

# NEW WEIGHTED PREDICTION ARCHITECTURE FOR CODING SCENES WITH VARIOUS FADING EFFECTS

## *Image and Video Processing*

Sik-Ho Tsang, Yui-Lam Chan and Wan-Chi Siu

*Centre for Signal Processing, Department of Electronic and Information Engineering  
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

Keywords: Brightness Variations, Weighted Prediction, Multiple Reference Frames, H.264.

**Abstract:** Weighted prediction (WP) is one of the new tools in H.264 for encoding scenes with brightness variations. However, a single WP model does not handle all types of brightness variations. Also, large luminance difference induced by object motions would mislead an encoder in its use of WP which results in low coding efficiency. To solve these problems, a picture-based multi-pass encoding strategy, which extensively encodes the same picture multiple times with different WP models and selects the model with the minimum rate-distortion cost, has been adopted in H.264 to obtain better coding performance. However, computational complexity is impractically high. In this paper, a new WP referencing architecture is proposed to facilitate the use of multiple WP models by making a new arrangement of multiple frame buffers in multiple reference frame motion estimation. Experimental results show that the proposed scheme can improve prediction in scenes with different types of brightness variations and considerable luminance difference induced by motions within the same sequence.

## 1 INTRODUCTION

Block-based motion estimation and compensation are among the most popular approach to reduce temporal redundancy in video coding by estimating a motion vector of each block between successive frames. It assumes that brightness between frames is constant during motion estimation and compensation, changes between video frames may be caused by object movements or camera motions rather than brightness changes between frames. When brightness variations, such as fade-in/out effects and camera flashes, occur between successive frames, the motion estimation cannot be accurately performed. True motion vectors cannot be obtained which may increase the amount of the prediction errors. Inter modes with large distortion or even intra modes would be chosen mostly through rate-distortion optimization (RDO) as the optimal modes. Consequently, coding efficiency may be reduced in the presence of brightness variations.

To solve this problem, weighted prediction (WP), in (Joint Model 15.1, 2009), has been adopted in the Main and Extended Profiles of an H.264/AVC video coding standard to enhance motion compensation in

scenes with brightness variations by modifying the original reference frame  $P_{ref}$  with a multiplicative weighting factor  $W$  and an additive offset  $O$  (Joint Model 15.1, 2009, Boyce., 2004, Kato and Nakajima, 2004, Aoki and Miyamoto, 2008, Zhang and Cote, 2008). With the use of WP, the weighted reference frame  $\tilde{P}_{ref}$  becomes

$$\tilde{P}_{ref} = W \times P_{ref} + O \quad (1)$$

To alleviate the problems of brightness variations,  $\tilde{P}_{ref}$  is used instead of  $P_{ref}$  in motion estimation and compensation. In the process of motion estimation, the sum of absolute differences (SAD) using WP is

$$SAD = \sum |P_{curr}(i, j) - \tilde{P}_{ref}(i + x, j + y)| \quad (2)$$

where  $(i, j)$  is the pixel location in the given block of  $P_{curr}$ , and  $(x, y)$  denotes the candidate motion vector. Hence, in sequences with brightness variations,  $P_{curr}$  is more strongly correlated to  $\tilde{P}_{ref}$  than to

$P_{ref}$  itself. By referencing  $\tilde{P}_{ref}$ , fewer bits are needed to encode  $P_{curr}$ .

Nevertheless, different types of fading effects may happen in different video segments. This always occurs in nowadays movies that contain many special effects. In this case, one single WP model is not sufficient to support sequences with diverse fading effects. In addition, a pre-processing fade detector is always needed to discriminate scenes contains fading from the scenes with large luminance differences induced by motions in order to avoid using WP wrongly. Notwithstanding, an accurate fade detector is a difficult task. In H.264, multi-pass encoding allows videos to be encoded multiple times so as to make the decision for the use of WP and the selection of WP models that keeps the best quality. This approach, however, induces high computational complexity.

