# A FAST INTRA ENCODER OF FRAME-COMPATIBLE FORMAT BASED ON CONTENT SIMILARITY FOR 3D DISTRIBUTION

*Zhuoying Zeng[1], Xin Jin[2], and Satoshi Goto[1]*

[1]Graduate School of Information, Production and System, Waseda University
[2]Information Technology Research Organization, Waseda University
2-7 Hibikino, Wakamatsu, 8080135, Kitakyushu, Fukuoka, Japan
phone: + (81) 90-64286738, email: zyzeng@asagi.waseda.jp, xjin@aoni.waseda.jp, goto@waseda.jp

## ABSTRACT

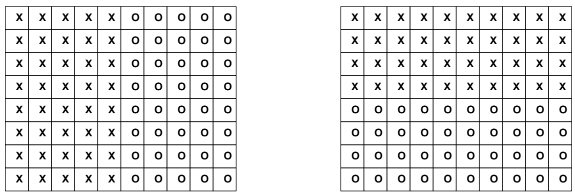
*Frame-compatible format, packing two neighbouring views into one frame, is considered as a promising solution for 3D distribution in the existing system. In this paper, a fast Intra encoder is designed for frame-compatible format coding based on the content similarity and proposed to reduce the computation complexity. With a relative shift, statistics analysis proves the high prediction correlation between the two packed views and qualifies the first coded view provide prediction mode as the reference for the second view prediction. The proposed scheme enables the prediction of the second view's MB only perform candidate Intra modes and candidate directions of each, according to the reference MB in the first view. An average of 75.69% complexity reduction of encoding the second view can be achieved with comparable compression efficiency. With no requirement to change the decoder of existing system and there is little effort for this encouraging format, this proposed algorithm improves the efficiency of frame-compatible format coding, and contributes to the delivery of 3D service.*

## 1. INTRODUCTION

3D video is rendered as the most promising service in the evolvement of video technology, along with the current high definition video delivery. As it can provide people with substantial visual enjoyment by imitating the real world, the delivery of stereoscopic video to household through various channels has become as an irresistibly hot tendency.

Frame-compatible format is a class format based on spatial multiplexing of a pair of stereo views in a single frame [1]. There are various options to arrange the interleaved or packed views in one frame. The most common arrangement schemes are shown in Fig.1, while the packed left and right view in one frame-compatible format is called as plane_0 and plane_1 respectively. Additionally, the two packed views are usually sub-sampled. Especially, the quincunx sampling technique has drawn a considerate attention.

By packing the two views into one frame in one sequence, this format compensates the double bandwidth constraint caused by introducing a second stereo view encoding. The second advantage is that this format is completely compatible to the conventional video codec, such as MPEG-2 [2] and H.264/AVC [3]. Moreover, it contributes to cutting down the tremendous investment to update the equipment and infrastructure for 3D service promotion.



(a) (b)

Figure 1 – Most common frame-compatible format packing arrangement [4]. (a) side-by-side packing arrangement, plane_0 and plane_1 is right and left arranged, respectively, (b) top-bottom packing arrangement plane_0 and plane_1 are up and down arranged, respectively, x indicates MB of plane_0, o indicates MB of plane_1.

Regarding to the current applications approval for frame-compatible format, not only HDMI v1.4 has announced its support [5], but also H.264/MPEG-4 AVC adopts it as the amendment in Supplement Enhancement Information (SEI) [6]. What's more, the ambitious industry has announced their majority of service will base on this format. It is noted that the frame-compatible format has been rendered as one of the most prospective and practical approaches to spread out the 3D video distribution in the existing system.

In this paper, a fast Intra encoder for frame-compatible format based on content similarity in packed views is presented. Considering that Intra coding is the fundamental tool for all the encoders and also widely used in video editing or other video applications, the proposed fast algorithm is designed for Intra coding at first. Due to the short baseline between the two packed views' cameras, the content similarity could be detected. Such similarity could lead to a high correlation in prediction modes, including both partitions and directions. Based on the prediction correlation analysis on the two corresponding macroblocks (MBs) in plane_0 and plane_1 (as that shown in Fig. 1), the candidate sets of prediction modes for plane_1 encoding have been established. The fast Intra prediction scheme is proposed for the MBs in plane_1 by implementing the two Intra prediction partition candidates and the only two candidate directions of each, according to the encoded MB in plane_0 as the reference. Through applying the proposed algorithm solely on plane_1, which is only half of the whole frame, an average of 75.69% complexity reduction of plane_1's encoding can be achieved with comparable compression efficiency.

The rest of this paper is organized as follows. Prediction correlation between the corresponding MBs in the two packed views is analyzed in Section 2. Section 3 is the detailed description of the proposed fast algorithm. Experiment results are presented in Section 4, followed by the conclusion in Section 5.

