A Gradual Shot Change Detection using Combination of Luminance and Motion Features for Frame Rate Up Conversion

Sangchul Kim
Dept. of Computer Science and Engineering
Sogang University
Seoul, Republic of Korea
sckim@sogang.ac.kr

Hotak Hong
Dept. of Computer Science and Engineering
Sogang University
Seoul, Republic of Korea
hotakhong@sogang.ac.kr

Jongho Nang
Dept. of Computer Science and Engineering
Sogang University
Seoul, Republic of Korea
jhnang@sogang.ac.kr

***Abstract*— Frame rate up conversion (FRUC) is a method that automatically generates and inserts virtual frames between input frames for smoother video playback. When a simple FRUC is applied to a video clip, the image quality around shot boundaries is severely deteriorated. Therefore, shot boundary detection (SBD) is essential for ensuring the quality of video clips. Detection of gradual shot transition still remains a challenging issue. In this paper, we propose a gradual shot transition detection method that uses both an SBD based on motion-detection using multiple frames and an SBD based on luminance. Luminance-based SBD is robust to dynamic motions and sensitive to light effect; in contrast, motion-based SBD is sensitive to dynamic motions and robust to luminance effects. Both SBD methods are performed independently at first. In multi-frame luminance-based SBD, each algorithm is performed and the results are validated to detect shot boundaries by using distribution of luminance change. Threshold is adaptively changed depending on the borders of frame shots and gradual shot change is identified with multi-frame motion-based gradual shot transition detection. Our proposed method of shot transition boundary detection showed 97.36% recall and 92.61% precision on average in our experiment.**

Keywords-Shot Boundary Detection; Gradual Change Detection; Frame Rate Up Conversion

I. Introduction

Frame rate up conversion (FRUC) makes video playback smoother by inserting artificial interleave frames between source frames [1][2]. A shot is the unit of a video clip and shot boundary detection (SBD) algorithm needs to be performed to analyze the content of video sequences.

If two frames belonging to different shots are blended, an unnatural-looking frame is generated during FRUC, hence SBD is essential to FRUC. One outstanding challenge to SBD algorithms is detection of gradual shot transition, such as dissolve, fade in/out, wipe, and so on [3]. Histogram-based methods show solid performance for blunt shot cuts, but not for gradual shot transitions because brightness is changed smoothly with them [4][5][6]. This aspect was addressed in [7][8] with motion-based algorithms at expensive computational costs. With their methods, accuracy of shot boundary detection degenerated for shots with dynamic motions and when brightness changed minutely during a smooth shot transition with long duration.

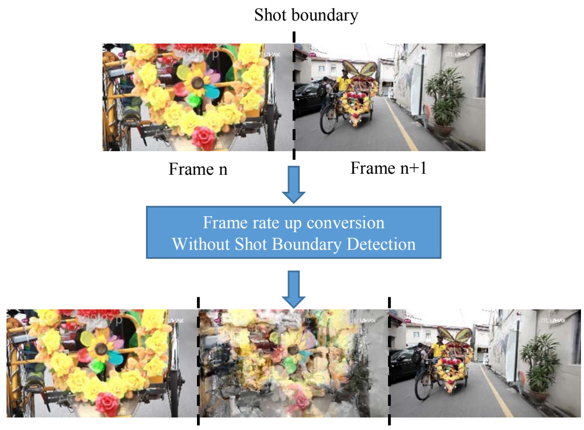

Figure 1.An Example of FRUC at Shot Boundary

Figure 2 illustrates the difficulty in shot boundary detection. FRUC performs motion estimation and it is easy to apply a motion-based SBD algorithm. In this paper, we propose a motion-based gradual shot transition detection method that prevents frame quality deterioration.