Therefore, in this paper, we propose to utilize the structure of multiple reference frames in H.264 in order to facilitate the use of multiple WP models. This can easily handle different brightness variations with better compensation in sequences. The rest of the paper is organized as follows. Section 2 gives a brief description of various WP models. A new WP referencing architecture is presented in Section 3. Experimental results are shown in Section 4. Finally, Section 5 concludes this paper.

## 2 CONVENTIONAL WP MODELS

There are four commonly used models in H.264. They are hereafter referred to as DC, Offset, LS and LMS, and their corresponding  $W$  and  $O$  are computed as follows:

$$W_{ref}^{DC} = \overline{P_{curr}} / \overline{P_{ref}}, O_{ref}^{DC} = 0 \quad (3)$$

$$W_{ref}^{OFF} = 1, O_{ref}^{OFF} = \overline{P_{curr}} - \overline{P_{ref}} \quad (4)$$

$$\begin{aligned} W_{ref}^{LS} &= \frac{\overline{P_{curr}} \cdot \overline{P_{ref}} - \overline{P_{curr}} \cdot \overline{P_{ref}}}{\overline{P_{ref}}^2 - \overline{P_{ref}}^2} \\ O_{ref}^{LS} &= \frac{\overline{P_{curr}} \cdot \overline{P_{ref}}^2 - \overline{P_{curr}} \cdot \overline{P_{ref}} \cdot \overline{P_{ref}}}{\overline{P_{ref}}^2 - \overline{P_{ref}}^2} \end{aligned} \quad (5)$$

$$\begin{aligned} W_{ref}^{LMS} &= \frac{\sum |P_{curr} - \overline{P_{curr}}|}{\sum |P_{ref} - \overline{P_{ref}}|} \\ O_{ref}^{LMS} &= \overline{P_{curr}} - W_{ref}^{LMS} \cdot \overline{P_{ref}} \end{aligned} \quad (6)$$

For the DC model (Boyce, 2004), in Eq. (3),  $W$  is estimated as the ratio of the mean value of the current frame  $\overline{P_{curr}}$  and the mean value of the reference frame  $\overline{P_{ref}}$ , and  $O$  is set to 0. This model has been proved to be favourable for coding the scenes with fade-in from black or fade-out to black only (Kato and Nakajima, 2004). In the offset model (Boyce, 2004),  $O$  is simply computed by the difference between  $\overline{P_{curr}}$  and  $\overline{P_{ref}}$  and  $W$  is set to 1, as in Eq. (4). This model is more suitable for the scenes with fade-in from white or fade-out to white as it can still estimate the slight difference between  $\overline{P_{curr}}$  and  $\overline{P_{ref}}$  as offsets when  $\overline{P_{curr}}$  and  $\overline{P_{ref}}$  become large. However both models suffer from low coding efficiency due to the clipping of pixel values from 0 to 255 in the area where the difference between pixel values and mean value of the frame is large. On the other hand, LS model (Kato and Nakajima, 2004), as in Eq. (5), optimizes an error function using a least square technique, and only works well in coding scenes with low motion activity since it depends on the mean of the product of the current frame and the corresponding pixels in the same position in the reference frame (Aoki and Miyamoto, 2008). For the LMS model (Aoki and Miyamoto, 2008, Zhang and Cote, 2008), in Eq. (6),  $W$  and  $O$  are derived theoretically based on the fade effect equation which is fading from/to a fixed colour only. Otherwise it might not work well.

It is expected that no single WP model can cope with all situations of brightness variations. And it is useless to apply any kinds of WP models on scenes without any brightness variations. Thus multi-pass WP encoding strategy was adopted in H.264 to encode the same picture multiple times including without the use of WP and using various WP models (Tourapis et al., 2005). The encoder then selects the mode with the minimum picture-based rate-distortion cost. Though the multi-pass WP encoding can choose the optimal WP models for coding scenes with brightness variations and also avoid misleading use of WP for coding scenes with large luminance difference induced by large object motions, computational complexity is extremely high. Consequently, it is not practical and become impossible for real-time encoding.

