# CODING OF FADED SCENE TRANSITIONS

*Dong Tian*[1], *Miska M. Hannuksela*[2], *Ye-Kui Wang*[1], *and Moncef Gabbouj*[3]

[1]Tampere International Center for Signal Processing (TICSP), Tampere, Finland
[2]Nokia Mobile Software, Tampere, Finland
[3]Tampere University of Technology, Finland

## ABSTRACT

Coding of a scene transition is often a challenging problem, from compression efficiency point of view, because motion compensation may not be a powerful enough method to represent changes between pictures in the transition. This paper proposes a overlay coding technique for coding faded scene transitions. As shown by extensive simulations, over 50% bit-rate savings in both cross-fades and through-black fades compared to earlier techniques can be achieved. Overlay coding suits situations where video is edited manually or automatically.

## 1. INTRODUCTION

Many types of video material, such as news, movie trailers, and music video, contain frequent scene changes. Abrupt scene cuts are most common, but sometimes transitions, such as fades or wipes, are applied. Coding of a scene transition is often a challenging problem, from compression efficiency point of view, because motion compensation may not be a powerful enough method to represent the changes between pictures in the transition.

Herein, we categorize scene transitions to abrupt, masked, faded, and hybrid ones. Abrupt transitions are such that there is no transition period and no picture where the contents of the two subsequent scenes are present at the same time, and therefore abrupt transitions are in fact scene cuts. Masked transitions are such that the second scene spatially uncovers from the first scene in a gradual manner. All pictures are displayed at full intensity. Examples of masked scene transitions include box-in, box-out, wipes, and splits. Faded transitions are such that the pictures of the two scenes are laid on top of each other in semi-transparent manner, and the transparency of the pictures at the top gradually changes in the transition period. Examples of faded transitions include a normal cross-fade and fade through (to and from) black. Hybrid scene transitions are a combination of masked and faded transitions.

The main goals of the ongoing JVT/H.26L standardization effort [1] are the definition of a simple and straightforward video coding design to achieve enhanced compression performance and provision of a "network-friendly" packet-based video. The video coding layer (VCL) design has achieved a significant improvement in rate-distortion efficiency - providing nearly a factor of two in bit-rate savings against existing standards.

In this paper, the overlay coding method is proposed for coding of cross-fades and through-black fades. Compared to earlier faded scene transitions coding techniques, coding efficiency is remarkably improved. The paper is organized as follows. An overview of earlier coding techniques for faded scene transitions is given in section 2. Section 3 presents the overlay coding technique and its application in cross-fades and through-black fades. Simulation results are shown in section 4 and section 5 concludes the paper.

## 2. EARLIER CODING TECHNIQUES

### 2.1. Conventional Coding

Conventionally, coding of faded scene transitions is done as follows. The pictures within the transition period are composed prior to encoding. After composing, the pictures in the transition period are encoded normally. However, the compression efficiency is moderate [2], because conventional motion compensation is not capable of handling intensity changes or mixing of two scenes.

### 2.2. Weighted Averaging of B-Pictures

Lillevold [2] proposed to scale or weight the calculation of the prediction block based on the distance between a B-picture and its anchor pictures, which can be used to improve the coding efficiency within faded scene transitions. The following formula for the prediction block was proposed:

$$P = (Pp \times (TRD - TRB) + Pf \times (TRB)) / TRD \qquad (1)$$

where $Pp$ is the prediction block from the previous reference picture, and $Pf$ is the prediction block from the future or subsequent reference picture. $TRD$ is the temporal distance between the temporally previous and next reference pictures, and $TRB$ is the temporal distance between the current picture and the previous reference picture.

There are two approaches to code a scene transition with the method. One is to use a normal frame pattern, such as PBBB. The other is to code the whole transition as B-pictures whose reference pictures are the closest non-faded pictures. Hereinafter, we refer to the former approach as the weighted B-picture scene transition coding with normal frame pattern (normal WB-SCT coding), and to the latter approach as the weighted B-picture scene transition coding with B-frames across the transition period (all-B WB-SCT coding).