## 2. ANALYSIS OF PREDICTION CORRELATION

### 2.1 Content Similarity

Since the stereoscopic display illustrates 3-D visual effect from a pair of 2-D views, the frame-compatible format packs the two neighbouring views into one frame after subsampling. Due to the short baseline distance between the two neighbouring cameras, the same or approximately same contents with a left/right shift exist in the two packed views. Such similar content could result to basically the same prediction type including MB partition type and prediction direction. This kind of characteristic is considered as the content similarity. Fig.2 shows one frame of a top-bottom packed frame-compatible format sequence as an example. The two MBs in red contain the same content with a quite short horizontal shift (highlighted by the horizontal part of yellow line). With the shift, when encode the MB in plane_1, the reference MB in plane_0 could be addressed. Regarding to the views which are captured by the cameras in vertical and 2-D arrangement, the shift turns to be vertical and 2-D, respectively.


Figure 2 – Content similarity between the two packed views in top-bottom arrangement.

As the proposed fast algorithm is based on the content similarity in frame-compatible format, the first essential step is to specify the shift between two packed views.

### 2.2 The Relative Shift Obtaining Computation

It is obviously practical to obtain the shift between the left and right view when in hold of the camera parameters and the depth view. However, the ground truth depth image is not available for each source of stereo sequence. Moreover, the generation of depth map from captured multi-view video sequences is very challenging and unstable [7]. Therefore, in our proposal, a fast matching measurement based on the main objects in the first frame is designed, targeting for an approximate and determinate shift to ensure the correlation accuracy among corresponding per-MB. Therefore, we called it shift instead of disparity, although the shift corresponds to the disparity in the stereovision system.

Since the central part of area generally illustrates the main objects, the proposed matching measurement searches the two central rows of corresponding MBs in the two packed views respectively. Fig.3 shows the searching scheme by taking the top-bottom arrangement as an example. The current processing MB in plane_1(bottom) is symbolized as $MB_i$, where $i \in (0, \frac{imagewidth}{16})$. Each pixel in $MB_i$ is designated as $(x_{i_j}, y_{i_k})$, and the top-left pixel is donated as $(x_i, y_i)$. $MB_t$ represents each MB in the central row of plane_0. Each pixel of $MB_t$ is denoted as $(x_{t_j}, y_{t_k})$ and top-left pixel is $(x_t, y_t)$. The $MB_i{}'$ is the MB with the minimum SAD to $MB_i$, according to Eq.1 and Eq.2. And $s(x,y)$ means the value of pixel $(x,y)$. The searching region is $x_i \pm x_r$ in $MB_t$ as Eq.3 shows. $x_r$ suggests the searching range. As the top-left pixel of $MB_i{}'$ is $(x_i', y_i')$, the shift between $MB_i{}'$ and $MB_i$ is defined as $\Delta x_i$, according to Eq.4. For each MB in the central row in plane_1, we retrieve such $\Delta x_i$ of each MB. $\overline{\Delta x}$ is the average value of all the $\Delta x_i$, in which $i$ ranges from 0 to $\frac{imagewidth}{16}$.Thus, $\overline{\Delta x}$ is the relative shift we called in our proposal and used for the fast encoding of all the frames.

$$SAD(MB_i|MB_t) = \sum_{j,k=0}^{15} \left| s(x_{i_j}, y_{i_k}) - s(x_{t_j}, y_{t_k}) \right| \quad (1)$$

$$MB_i' = \arg\min(SAD(MB_i|MB_t)) \quad (2)$$

$$x_t = x_i \pm x_r \quad (3)$$

$$\Delta x_i = x_i - x_i' \quad (4)$$

Regarding to the views captured by the vertical or 2-D arranged cameras, the shift turns to be $\Delta y$ or the pair of $(\Delta x, \Delta y)$ accordingly.

Since the minimum unit size of encoded block is accurate to 4x4, the step of $x_r$ suggests be 4. If it encounters the situation when the relative shift is not an integer multiple of MB size, the solution will be discussed further in Section 3.

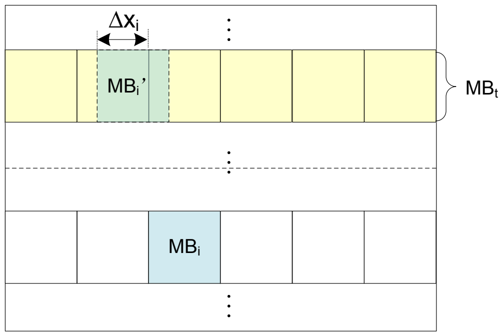



Figure 3 – Matching scheme in top-bottom with horizontal camera arrangement.

### 2.3 Prediction Correlation Analysis

With the help of the relative shift, the prediction correlation between the two packed views is analyzed by retrieving the prediction types encoded from the H.264/AVC compressed bitstreams. Five frame-compatible format sequences packed in top-bottom are generated from: Book_Arrival (view 1 and view 2) [8], Leaving_laptop (view10 and 11) [8], Lovebird1 (view 3 and 4) [9], Alt_Moabit (view 3 and 4) [8] and Newspaper (view 5and 6) [9]. For each sequence, 100 frames are encoded over four QP values.