### 3 THE PROPOSED SCHEME

The H.264 video coding standard supports multiple reference frame motion estimation (MRF-ME) to improve coding performance, with a reference picture index (*ref\_idx*) coded to indicate which of the multiple reference frames is used. If MRF-ME is enabled, more than one WP parameter sets are sent per slice. Figure 1 shows a block diagram to illustrate the use of WP with MRF-ME in an H.264 encoder. A single WP model, say DC model, is applied to multiple reference frames from  $f_0$  to  $f_4$  for motion estimation. Different weighting factors,  $W_{f_0}^{DC}$ ,  $W_{f_1}^{DC}$ ,  $W_{f_2}^{DC}$ ,  $W_{f_3}^{DC}$ , and  $W_{f_4}^{DC}$ , are computed according to Eq. (3) for  $f_0$  to  $f_4$ , and their weighted reference frames ( $\tilde{P}_{f_i}$ , where  $i=0,1..4$  in this example) are placed in the multiple frame buffers for motion estimation and compensation, as depicted in Figure 1. The WP parameter set applied to the current MB is indicated in *ref\_idx*. By doing so, the decoder can recognise the WP parameter set correctly. Since *ref\_idx* is already available in the bitstream, the use of this index to indicate which WP parameter set for each MB can avoid the need of additional bits, which increases coding efficiency. But, as aforementioned, different WP models are suitable for different kinds of brightness variations or fading effects. The selection of WP models in advance of WP parameter estimation in a real-time encoding system may not be possible for a variety reasons. First, a complicated process is needed to detect the existence and types of brightness variations in order to select the most probable WP model. Second, most of the brightness variation detection algorithms in the literature depend on a relatively long window of frames to analyse enough statistics for an accurate detection (Zhang and Cote, 2008, Alattar, 1997). Third, a multi-pass encoding strategy by trying all WP models as well as coding without the use of WP for each coding picture is needed if there is no brightness variation detection. However, computational complexity is impractically high.

Therefore, we examine a way to jointly use of multiple frame buffers and weighted prediction such that more than one *ref\_idx* can be associated with a particular reference picture. Figure 2 shows the new arrangement of the multiple frame buffers for the proposed scheme. In this figure, different parameter sets, ( $W_{f_0}^{DC}, O_{f_0}^{DC}$ ), ( $W_{f_0}^{OFF}, O_{f_0}^{OFF}$ ), ( $W_{f_0}^{LS}, O_{f_0}^{LS}$ ), and ( $W_{f_0}^{LMS}, O_{f_0}^{LMS}$ ) are estimated between the current frame and  $f_0$ . And their weighted reference frames

are stored in the multiple frame buffers for motion estimation. This scheme allows different MBs in the current frame to employ different WP parameters even when predicted from the same reference frame. It is noted that the original reference frames without WP, from  $f_0$  to  $f_4$ , are also stored in the frame buffers, and they are beneficial to code the scenes without brightness variation. In the presence of brightness variations, the correlation between weighted reference frames (WRFs) and current frame is larger than that between temporal reference frames (TRFs) and current frame. Those WRFs, rather than the nearest temporal reference frame  $f_0$  or other TRFs, tend to be used as the reference even though WRFs has a greater *ref\_idx* than  $f_0$  and other TRFs. With the help of this arrangement, we can handle diverse fading scenes appropriately. In addition, the proposed scheme allows multiple WP models to be used in the current frame. It can also potentially improve the prediction for the scenes with local brightness variations.

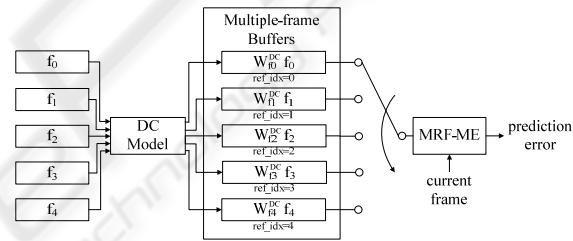


Figure 1: Conventional WP using DC model in H.264.