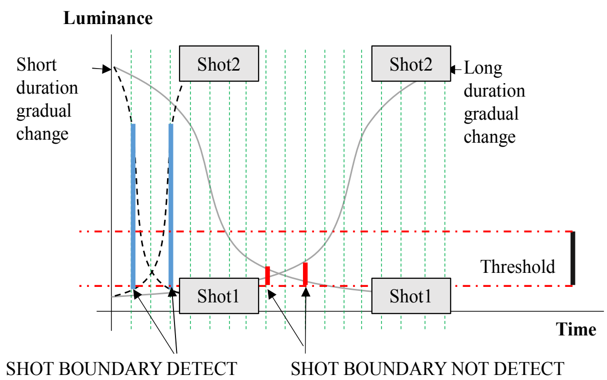

Figure 2. An illustration of reason why shot boundary detection is difficult when gradual change occur

The proposed algorithm performs motion-based gradual shot transition and uses luminance-based SBD to compensate for errors that frequently occur for dynamic shots. In a gradual shot transition, a sequence of images change smoothly, hence multiple frames are used for verification. The proposed method showed 97.36% recall and 92.61% precision on average in our experiment. For

Gradual change detection (GCD) in FRUC, recall is very important, hence the result of the proposed method is an important contribution. This paper is organized as follows. Section 2 discusses related work. Section 3 presents the proposed method, followed by experiment and analysis in Section IV. Section V concludes the paper with suggestions for future work.

II. Related Works

A. Motion Compenstation Frame Rate Up Conversion

FRUC inserts artificially generated interleave frames between source frames for smoother video playback. Motion-compensated FRUC is used in most research work [2]. Figure 3 shows how MC-FRUC estimates motion vectors for each block for two adjacent frames. MC-FRUC generates interleave frames by placing pixels in source and reference frames according to the identified motion vectors. Latest studies on MC-FRUC seek to improve the accuracy of estimated motion vectors for improved image quality, but the topic of detecting gradual shot transition has been remained largely unexplored. In [9], multi-block histogram-based shot-boundary detection algorithm was proposed for FRUC. However, generating histograms for multiple blocks for high-resolution images is computationally expensive and it has not been verified whether it is effective for detecting gradual shot transitions.

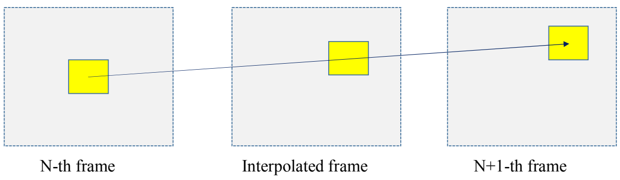

Figure 3. An example of motion compensation frame rate up conversion

B. Shot Boundary Detection

In an SBD algorithm, shot boundaries are detected by identifying how unrelated two adjacent frames are. This is done by extracting features from frames; if the features of two adjacent frames are substantially different, then a shot boundary is declared. In [6][7], the threshold value is adaptively adjusted depending on the nature of the frames for improved accuracy. The state-of-the-art methods produce almost perfect results, however, gradual shot transition has remained unaddressed so far.

C. Global Histogram based Gradual Change Detection

In global histogram-based gradual change detection, patterns found in gradual changes have been captured in histogram models; if similar histogram patterns are detected, then the frames are declared as a gradual change. Models for gradual change detection have been suggested as in [4][5], but they fitted only for a specific case of dissolve transition, so the scope it covered was not sufficiently large.

D. Motion based Shot Boundary Detection

Distribution of pixel-based features can be widespread for certain motions of object or camera. It causes for the detector to inaccurately detect a shot. In [7][8], this issue was addressed by comparing histograms of regions before and after motion by following estimated motion vectors. If the estimated motion vectors are accurate, the result has a superior quality over global histogram-based methods; but motion vector estimation is usually inaccurate so that the produced quality is not as good as when a global histogram-based method is used. Furthermore, accurate motion-based estimation is computationally very expensive.

III. Proposed Gradual Change Detection

Figure 4 illustrates the flowchart of the proposed algorithm. In the algorithm, two adjacent frames are given as input and their motion feature and pixel-based global feature are extracted. SBD is performed for each feature; if shot boundary is detected as 'miss,' the border between gradual change and shot are detected by using an algorithm that compensates for each feature.

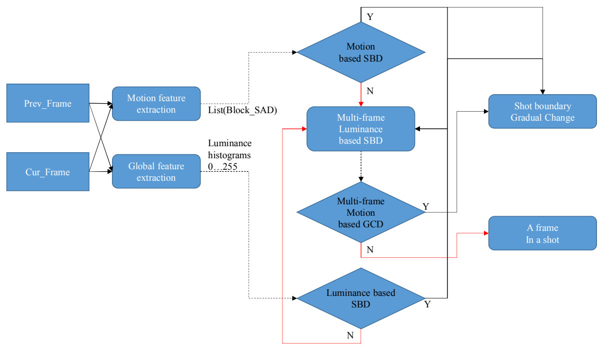

Figure 4. Flow chart of proposed gradual change detection algorithm.

