東海學報 48 卷(2009):27-36
中華民國九十八年十一月出版

Film-to-Video Conversion with Shot Cut Detection

Chishyan Liaw, Chung-chi Lin, and Ching-tsorng Tsai
Department of Computer Science, Tunghai University
{liaw, cclin, cttsai}@thu.edu.tw

Abstract

A film-to-video conversion technique with shot cut detection is proposed and its performances are examined. Shot cut happens fairly often in film broadcasting and it tends to destabilize the quality of performance while motion compensated interpolation technique is utilized. In the proposed method, the shot cut detection is executed at the beginning of motion compensation of frame-to-video conversion to ensure that the interpolated information can be used correctly. And it utilizes the low computational complexity of frame-to-video conversion technique to promote higher quality video sequences. The simulation results show that the proposed method exhibits better performances than other methods.

Keywords: *frame-to-video conversion, shot cut, 3:2 pulldown*

1 Introduction

The traditional NTSC system uses the interlaced scan technique to display video sequence. However, new devices like LCD displays, personal computer monitors, and HDTV require a progressive scan format. The NTSC system will still exist with new devices for a period of time before being replaced. The scan rate of TV is 60 fields per second, i.e. 30 frames per second; on the other hand, DVD player uses progressive scan technique with scan rate of 24 frames per second, 25 frames per second, 29 frames per second, etc. Therefore, displaying progressive scan images on the NTSC system is a problem before it phases out.

In order to watch movies or filmed programs on TV, the film-to-video conversion is necessary. There are numerous processing techniques for film-to-video conversion. The most common approach is 3:2 pulldown [1] which converts original progressive 24 frames/second to 60 fields/second, as shown in Fig. 1. It divides each even frame into two fields and decomposes every odd frame into three fields. Thus, every four frames are split into ten fields arranged in the order of odd field follows by even field. From Fig. 1, fields 1^o and 2^e are merged into a frame and similarly, fields 2^o and 3^e are combined into another frame. Video images will appear clawing effect and artifacts effect, as shown in Fig. 2 and Fig. 3, if there is fast movement or shot cut since fields are from different sources.

In order to reduce motion clawing effect and artifacts effect, we develop a motion compensated interpolation technique to obtain high quality results with low computation. Motion compensated interpolation [2-5] searches for two most similar macroblock in the two successive frames and calculates its motion vector (MV) to form a new frame. However, this approach is difficult to obtain a good result without reliable motion estimation.

All motion compensated interpolation methods focus on the strategy of handling the occurrence of motion but the shot cut which happens fairly often in video sequence is ignored. Nevertheless, shot cut tends to destabilize the quality of performance while motion compensated interpolation technique is utilized. Consequently, the artifacts effect appears in most cases. In order to have reliable motion estimation, shot cut detection is necessary. Moreover, the issue of shot cut needs to be addressed with the motion compensated interpolation process.

In general, shot cut in video sequences occurs abruptly and it is highly possible to retrieve wrong messages between two consecutive frames and produce artifacts effect by motion compensated interpolation process. This paper presents a motion compensated frame-interpolation technique to obtain high quality results without excessive computation. The technique includes shot cut detection and frame-interpolation with film-to-video conversion.

2 Related Works

2.1 Motion Compensation

Before motion compensated interpolation, motion estimation (ME), is based on the contents of two successive frames to find an adaptive motion vector which is used as the information of interpolation, as shown in Fig. 4. There are some motion estimation methods like full search, fixed pattern search, etc. Full search method has the best result but it has much calculation effort. In order to save the calculation time, fixed pattern search like three-step search [6], four-step search [7], diamond search [8], etc. were proposed. These methods save calculation time but the results are not optimized. To solve the problem of fixed pattern search, the searching method with dynamically adjusting searching range like adaptive rood pattern with zero motion vector (ARP-ZMV) [9] was proposed. It uses zero motion vector to reduce the area where needs estimation.

2.2 Motion Compensation

The most popular method of film-to-video conversion is 3:2 pulldown which may have clawing effect or artifacts effect after the process. Thus, Hilman [10] proposed a method that converts 24 frames/sec to 30 frames/sec then by interlaced scan, 30 frames/sec is broadcast in 60 fields/sec. This method eliminates clawing effect and artifacts effect and reduces motion jerkiness between frames. It uses motion compensated interpolation to interpolate a new frame and uses three-step search to estimate motion. The conversion from 24 frames/sec to 30 frames/sec is depicted in Fig. 5 where two new frames are created from four consecutive frames. But the scent cut is neglected when it is applied to improve the problems of 3:2 pulldown. Therefore, the information on two consecutive frames causes the serious block effect on interpolated frame.