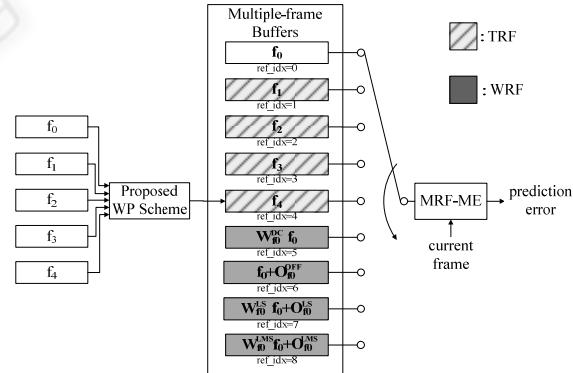


Figure 2: Proposed WP scheme.

Since fading is always applied over a few seconds, the correlation between frames that uses a particular WP model remains reasonably high. Whenever fading exists, a new reordering mechanism can further reduce the bits to encode *ref\_idx*. By using this property, the proposed scheme determines which WP model is likely to be used in the 8x8 blocks of the previously encoded frame, and

the probabilities of using different WP models are calculated. The reference picture list used in the current frame is re-ordered based on these probabilities so that the WP model that is most likely to be used is first in the list. This allows using shorter codes for  $ref\_idx$  in the encoded bitstream, which results in further decreasing the bitrate. For this re-ordering mechanism, TRFs and WRFs are swapped such that smaller  $ref\_idx$  can be assigned to WRFs that is always to be referred for bitrate reduction in the presence of fading effects. Within  $f_0$  and WRFs, the most frequently used reference frame, which gives the best prediction, is first in the list to lower the required bitrate. On the other hand, in the scenes without brightness variation, the nearest temporal reference frame  $f_0$  and other TRFs are put first in the list. That is, TRFs are assigned the smaller  $ref\_idx$ .

In JM15.1 (Joint Model 15.1, 2009), a partial distortion search strategy is used in nearly all motion estimation options. The partial distortion search provides same results as full search with reduced complexity (Chan et al., 2004, Hui et al., 2005). It rejects impossible candidate motion vectors by means of half-way stop technique with partial distortion comparison to the current minimum distortion in a pixel-wise basis. If the current minimum distortion is computed sooner, the impossible candidates will be eliminated faster, which results in deceasing encoding time. In fact, the reordering mechanism used in the proposed scheme is also beneficial to partial distortion search since the most likely reference frames first is put first in the list. Consequently, the minimum distortion may be computed sooner. In contrast, the multi-pass approach cannot take this advantage since each pass is an independent search.

## 4 SIMULATION RESULTS

Sequences including “Akiyo”, “Foreman”, “M&D”, and “Silent” with the frame size of 352x288 were used to conduct the experiments. All experiments were conducted using IPPP... structure, quarter-pel full search motion estimation with search range of  $\pm 32$  pixels, RDO with all seven inter modes as well as intra modes, and CABAC using four different QPs (20, 24, 28, and 32). As shown in Figure 3, four kinds of fading effects, which are two-second long 60 frames each, were applied to sequences at different segments. 300 frames were encoded in total for each sequence including a segment with no fading effects.

We incorporated our proposed scheme (Proposed) into the H.264 reference software JM 15.1 (Joint Model 15.1, 2009). The conventional WP with different models such as DC, Offset, LS, and LMS were adopted for performance comparison. They are denoted by WP-DC, WP-OFF, WP-LS and WP-LMS, respectively. The scheme without WP (Without WP) was also included in all experiments.

Besides, we also implemented the multi-pass encoding approach (WP-MultiPass) for comparison. In WP-MultiPass, five-pass encoding (Without WP, WP-DC, WP-OFF, WP-LS and WP-LMS) was adopted.