### 2.3. Interpolative Motion Compensation

An interpolative motion compensation method was proposed in [3]. Two reference frames, $r_1$ and $r_2$, in the multi-frame buffer are used for the interpolation. First, motion compensation signals are generated for each of the two reference frames. Then, a weighted sum of the motion compensated signals is calculated as the prediction signal. The weighting factors $w_1$ and $w_2$ are used for reference frames $r_1$ and $r_2$, respectively. The weighting factor pair (2, -1) is useful for video sequences with linear fading. The brightness factor of fading at the current frame ($b_0$) equals to the extrapolation of those for reference frames $r_2$ and $r_1$ ($b_2$ and $b_1$), independent of the fading factor (the brightness changing ratio). That is, $b_0=2b_1-b_2$.

### 2.4. Generalized Weighted Averaging of B/MH-Pictures

Weighted averaging of B-pictures can only be applied to "real" B-pictures whose reference frames surround the B-picture temporally. Interpolative motion compensation works optimally only if the picture rate is steady. It was shown in [4] that there is a generalized formula for pixel amplitude scaling according to temporal distance for any B/MH-pictures [1] without any constraints on temporal frame positions or frame rate. The expression yields the weighted averaging of B-pictures and interpolative motion compensation if their constraints above are fulfilled.

## 3. OVERLAY CODING

### 3.1. General

In the overlay coding, pictures in the first scene (called bottom/source pictures) are coded independently from pictures in the second scene (called top pictures). Any frame pattern (e.g. PPP, PBP, or PBBP) can be used. The reconstructed pictures from the two scenes are stored in the multi-picture buffer to enable efficient motion compensation during the transition.

In case of cross-fades, overlapping component images are overlaid so that the top picture is partially transparent. While in through-black fades, the top pictures are appended to the end of the source pictures to fulfill the transition.

Since the bottom and top pictures should be available for the encoder, overlay coding suits for situations where video is edited manually or automatically.

### 3.2. Analysis of Faded Subsequences

Both through-black fading and cross-fading sequences can be decomposed into to-black and from-black subsequences. The fading component subsequences can produce through-black fade or cross-fade by merging or being overlaid. The to-black and from-black fades can be formularized as:

$$P_F(i) = F(i) \cdot P_O(i) \qquad (2)$$

where $i$ denotes the picture index within a scene transition, $P_O$ stands for the original picture, $P_F$ denotes the faded picture, $F(i)$ denotes the fading function and should fall within [0, 1].

In the JVT/H.26L testing model, the following formulas can be used for each pixel to produce the fading effect:

$$\begin{aligned} y_F &= F \times y_O \\ c_F &= F \times (c_O - 128) + 128 \end{aligned} \qquad (3)$$

where $y_F$ and $c_F$ are the luminance and chrominance faded pixels, $y_O$ and $c_O$ are the corresponding luminance and chrominance of the original pixels, respectively, and $F$ is the amplitude fading factor in the range [0, 1].

There may be many kinds of fading models. The following model will produce a uniform linear to-black fading effect:

$$F(i) = \frac{L-1-i}{L-1} \qquad (4)$$

where $L$ denotes the fading length in number of pictures and $i$ is the the picture number within the fade.

When coding the faded component subsequence, the different picture brightness (or pixel value amplitude) in reference frames and in the current frame drops the efficiency of motion compensation. Thus, we are trying to adjust the brightness in reference frames so that they have the same brightness as in the current frame.

We assume that the fading function $F(i)$ can be expressed as follows:

$$F(i) = S(i,j) \cdot F(j) \qquad (5)$$

where $S(i, j)$ is a scaling factor between $F(i)$ and $F(j)$. $S(i, j)$ will help to adjust the reference frames to have the same brightness before predicting the current frame. That is, the $j$th reference frame, $P_F(j)$, is first transformed as:

$$P'_F(j) = S(i,j) \cdot P_F(j) \qquad (6)$$

For the uniform linear to-black fading as indicated by (4), the scaling function is:

$$S(i,j) = \begin{cases} \frac{L-1-i}{L-1-j}, & L-1-j \neq 0 \\ 1, & otherwise \end{cases} \qquad (7)$$

### 3.3. Overlay Coding

Basically, the original bottom and top pictures can be encoded normally and independently. We refer to this technique as the simple overlay coding.