3 Related Works

Scent cut is usually neglected when the strategy of film-to-video conversion is applied to improve the problems of 3:2 pulldown. Thus, the information on two consecutive frames causes distortion on interpolated frame. In our proposed method the first step of film-to-video conversion is shot cut detection, which is to ensure that the motion compensated interpolation can be used at correct position of frames. The proposed shot cut detection is depicted in Fig. 6. Frames 1 and 2 are used as shot cut detection. A new frame is generated by motion compensated interpolation according frames 1 and 2 if no shot cut is found, as shown in Fig. 6(a). Otherwise, frames 3 and 4 are used to create a new frame, as shown in Fig. 6(b).

3.1 Shot Cut Detection

Within the same shot, a scene change may cause a frame to differ from its consecutive frame due to one or more factors, e.g. shot cut, lighting change, focal length change, large object movement, and dissolving. To detect a scene change between two consecutive frames, a dissimilarity measure between two frames must be defined, which is mainly based on histogram-based method [11-13] and pixel-based method [13,14]. We explored several types of histogram-based algorithms and found that the dissimilarity measure, denoted as d_h, has better performance than other measures to define a scene change [12]. The following $d_h(F_n,F_{n-1})$ equation is used to detect the condition of scene change between two adjacent frames F_n and F_{n-1},

$$d_h(F_n,F_{n-1}) = \sum_{j=0}^{m} \begin{cases} \dfrac{(H_n(j)-H_{n-1}(j))^2}{\max(H_n(j),H_{n-1}(j))}, & \text{if } (H_n(j)\neq 0) \text{ or } (H_{n-1}(j)\neq 0) \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where $H_n(j)$ and $H_{n-1}(j)$ are the bin value of the histogram of the frame F_n and F_{n-1}, respectively, and m is the gray level of histogram. The value of $d_h(F_n,F_{n-1})$ will present a large value when scene changes. Afterward, a threshold, T_1, will be used to determine whether a scene change is involved or not.

Not all scene changes cause distortion on new interpolated frame but basically, shot cut will lead distortion. The contents do not change suddenly in other scene changes like lighting change, focal length change, large object movement, and dissolving so the images in two consecutive frames are similar and thus, distortion is not serious. The proposed method will detect shot cut precisely along with formula d_h and the method of evaluating dissimilarity pixels in successive frames.

The evaluating dissimilarity pixels method proceeds when scene change detected. Let $F_{n-1}(x,y)$, $F_n(x,y)$ and $F_{n+1}(x,y)$ denote as the previous, current, and the next frame, respectively, where the two-dimensional spatial indices (x,y) are x=1, 2, …, W and y=1, 2, …, H, and W and H are the width and height of the frame, respectively. The absolute pixel difference value (PD) of two adjacent frames is obtained from the absolute difference value between F_{n-1} and F_n frames. Thus, PD is defined as

$$PD(x, y) = \left| F_n(x, y) - F_{n-1}(x, y) \right|. \qquad (2)$$

$PD(x,y)$ is used to define the dissimilarity pixel. The dissimilarity pixels statistics value (DPS) is obtained by

$$DPS = \begin{cases} DPS + 1 & \text{if } PD(x, y) > Th \\ DPS + 0 & \text{otherwise,} \end{cases} \qquad (3)$$

where a threshold parameter Th needs to be set in advance. The value of DPS will present a value great than half of total pixels of a frame if two frames are not similar.

Eqs. (1) and (3) are used to detect the condition of shot cut between two adjacent frames F_{n-1} and F_n. The value of $d_h(F_n,F_{n-1})$ will present a large value when scene changes and the percentage of dissimilarity pixels of two adjacent frames will be great than 50% if shot cut happens in video sequences, as shown in Fig. 7.