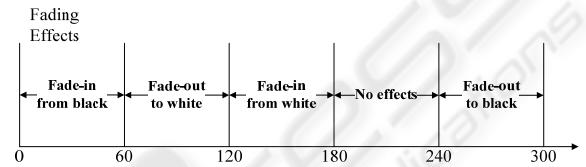


Figure 3: Different fading effects applied along sequences.

Figure 4 shows the rate-distortion (RD) performance of different schemes for the “Foreman” sequence from frame 0 to 59, which contains a fade-in from black effect. The proposed scheme shows the best RD performance in comparison with the other schemes. Gain is obtained of at most 2.8dB compared to ‘Without WP’. This gain is also significant as compared with other four conventional WP models. It is noted that the performance of our proposed scheme is even better than that of “WP-MultiPass”. It is due to the reason that “WP-MultiPass” can only use one WP model on slice level though it can choose the best WP model for each frame, whilst our proposed scheme can choose the best WP model on MB level which helps to improve the coding efficiency. It can be explained by Table 1 in which the distribution in percentage of referencing the temporal nearest reference without WP ( $f_0$ ), TRFs, WRFs and also being intra-encoded of 8x8 blocks for “Foreman” at QP=20 is shown for the proposed scheme. The use of WRFs is dominant throughout the fading period, as shown in Table 1. As expected, WRFs can help to get smallest cost through RDO when there is fade-in from black effect. For other fading effects within the sequence, similar results are shown in Table 1. Summary results of different schemes for different sequences using BD-PSNR and BD-Bitrate measurement (Bjontegaard, 2001) compared to ‘Without WP’ are included in Table 2. From this table, it can be easily seen that the proposed scheme can significantly outperform other schemes.

To demonstrate the ineffectiveness of the other single model scheme over high motion scenes without brightness variation, a fast camera panning motion in the “Foreman” sequence (frame 180 – frame 239) was encoded by different schemes. The RD performances are shown in Figure 5. ‘WP-LS’ obtains the worst performance due to its motion-sensitive characteristics. It is at most 2.3dB drop comparing with ‘Without WP’ while other single model scheme such as ‘WP-DC’, ‘WP-OFF’, and ‘WP-LMS’ also performs unsatisfactory. It is due to the fact that large luminance difference induced by object motions may mislead the encoder in its use of WP and it causes irrelevant WP parameter sets to be computed. Using wrong weighted parameters in motion estimation is likely to get a larger RD cost comparing with that of ‘Without WP’. It results in lower coding efficiency. On the other hand, our proposed scheme can prevent this situation and keep similar RD performance with ‘Without-WP’ and ‘WP-MultiPass’. Table 1 also shows the number of 8x8 blocks referencing WRFs is the least from frame 180 to frame 239 which has a camera panning effect. Hence, from Figure 5 and Table 1, our proposed scheme can handle all types of brightness variations and object motions effectively Figure 6 then shows the overall RD performances of different schemes for the whole “Foreman” sequence, which contains various fading effects and fast camera panning, our proposed scheme can achieve the best coding performance even comparing with “WP-MultiPass”.

Even though “WP-MultiPass” shows relatively good performance in Figure 4 to Figure 6, it needs high computational complexity because the encoding has to be performed five times per frame to calculate the RD costs. To further show the complexity of the proposed scheme, the encoding time reduction of various schemes in comparison with “WP-MultiPass” for all test sequences are measured and tabulated in Table 3. The experiments were preformed on an Intel Xeon X5550 2.67GHz computer with 12GB memory. As a partial distortion search strategy is used in the full search inside JM15.1, faster encoding time is expected when the minimal distortion is obtained earlier. As shown in Table 3, the reduction of encoding time of proposed scheme is about 10% compared to other single-model WP schemes, but it achieves a remarkable RD improvement. On the other hand, the proposed scheme can still achieve averagely 70.66% reduction of encoding time compared with “WP-MultiPass”.

According to these results, the proposed scheme obtains the best RD performance among all schemes with relatively low computational complexity.

