth data collected manually in these five segments are shown in Table 1. Precision and recall metrics are used for evaluation of the proposed methodology. The precision and recall are defined as

$$\mathrm{Re}call\ \% = \frac{N_c}{N_c + N_m} \times 100 \tag{9}$$

$$\mathrm{Pr}ecision\ \% = \frac{N_c}{N_c + N_f} \times 100 \tag{10}$$

Where N_c is the number of correct alarms, N_m is the number of missed alarms and N_f is the number of false alarms. The following Table II shows the precision and recall values for Test segment 1 and test segment 2.

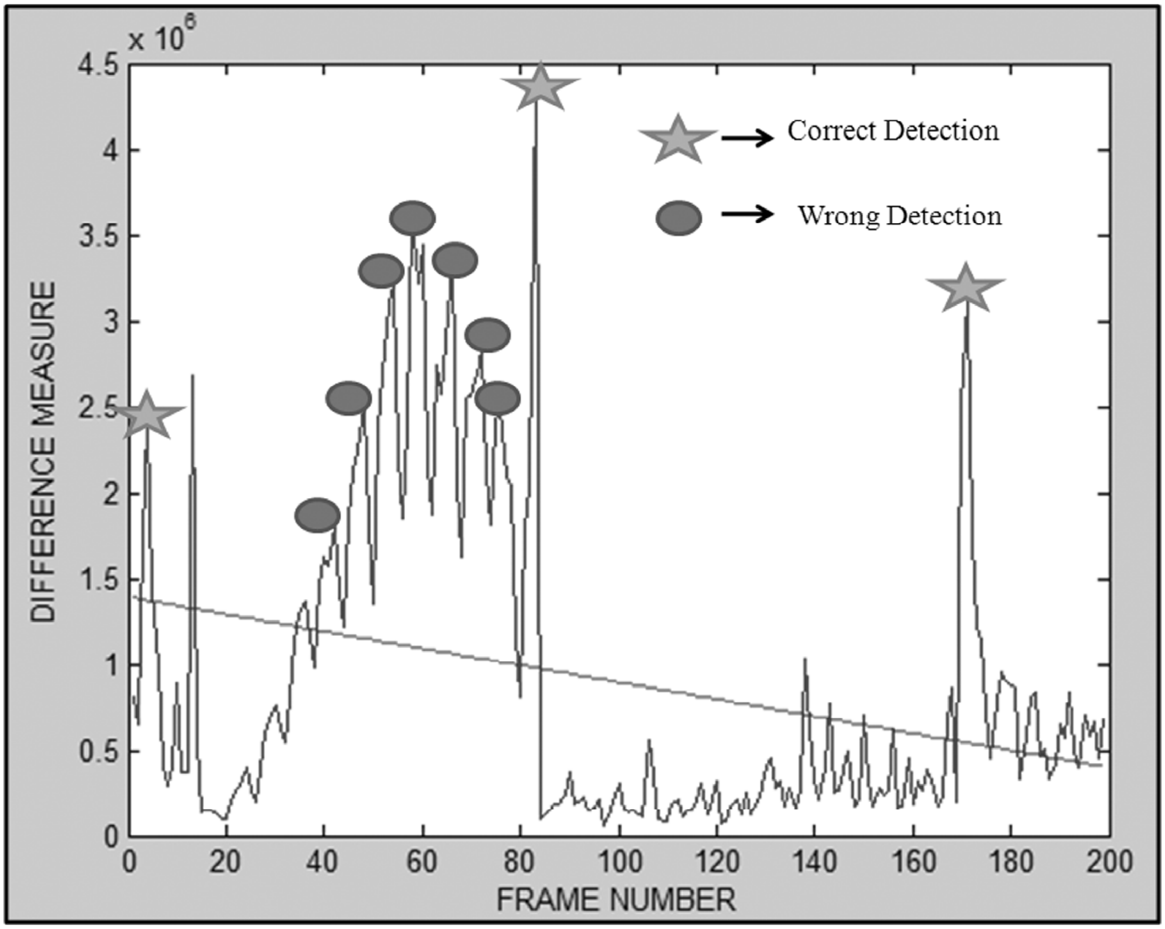


Figure 4: Shot boundary detection using orthogonal vectors [8]