3.2 Shot Cut Detection

The proposed motion compensated interpolation method interpolates a new frame in every four frames by motion compensated interpolation. The first step of motion compensated interpolation method is motion estimation, which is based on two adjacent frames to find a proper macroblock motion vector. According to the contents of two consecutive frames, motion estimation proceeds to find a proper motion vector. The proposed method presents a dynamical searching range that uses the predicted MV, MV of leftblock, and four search points of adaptive rood pattern with zero motion prejudgment (ARP-ZMP). The algorithm takes static vector into account and refers the motion vector calculated before to estimate the possible motion vector. Thus, it decreases the calculation effort and adjusts the position in searching similar macroblocks to find better image quality. Static vector, zero motion, is recognized when the value of the sum of absolute difference (SAD) is less than a threshold value T_2. That is, the macroblock of current frame is similar to the corresponding position of reference frame. Thus, the information of that macroblock can be interpolated on the corresponding position of the new frame. If the SAD value of minimal matching error (MME) larger than T_2, then ARP motion estimation proceeds. The algorithm of ARP-ZMP is

Step 1: Compute the matching error (SAD_{center}) between the current macroblock and the macroblock at the same position in the reference frame (i.e., the center of the current search window).

if $SAD_{center} < T_2$

    $MV_{target} = [0\ 0]$,

    Stop;

else

    if the current macroblock is a leftmost boundary macroblock,

        $Range = 4$;

    else

        $Range = \max\{|MV_{LeftBlock}(x)|, |MV_{LeftBlock}(y)|$

        Go to Step 2.

Step 2: Align the center of ARP with the center point of the search window and check its four search points and the position of the predicted MV to find out the current *MME* point.

if $SAD_{MME} < T_2$

$MV_{target} = [MV_{MME}(x)\ MV_{MME}(y)]$,

Stop;

else

Go to Step 3.

Step 3: Set the center point of the unit-size rood pattern (URP) at the *MME* point found in the previous step and check its points. If the new *MME* point is not incurred at the center of the current URP, repeat this step; otherwise, the MV is found.

4 Simulation Results

The proposed algorithm is coded in C++ and executed on personal computer. To evaluate the performance of film-to-video conversion with shot cut detection, three threshold values used in the simulations are set empirically to: $T_1 = 10000$, $T_2 = 512$, and $Th = 15$. The test video sequences are illustrated in Table 1. The performances of our proposed algorithm with 3:2 pulldown and Hilman [10] are analyzed. The results of test sequences indicate that the proposed method exhibits a better performance than other methods in various video sequences. Fig. 8 and Fig. 9 demonstrate that our proposed method with shot cut detection presents a more pleasing visual quality. For example, artifacts effect and block effect are eliminated in Fig. 8(d); artifacts effect is eliminated and line crawling is reduced in Fig. 9(d).

5 Conclusion

In this paper, we proposed a new shot cut detection method for frame-to-video conversion. The occurrences of shot cut affect the quality of frame-to-video conversion seriously if they are not processed properly. The proposed shot cut detection scheme is executed at the first stage of motion compensation of frame-to-video conversion to ensure that the interpolated information can be used correctly. The results of our experiments show that our method reduces motion clawing effect and artifacts effect and the quality of pictures can be improved by shot cut detection in frame-to-video conversion technique.

Reference

[1] Lagendijk, R.L and Sezan, M.I.
1992 "Motion compensated frame rate conversion of motion pictures," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 453-459.

[2] Zhai, J., Yu, K., Li, J., and Li, S.
2005 "A low complexity motion compensated frame interpolation method," *IEEE International Symposium on Circuits and Systems, ISCAS 2005*, vol. 5, pp. 4927 – 4930.

[3] Ha, T., Lee, S., and Kim, J.
2004 "Motion compensated frame interpolation by new block-based motion estimation algorithm," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 2, pp. 752 – 759.

[4] Lee, S.H., Kwon, O., and Park, R.H.
2003 "Weighted-adaptive motion-compensated frame rate up-conversion," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 3, pp. 485 – 492.

[5] Al-Mualla, M.E.
2003 "Motion field interpolation for frame rate conversion," *International Symposium on Circuits and Systems, ISCAS '03*, vol. 2, pp. 652-655.

[6] Koga, T., Iinuma, K., Iijima, A., and Ishiguro, T.
1981 "Motion-compensated interframe coding for video conferencing," in *Proc. NTC81*, New Orleans, LA, pp. C9.6.1-9.6.5.

[7] Po, L.M. and Ma, W.C.
1996 "A novel four-step search algorithm for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 313-317.

[8] Tham, J.Y., Ranganath, S., Ranganath, M., and Kassim, A.A.
1998 "A Novel Unrestricted Center-Biased Diamond Search Algorithm for Block Motion Estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 369-377.

[9] Nie, Y. and Ma, K.K.
2002 "Adaptive Rood Pattern Search for Fast

Block-Matching Motion Estimation," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1442 - 1449.