When observing the to-black fading component subsequence, it will be found that less and less content is contained in the faded pictures, thus less bit-rate would be likely needed for their representation. However the normal motion compensation is deficient due to the different brightness or pixel value amplitude in the fading pictures. By scaling the reference frame with function *S(i, j)*, we will get the pixel amplitude weighting overlay coding.

When coding the faded component subsequence, it is generally more beneficial to predict a low-amplitude picture from a high-amplitude picture compared to the other way around. This is due to the fact that a high-amplitude picture preserves details better than a low-amplitude picture. Thus, it may be beneficial to code from-black fades in temporally reverse order if memory capabilities and latency requirements of the system allow. Instead of coding the low-amplitude first frame as intra, the first full-amplitude frame in the second scene is selected as an intra-frame. Then, the fractional-amplitude frames are coded using the high-amplitude frames as references. We refer to this technique as the bottom-top pixel-amplitude-weighting overlay coding (BT-PAW overlay coding).

Another possibility of avoiding prediction of a high-amplitude picture from a low-amplitude one is to select the faded bottom subsequence and the original top subsequence as the component subsequences. Then, only the pixel amplitude values of the second scene need to be scaled in run-time. We refer to this technique as bottom pixel-amplitude-weighting overlay coding (B-PAW overlay coding).

When coding the fading component images, generalized weighted averaging of B-pictures can be used to further improve compression efficiency.

### 3.4. Bit-Rate Scalability

The scene transition can be coded scalably or non-scalably in terms of bit-rate. In scalable coding of scene transition, top component pictures can be discarded. If top component picture are not transmitted or decoded, cross-fade or through-black fade becomes an abrupt scene cut.

Note that B-pictures can provide scalability in terms of bit-rate as well. However, disposal of B-pictures in cross-fades causes a fluctuation in picture rate, which is typically considered somewhat annoying. Thus, the overlay coding also outperforms B-picture based fades coding from the scalability point of view.

### 3.5. Random Access to Scene Transitions

Random access to coded video streams is one of the important issues in JVT/H.26L applications. Overlay coding typically provides a random access point, because it is logical to code an intra frame at the beginning of the second component sequence of a scene transition. Thus, an overlay-coded scene transition can be accessed randomly, when only the second component sequence of the transition is decoded.

## 4. SIMULATIONS

### 4.1. General

Extensive simulations are performed to compare weighted averaging of B-pictures and overlay coding techniques under the JVT software JM-1.4 [5]. Uniform linear cross-fade and through-black fade are selected as the fading model, in which the component subsequences follow the uniform linear to-black or from-black model.

First, we compare normal WB-SCT coding and all-B WB-SCT coding in order to see which one is a better comparison point for further tests. As a result, it turns out that normal WB-SCT outperforms all-B WB-SCT. Then we search for an optimal number of B-pictures in order to reach the best performance with weighted B-picture averaging. These tests help us to set the testing conditions in Table 1.

**Table 1 Summary of the testing conditions.**

| | |
|---|---|
| R-D optimization | used |
| MV search range | ±32 |
| B picture number | 3 for QCIF, 2 for CIF |
| Frame Rate | 15Hz for QCIF, 30Hz for CIF |
| CABAC | used |
| Motion Compensation | up to ¼ pel accuracy |
| Number of reference pictures | 5 |
| QP | 16, 20, 24, 28 |
| Scene transition begins at | 5th P picture |

The 5th P picture is selected as the scene transition starting point to guarantee that there are enough reference pictures for all pictures in the scene transition, since five reference frames are used in the codec [5]. Four quantization parameters (QPs) are used to plot the rate-distortion curves.

We compare normal WB-SCT, simple overlay coding, B-PAW overlay coding, and BT-PAW overlay coding.

### 4.2 Simulation Results

We selected the following transitions as examples to show the performance of overlay coding methods: Tempete-to-

Coastguard in CIF as a 0.25-second through-black fade transition and Foreman-to-Carphone in QCIF as a 1-second cross-fade. Figure 1 and Figure 2 show their R-D curves. The Bjontegaard [7] average PSNR improvement and bit-rate savings are presented in Table 2. More results can be found in [6].