A. Luminance Based Shot Boundary Detection

If estimated motions are inaccurate or motions are dynamic in a motion-based SBD method, frames can be incorrectly detected as shot boundaries. False alarm rate increases in such a case. The image quality of resultant frame sequence can deteriorate because if FRUC is applied in this case because shot boundaries are declared incorrectly. To correct this, similarity of luminance histogram is referenced for shot boundary detection. Equation (1) is used for finding luminance histogram. The number of luminance values between 0 and 255 is counted and it is normalized for the frame size.

$$\begin{aligned} &hist_{frame}(l) = n(\mathrm{Pixel}(x,y)) / (\mathrm{width} * \mathrm{height}) \\ &Pixel(x,y) = 1, \text{if } 0 \le x < \mathrm{width}, 0 \le y < \mathrm{height} \text{ and } 0 \le l < 256 \end{aligned} \quad (1)$$

To determine the similarity between two frames, we used xi-square validation, which is commonly used for finding out similarity of two histograms. Equation (2) is a xi-square goodness-fit equation; g is substituted with $hist_{prev\_frame}$ of the previous frame is and h with $hist_{cur\_frame}$ of the current frame.

$$\chi^2(g,h) = \frac{1}{2}\sum_i \frac{(g(i)-h(i))^2}{g(i)+h(i)} \quad (2)$$

Equation (3) determines whether to declare a shot boundary; if the predefined threshold ω_l of luminance histogram is exceeded, then the two adjacent frames are determined to be unlikely to be related, so $SB_l(\omega_l)$, the shot boundary from luminance-based detector, is set true.

$$\begin{aligned} &SB_l(\omega_l) = \text{true} \\ &, if\ \chi^2(hist_{prev\_frame}, hist_{cur\_frame}) > \omega_l \end{aligned} \quad (3)$$

B. Motion based Shot Boundary Detection

Larger number of similar blocks as determined in motion estimation implies that the two input frames are highly related. Recreating histogram for each block incurs extra computational cost that varies depending on frame resolution and block size, we recycle the results from motion estimation used in FRUC and find gradual change. Equation (4) determines similarity between blocks determined during motion estimation.

$$\begin{aligned} &\min SAD(i,j) = \\ &\sum_{(x,y)\in block_{i,j}} |prev\_frame(x,y) - cur\_frame(x+vx, y+vy)| \quad (4) \\ &simBlockRatio(\omega_m) = n(\min SAD(i,j) < \omega_m) / \text{n}(\forall block) \end{aligned}$$

The term $\min SAD(i,j)$ is the Sum of Absolute Difference (SAD) value to determine the motion vector (vx, vy). To determine similarity of the whole frame, $simBlockRatio$ is used, and this is obtained by counting the blocks that exceed threshold ω_m followed by normalization with $\text{n}(\forall block_{i,j})$, which is the total number of blocks. Equation (5) is used in determining shot boundaries; if $simBlockRatio(\omega)$ as calculated in Equation (4) exceeds a specific value α, then it is assumed that the two frames are dissimilar, and as a result, the shot boundary from motion-based detector, $SB_m(simBlockRatio(\omega_m), \alpha)$, is set true.

$$\begin{aligned} &SB_m(simBlockRatio(\omega_m), \alpha) = \text{true} \\ &if\ simBlockRatio(\omega_m) > \alpha \end{aligned} \quad (5)$$

C. Multiframe based Luminance Shot Boundary Detection

When brightness changes significantly in a shot, inter-frame luminance fluctuates greatly and both luminance-based or motion-based SBDs could easily false-alarm shot boundaries. If any one of $SB_l(\omega_l)$, $SB_m(\omega_m)$ is false, then the frame needs to be inspected again for shot boundary. Such shots generally tend to have wide-spreading luminance change. To verify this hypothesis, we detected shot boundaries $SB_{ml}(\dot{\chi}^2)$ as shown in Equation (6) for contiguous frames with the size of window and using calculating distribution of $\chi_i^2(g,h)$.