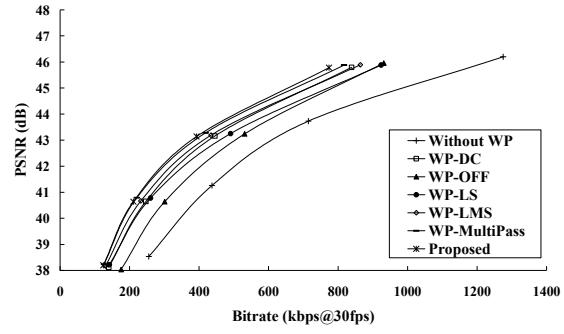


Figure 4: RD performances of different schemes for “Foreman” from frame 0 to 59 with fade-in from black effect.

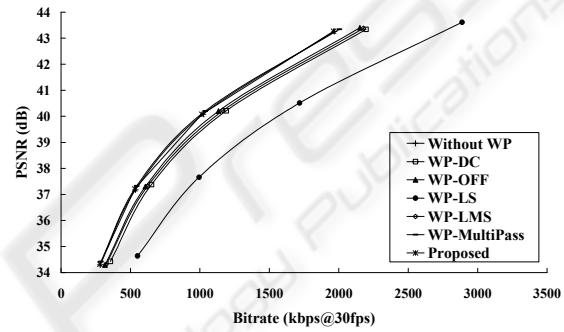


Figure 5: RD performances of different schemes for “Foreman” from frame 180 to 239 with fast panning camera movement.

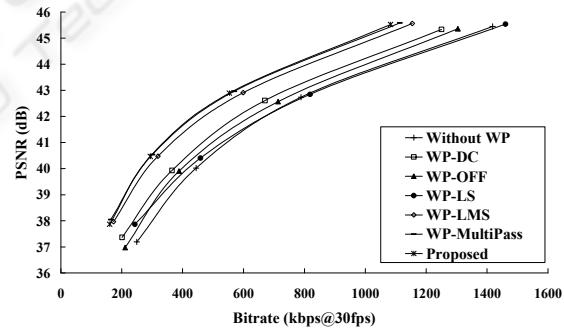


Figure 6: Overall RD performances of different schemes for “Foreman” from frame 0 to 299.

## 5 CONCLUSIONS

In this paper, weighted prediction utilizing a multiple reference frames architecture and a new reference reordering mechanism has been proposed. Our proposed scheme can efficiently handle sequences with different types of brightness variations. It can be concluded from the experimental results that the proposed scheme can outperform any conventional WP models in scenes

with different types of fading effects, and select an appropriate WP model. Results also show that it is even better than the multi-pass WP encoding strategy with reduced complexity.

Table 1: Distribution (%) of 8x8 blocks referencing  $f_0$ , TRFs, WRFs and being intra-encoded for “Foreman” at QP20 using proposed scheme.

Frame No.	Fading Effects	$f_0$	TRFs ( $f_i$ to $f_4$ )	WRFs	Intra
0-59	Fade-in from black	7.64	10.69	71.58	10.09
60-119	Fade-out to white	11.47	16.87	66.49	5.17
120-179	Fade-in from white	11.95	12.28	61.45	14.32
180-239	-	71.48	12.99	7.32	8.22
240-299	Fade-out to black	0.20	46.41	50.02	3.37

Table 2: BD-PSNR (dB) and BD-Bitrate (%) compared to H.264 without WP.

Methods	Akiyo	Foreman	M&D	Silent	Average
WP- DC	BD-PSNR	1.44	0.72	1.46	1.15 <b>1.19</b>
	BD-Bitrate	-24.52	-14.79	-27.46	-21.70 <b>-22.12</b>
WP- OFF	BD-PSNR	1.30	0.41	1.36	0.76 <b>0.96</b>
	BD-Bitrate	-21.50	-8.51	-24.52	-14.65 <b>-17.30</b>
WP- LS	BD-PSNR	3.15	0.16	2.39	1.81 <b>1.88</b>
	BD-Bitrate	-48.98	-3.08	-41.84	-32.55 <b>-31.61</b>
WP- LMS	BD-PSNR	3.54	1.60	2.87	2.71 <b>2.68</b>
	BD-Bitrate	-54.01	-31.01	-48.36	-44.06 <b>-44.36</b>
WP- Multi- Pass	BD-PSNR	3.83	1.87	3.11	2.87 <b>2.92</b>
	BD-Bitrate	-56.27	-35.34	-51.37	-45.68 <b>-47.17</b>
Proposed	BD-PSNR	4.09	1.91	3.30	2.93 <b>3.06</b>
	BD-Bitrate	-58.59	-35.87	-52.63	-46.49 <b>-48.40</b>