**Table 2 Bjontegaard average difference between RD curves of overlay methods and normal WB-SCT**

| - | | simple | B-PAW | BT-PAW |
|---|---|---|---|---|
| Tempete to Coastguard | Δ*PSNR* | 0.86 dB | 1.18 dB | 1.66 dB |
| | Δ*Bitrate* | -24.58% | -33.00% | -43.45% |
| Foreman to Carphone | Δ*PSNR* | 1.37 dB | 2.09 dB | 3.41 dB |
| | Δ*Bitrate* | -26.69% | -38.86% | -54.30% |

Overlay coding outperforms weighted B-picture averaging remarkably. Over 50% bit-rate savings in the scene transition period can be achieved. BT-PAW overlay coding performs better than B-PAW overlay coding that outperforms simple overlay coding.

Figure 3 presents the 12th picture of the Foreman-to-Carphone cross-fade to show that the subjective quality can be greatly improved at a similar bit-rate by using BT-PAW overlay coding.

## 5. CONCLUSIONS

This paper proposes overlay coding techniques for faded scene transition coding. Overlay coding is based on independent coding of component sequences in the scene transition and run-time composition of the fade. Extensive simulation results show that over 50% bit-rate savings can be achieved in the scene transition period compared to earlier coding techniques.

In addition to the superior compression performance, overlay coding provides the possibility to scale scene transitions in terms of bit-rate. Another advantage is that overlay coding provides "clean" random access points to faded scene transitions. Such random access points are impossible without an additional intra frame in earlier solutions.

## 6. REFERENCES

[1] Thomas Wiegand, Gary Sullivan, "Working Draft Number 2, Revision 8 (WD-2, rev 8)", Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, April 2002.

[2] Karl Lillevold, "Improved direct mode for B pictures in TML", ITU-T Video Coding Experts Group (Question 15), document Q15-K44, August 2000.

[3] Yoshihiro Kikuchi, Takeshi Chujoh, "Improved multiple frame motion compensation using frame interpolation", Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, document JVT-B075, Feb 2002.

[4] Miska M. Hannuksela, "Generalized B/MH-Picture Averaging", Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, document JVT-C077, May 2002.

[5] JVT Test Model Software, version 1.4, Feb 2002.

[6] Dong Tian, Miska M. Hannuksela, "Coding of Faded Scene Transitions", Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, document JVT-C076, May 2002.

[7] Gisle Bjontegaard, "Calculation of average PSNR differences between RD-curves", ITU-T Video Coding Experts Group, document VCEG-M33, March 2001

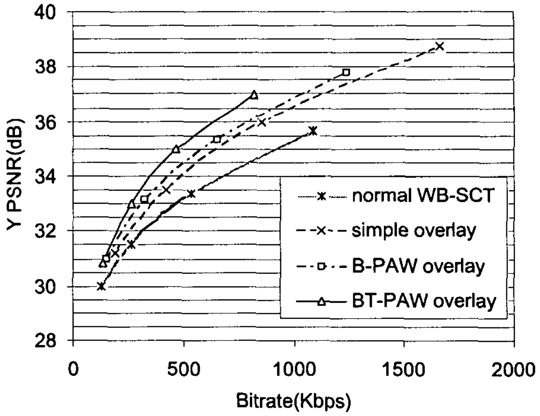

**Figure 1 Through-black fade from Tempete to Coastguard.**

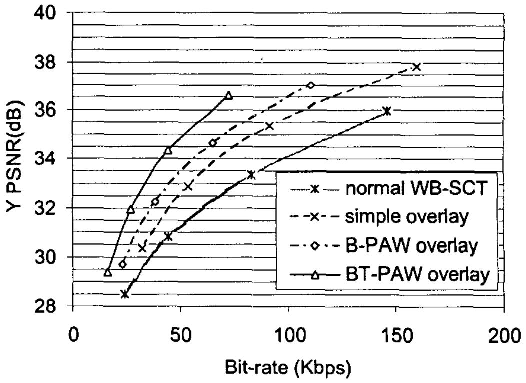

**Figure 2 Cross-fade from Foreman to Carphone.**

original

normal WB-SCT
44.4Kbps, 30.80dB

BT-PAW overlay
44.6Kbps, 34.33dB

**Figure 3 The 12th frame in the cross-fade .**