$$\begin{aligned} &E(\dot{\chi}_c^2) = \Big(\sum_{i=-window+cur}^{cur+window} \chi_i^2(g,h)\Big) / (2 * window) \\ &\text{var}(\dot{\chi}_c^2) = \sqrt{\frac{\sum_{i=-window+cur}^{cur+window} (E(\dot{\chi}_c^2) - \chi_i^2(g,h))^2}{2 * window}} \\ &SB_{ml}(\dot{\chi}_c^2) = \text{true, if var}(\dot{\chi}_c^2) < \omega_{ml} \end{aligned} \quad (6)$$

D. Multi Frame based Gradual Change Detection

As explained in Section I, the reason why gradual change detection is challenging is that threshold value has to adaptively adjust over a temporal window during a gradual change. When multiple shot boundaries are declared in frames within a predefined window, then it is likely that the there is a gradual change within the window.

$$\begin{aligned} &prev\_cnt(SB_{ml}(\dot{\chi}_c^2), window) = \sum_{i=-window}^{0} k, \\ &k = 1, if\ SBD(c+i-1, c+i) = SB_{ml}(\dot{\chi}_c^2)\ else\ k = 0 \end{aligned} \quad (7)$$

Equation (7) determines the distribution of shots as inspected by the method in Section III-C. The number of shot boundaries in the preceding and following frames is determined from the current frame. A higher number of shot boundaries implies that a gradual change has occurred. Hence, it is necessary to decrease ω_m as the number increases. To lower ω_m, weight is recalculated using Equation (8), a sigmoid function based on the number of shot boundaries is used to adjust weight to threshold.

```
function
GradualChangeDetect( ω_l, ω_m, α, frame_c, frame_c+1 ) {
  if( SB_l(ω_l) ==true&& SB_m(simBlockRatio(ω_m),α) )
  {
    //detected as a shot by both SBD
    //it is clearly shot boundary
    Return true;
  } else if( SB_l(ω_l) ==true){
    //only detected by luminance based SBD
    //do multi frame based Gradual Change Detector
      return SB_m(simBlockRatio(ω̇_m),α);
  } else if( SB_m(simBlockRatio(ω_m),α) ){
    //only detected by motion based SBD
    if( SB_ml(χ̇_c^2) ) {
        return true;
    } else
        return false;
  } else {
  if( SB_ml(χ̇_c^2) ) {
        return SB_m(simBlockRatio(ω̇_m),α);
    } else
        return false;
  }
}
```

Figure 5. The Pseudo Code of Proposed Gradual Change Detection Algorithm

The range of $\frac{prev\_cnt(SB_{ml}(\dot{\chi}_c^{2}), window)}{\text{window}}$ is between 0.0 and 1.0, so the range of the sigmoid function is also set at 0.0 and 1.0; the exponent of e is set at 6x instead of -x so that the the function yields a decreasing result as the number of shot boundary increases. The final threshold used for motion-based shot-boundary detection to declare a gradual change is $\dot{\omega}_m$. It has been designed to produce values between 0.0 and 1.0 with the sigmoid function. Figure 5 is the pseudo code for the proposed algorithm. A shot boundary is declared only when both motion-based and luminance-based decisions are positive. If only luminance-based decision is positive, then multi-frame motion-based shot boundary detection method is used.

$$norm\_sigmoid(x) = 1/1 + e^{6x}$$

$$\dot{\omega}_m = 0.5 + norm\_sigmoid\left(\frac{2*prev\_cnt(SB_{ml}(\dot{\chi}_c^{2}) - window)}{2*\text{window}}\right)*\omega_m \quad (8)$$

IV. Experimental Results

A. Experimental environment

To evaluate our proposed method, we tested several video clips that contain gradual changes. Table 1 charts details on the tested video clips: total frame counts and frame counts that contain gradual changes. We created video clip ____ for experiment, and video clips 2 and 3 were downloaded from YouTube. Hanatour is an advertisement video clip and it contains frames that alternate between night scenes and natural landscapes through dissolve effect. Avengers II Trailer has many fade-in/out effects. The frames are generally dark and contains dynamic motions. Imax_hd_test_japan has several dissolve and fade-in/out effects but the video clip is very static.