[10] Hilman, K., Park, H.W., and Kim, Y. 2000 "Using Motion-Compensated Frame-Rate Conversion for the Correction of 3:2 Pulldown Artifacts in Video Sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 6, pp. 869-877.

[11] Nagasaka A. and Tanaka, Y. 1992 "Automatic video indexing and full-video search for object appearances," *Visual Database Syst., II*, pp. 113-127.

[12] Sethi, I.K. and Patel, N. 1995 "A statistical approach to scene change detection," *SPIE*, vol. 2420, pp. 329–338.

[13] Huang, C.L. and Liao, B.Y. 2001 "A robust scene-change detection method for video segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.11, no. 12, pp. 1281-1288.

[14] Yi, X. and Ling, N. 2005 "Fast Pixel-Based Video Scene Change Detection," *IEEE International Symposium on Circuits and Systems, ISCAS* 2005, pp. 3443 – 3446.

Table 1. The test video sequences

Video Sequences	Frame size	No. of frames
TableTennis	320x240	599
Tennis	352x288	90
The Matrix	320x240	300
World Rally Championship part1	320x240	884
World Rally Championship part2	320x240	908

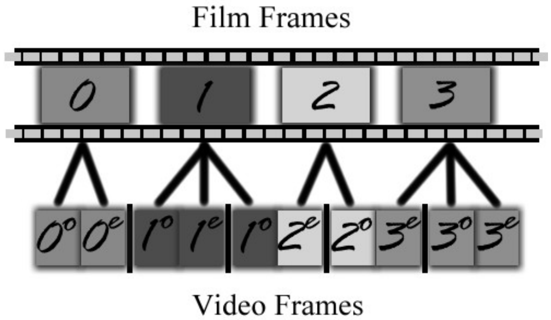


Fig. 1 Decomposition of source frames in 3:2 Pulldown.

(a) (b) (c)

Fig. 2 Table Tennis (a) frame #117 (b) frame #118 (c) result of 3:2 pulldown - clawing effect.

(a) (b) (c)

Fig. 3 Table Tennis (a) frame #261 (b) frame #262 (c) result of 3:2 pulldown - artifacts effect.

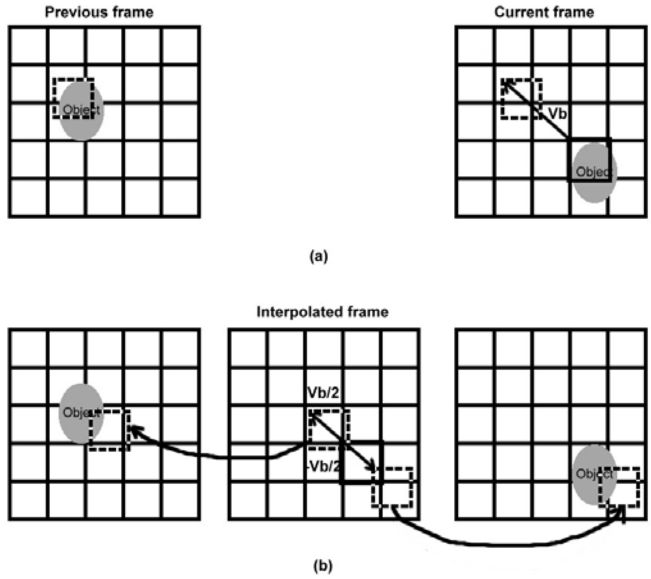


Fig. 4 Motion compensation (a) candidate motion vector, V_b, obtained by motion estimation (b) interpolation using the corresponding block from previous frame.