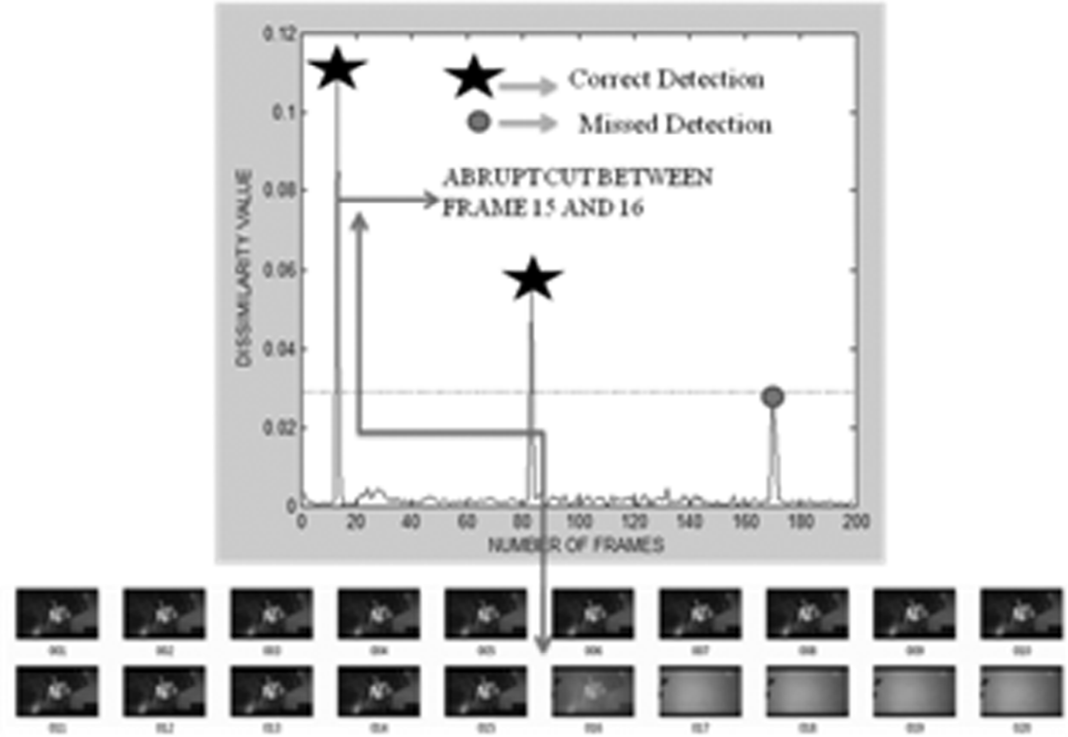


Figure 5: Cut detection analysis on segment 1

Table 1
Test Video Segments

Test Data	*Duration(s)*	*Frames*	*Abrupt cuts*
Segment 1	60	1799	12
Segment 2	56	1679	6
Segment 3	60	1799	10
Segment 4	52	1559	20
Segment 5	30	900	Nil
Total	258	7736	48

Table 2
Performance Evaluation

	Proposed method		*OrthogonalVectors [8]*	
Test Segment	*Precision*	*Recall*	*Precision*	*Recall*
Segment1	100%	66.66%	85%	60.34%
Segment2	70%	100%	70%	90%
Average	85%	83%	77%	75.17%

9. CONCLUSION

In this paper, an illumination invariant shot boundary detection based on NSCT has been proposed. The color channels are normalized to avoid mislead of illumination changes in cut detection of the test video. The NSCT coefficients from normalized domain are used to form the geometrical feature vector which supports the efficiency of the NSCT based shot boundary detection methodology. The performance of the proposed algorithm provides satisfactory results on highly challenging video segments. In future, the focus is to use NSCT based feature for detecting gradual shot boundaries under varying illumination and explosions in videos.

REFERENCES

[1] Gargi, Kasturi, and S.H. Strayer, “Performance Characterization of Video Shot Change Detection methods” IEEE Transactions on Circuits and Systems for Video Technology, CSVT-10(1), pp. 1-13, 2000.

[2] Jordi Mas and Gabriel Fernandez, “Video shot boundary detection based on color histogram”, TRECVID Workshop 2003.

[3] R. Zabih, J. Miller and K. Mai, “A feature-based algorithm for detecting and classifying production effects”, Multimedia System 7(2), pp.119 128., 1999.

[4] Hun-Woo Yoo, Han-Jin Ryoo and Dong-Sik Jang, Gradual shot boundary detection using localized edge blocks, Multimedia tools applications, vol. 28, pp.283-300, 2006 .

[5] Don Adjeroh, M. C. Lee, Banda, N. and Uma Kandaswamy, “Adaptive Edge-Oriented Shot Boundary Detection”, EURASIP, Journal on Image and Video Processing, 2009.

[6] Arthur L. da Cunha, Jianping Zhou, Minh N Do, “The Non Subsampled Contourlet Transform: Theory, Design and Applications”, IEEE Transactions on Image processing, 2006.

[7] J. Van de Weijer, C. Schmid, J.J. Verbeek, D. Larlus, “Learning colornames for real-world applications”, TIP 18, pp. 1512–1524, 2009.

[8] G.G. Lakshmi Priya, S. Domnic, “Edge Strength Extraction using Orthogonal Vectors for Shot Boundary Detection”, Procedia Technology, Vol. 6, pp. 247-254, 2012.