TABLE I. Our Test Sets

No	Video information		
	Video name	***# of total frames***	***# of gradual change frames***
1	Hanatour	690	70
2	Avengers II Trailer 0:00~1:17	1846	206
3	imax_hd_test_japan 0:00~1:00	1498	103

B. Analysis for our algorithm

Table 2 shows recall/precision for each video clip. The proposed algorithm showed a solid recall; Frame quality after FRUC severely deteriorates when SBD is false negative. The method proposed in this paper substantially reduces false negatives, hence recall is higher. Therefore, the proposed SBD algorithm is suitable for FRUC. Avengers II Trailer had the precision of 90.07 because luminance-based SBD and motion-based SBD incorrectly declared frames with dynamic lighting changes or dynamic motions (flashlights and chasing scenes) as shot boundaries. Sum of absolute difference (SAD) was used for motion features, which is very sensitive to changes in brightness, but high recall implies frame quality deterioration in FRUC would be substantially decreased.

TABLE II. Table of Recall/Precision of Our Algorithm

No	Video information		
	Video name	***Recall (%)***	***Precision (%)***
1	Hanatour	100	97.14
2	Avengers II Trailer 0:00~1:17	97.57	90.07
3	imax_hd_test_japan 0:00~1:00	96.77	93.20

C. Analysis for each modules

This section presents our performance analysis for each module in the algorithm. Table 3 charts the average recall/precision of video clips 1 to 3 in Table 1. L-SBD, M-SBD, Multi-frame L-SBD, and Multi-frame M-SBD are algorithms explained in Sections III-A, III-B, III-C, and III-D, respectively. L-SBD had lower recall because it recognizes the large differences in global feature as a shot boundary during a luminance effect. M-SBD had a significantly lower precision and this is attributable to the sharp increase in SAD when there are dynamic motions, which are difficult to estimate for motion vector. This is evident when a large object quickly appears or disappears.

Multi-frame M-SBD addressed these issues by adjusting thresholds adaptively during gradual changes, resulting in improved recall. This has a better precision value, so it could be used for improving the quality of frames generated in FRUC.

TABLE III. Table of Average Precision/Recall for each Modules

No	Video information		
	Applied Module Types	***Recall (%)***	***Precision (%)***
1	L-SBD	88.98	98.41
2	M-SBD	97.43	56.44
3	L-SBD + M-SBD + Multiframe L-SBD	93.37	94.19
4	L-SBD + M-SBD + Multiframe L-SBD +Multiframe M-SBD	97.36	92.61

V. Conclusion and Future Works

The method proposed in this paper recycles the motion estimation data calculated during FRUC. so it combines the benefits of M-SBD, which requires only a small additional computational resources while achieving high recall, and L-SBD, which generally has a high precision despite its weakness towards lighting effects. The proposed method has an increased recall since threshold is adaptively adjusted in multiple consecutive frames in order to detect gradual changes. A high recall for SBD greatly enhances FRUC quality. There are many types of thresholds and setting the initial threshold is important, so our future work will focus on automatically determining initial threshold.

ACKNOWLEDGMENT

This work was supported by ICT R&D program of MSIP/IITP. [R0126-15-1112, Development of Media Application Framework based on Multi-modality which enables Personal Media Reconstruction]

REFERENCES

[1] J. Nang, S. Kim, and H. Lee, "Classifying Useful Motion Vectors for Efficient Frame Rate Up Conversion of MC-DCT Encoded Video Streams," Journal of Information Science and Engineering, vol. 30, no.06, 2014, pp.1755-1771.

[2] S. Kim, D. Oh and J. Nang, "A New Path Based Interpolation using Object Motion for Frame Rate Up Conversion," (Accepted at in proc. of ICCE-Berlin 2015).

[3] J. Yuan, H. Wang and L. Xiao, "A Formal Study of Shot Boundary Detection", IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no.2, pp.168-186

[4] R. Lienhart, S. Clara, "Reliable Dissolve Detection," in proc. of SPIE 2001, pp.219-230.

[5] A. Hampapur, T. Weymouth, R. Jain, "Digital Video Segmentation," in proc. of MULTIMEDIA'94, pp. 357-364.

[6] C. Su, H. Liao, H. Tyan, K. Fan and L. Chen, "A Motion-tolerant dissolve detection algorithm," IEEE Transactions on Multimedia, vol.7, no.6, pp.1106-1113.

[7] A. Amel, B. Abdessalem and M. Abdellatif, "Video Shot Boundary Detection Using Motion Activity Descriptor," Journal of Telecommunications, vol.2, no.1, pp.54-59.

[8] A. Hameed, "A Novel Framework of Shot Boundary Detection for Uncompressed Videos," in proc. of ICET 2009, pp.274-279.