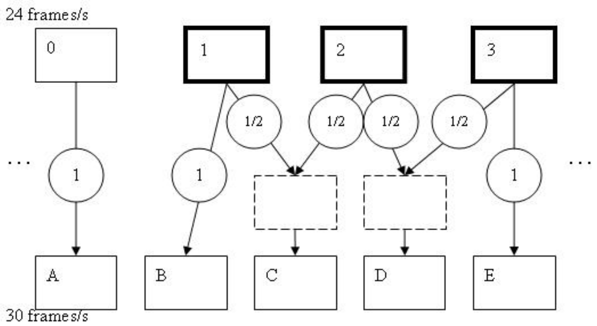
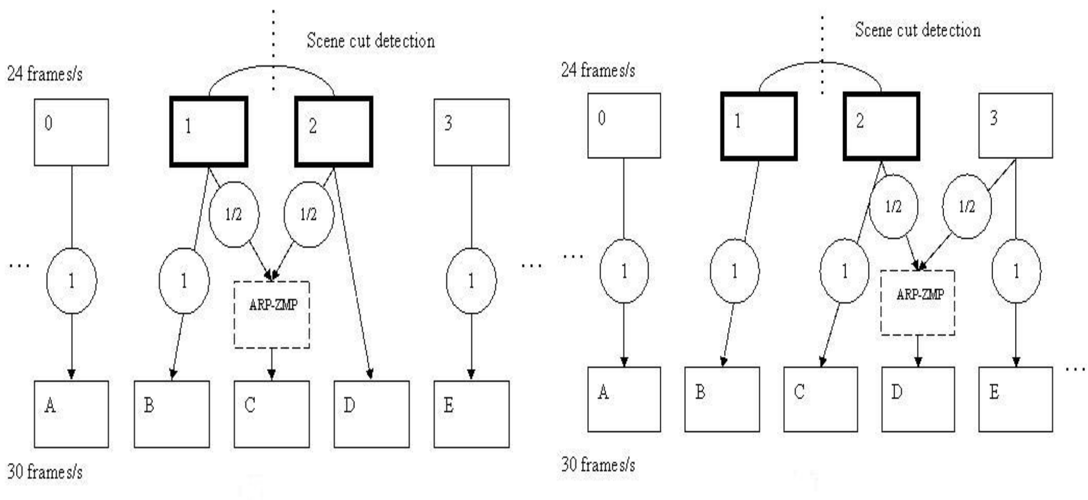
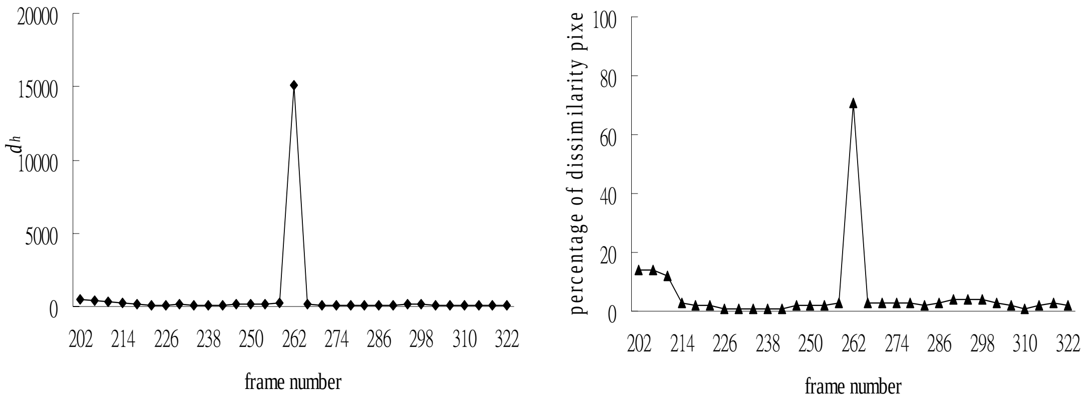


Fig. 5 The method of 24-to-30 fps conversion of Hilman [10].



(a) (b)

Fig. 6 Proposed frame-to-video conversion method (a) motion compensation with shot cut non-detected (b) motion compensation with shot cut detected.



(a) (b)

Fig. 7 Shot cut detection (a) value of d_h (b) percentage of dissimilarity pixel.

(a)

(b)

(c)

(d)

Fig. 8 Table Tennis (a) original frame #260 to frame #263 (b) result of 3:2 pulldown (c) result of Hilman [10] (d) result of proposed method.

(a)

(b)

(c)

(d)

Fig. 9 World Rally Championship part1 (a) original frame #212 to frame #215 (b) result of 3:2 pulldown (c) result of Hilman [10] (d) result of proposed method.

Film-to-Video Conversion with Shot Cut Detection

廖啓賢　林正基　蔡清欉
東海大學資訊工程學系

摘　要

場景切換(Shot Cut)在電視影片中是經常發生之現象，本論文針對數位視訊播放系統提出一套包含場景切換(Shot Cut)偵測的電影頻率轉成電視影片頻率(film-to-video)的插補技術。事先偵測是否有場景切換，再進行動作補償插補出新的frame以避免錯誤的Artifacts error失真結果。而本方法的計算量低，而能得到比其他方法更好的實驗結果。