Table 3: Change of Encoding Time (%) compared to “WP-MultiPass”.

Methods	Akiyo	Foreman	M&D	Silent	Average
Without WP	-71.75	-77.42	-70.98	-74.25	<b>-73.60</b>
WP-DC	-79.61	-80.37	-81.14	-80.27	<b>-80.35</b>
WP-OFF	-78.70	-79.65	-81.02	-79.37	<b>-79.68</b>
WP-LS	-82.39	-78.18	-81.33	-80.55	<b>-80.61</b>
WP-LMS	-83.99	-82.25	-83.09	-82.40	<b>-82.93</b>
Proposed	-71.58	-70.38	-70.86	-69.84	<b>-70.66</b>

## ACKNOWLEDGEMENTS

The work described in this paper is partially supported by the Centre for Signal Processing, Department of EIE, PolyU and a grant from the Research Grants Council of the HKSAR, China (PolyU 5120/07E).

## REFERENCES

- Joint Model 15.1, 2009. H.264/AVC Reference Software.  
 From <http://iphome.hhi.de/suehring/tm1/download/>.
- Boyce, J. M., 2004. Weighted prediction in the H.264/MPEG AVC video coding standard. In *Proceedings of the 2004 International Symposium on Circuits and Systems (ISCAS'04), Vancouver, Canada, May 2004, vol. III, pp. 789-792.*
- Kato, H., Nakajima, Y., 2004. Weighted factor determination algorithm for H.264/MPEG-4 AVC weighted prediction. In *IEEE 6th Workshop on Multimedia Signal Processing, Siena, Italy, 2004, pp. 27-30.*
- Aoki, H., Miyamoto, Y., 2008. An H.264 weighted prediction parameter estimation method for fade effects in video scenes. In *15th IEEE International Conference on Image Processing (ICIP'08), San Diego, California, U.S.A., Oct. 2008, pp. 2112-2115.*
- Zhang R., Cote, G., 2008. Accurate parameter estimation and efficient fade detection for weighted prediction in H.264 video compression. In *15th IEEE International Conference on Image Processing (ICIP'08), San Diego, California, U.S.A., Oct. 2008, pp. 2836-2839.*
- Tourapis, A. M., Suhring, K., Sullivan G., 2005. H.264/MPEG-4 AVC Reference Software Enhancements. In *14th Meeting of Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Hong Kong, PRC China, Jan. 2005, document JVT-N014.*
- Alattar, A. M., 1997. Detecting fade regions in uncompressed video sequences. In *22nd IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997 (ICASSP'97), Munich, Germany, Apr. 1997, pp. 3025-3028.*
- Chan, Y. L., Hui, K. C., Siu, W. C., 2004. Adaptive Partial Distortion Search for Block Motion Estimation. In *Journal of Visual Communication and Image, U.S.A., Nov. 2004, Vol.15, Issue 4, pp. 489-506.*
- Hui, K. C., Siu, W. C., Chan, Y. L., 2005. New adaptive partial distortion search using clustered pixel matching error Characteristic. In *IEEE Transactions on Image Processing, May 2005, Vol.14, No.5, pp.597-607.*
- Bjontegaard, G., 2001. Calculation of average PSNR differences between RD-curves. In *13th Meeting of ITU - Video Coding Experts Group (VCEG), Austin, Texas, USA, Apr. 2001, document VCEG-M33.*