According to the H.264 standard, there are three Intra MB partition types: Intra_4x4, Intra_8x8, and Intra_16x16. For Intra_4x4 and Intra_8x8, 9 prediction directions are defined, respectively. And for Intra_16x16, there are 4 prediction directions defined. Partition correlation between the corresponding MBs in plane_0 and plane_1 is firstly analyzed through comparing the recorded partitions of each two corresponding MBs, with the help of the relative shift $\overline{\Delta x}$. The occurrence probabilities of plane_1's MB partition types are ranked under the condition when the partition type of $MB_i'$ in plane_0 is determinate. All the probabilities of the partitions shown in Fig.4 are averaged value of 5 sequences' encoding results over 4 QP values.

Fig.4 depicts that the probability of the two corresponding MBs in the two packed views taking the same partition is averaged at 67.29%. In Fig.4(a), when the plane_0 MB takes Intra_16x16, the probability of corresponding MB in plane_1 taking the same partition type is averaged at 76.75%. On the other hand, from Fig.4 (b) and (c), it can be observed that the probability of the partition ranking as the second is relatively large. Therefore, we collect together the first two partitions with higher probabilities, and it can achieve average 88.17% correlation compared with the reference partition.

In view of the probability order of MB partitions in plane_1, the occurrence probability of Intra_16x16 naturally increases along with the encoding QP value increases. On the other hand, the second most probable partition differs a little in high QP from low QP encoding conditions, while the reference partition is Intra_8x8. More precisely, under the condition when the reference MB partition is Intra_8x8, the probability of Intra_16x16 taken by plane_1's MB is 4.67% when QP=22, and 7.69% when QP=37 respectively. Considering the importance of the first two partitions with higher probability, the partition candidates set of our proposed algorithm are established as Table.1shows.

Meanwhile, the prediction direction correlation is analyzed similarly to that of the prediction partition. Each 4x4 block's direction under each partition in both views has been recorded and compared with the determinate reference block direction. Like Fig.5 (a) shows, when the plane_0 blocks' prediction mode is direction 0 (Horizontal direction) of Intra_4x4, each occurrence probability of the 9 directions of Intra_4x4 taken by the corresponding block in plane_1 has been retrieved and ranked. Fig.5 (b) shows the case when the reference blocks in plane_0 take direction 3 (Diagonal_down_left direction) of Intra_8x8, and the probability rank of the 9 directions of Intra_8x8. Based on such kind of conditional probability analysis on plane_1 blocks' directions to the reference direction under the same partition (including 9 directions of Intra_4x4 and Intra_8x8 respectively, and 4 directions of Intra_16x16), the two directions with higher occurrence probability are selected to be the candidate directions for the plane_1's block prediction adopted by the proposed algorithm.

Finally, the two candidate directions for plane_1's block according to the reference direction of plane_0's block are listed in Table.2.

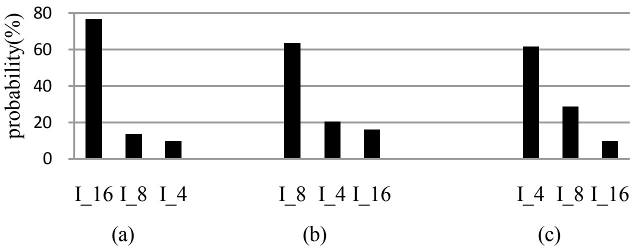



Figure 4 – Average probability of partition correlation between the corresponding MBs in top-bottom arrangement. (a):Plane_1 MB partition modes probability order when reference MB ($MB_i'$) is Intra_16x16. (b):Plane_1 MB partition modes probability order when $MB_i'$ is Intra_8x8. (c): Plane_1 MB partition modes probability order when $MB_i'$ is Intra_4x4.

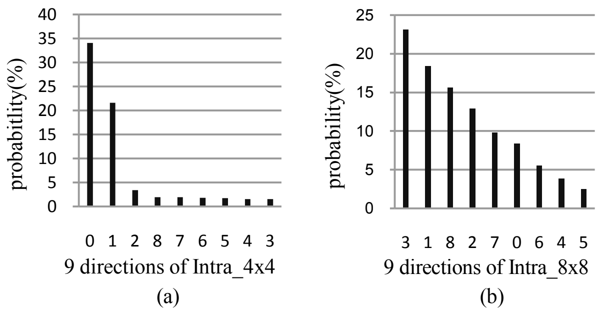



Figure 5 – Examples of direction correlation of top-bottom's corresponding MBs.(a) Probability rank of 9 directions of Intra_4x4 of plane_1 block when reference block in plane_0 is Intra_4x4 direction 0. (b) Probability rank of 9 directions of Intra_8x8 of plane_1 block when reference block in plane_0 is Intra_8x8 direction 3.

Table 1 –Partition candidate set for fast algorithm.

| Plane_0 mode | Candidates for plane_1 MB |
|---|---|
| Intra_4x4 | Intra_4x4, Intra_8x8 |
| Intra_8x8 | Low QP: Intra_8x8, Intra_4x4<br>High QP: Intra_8x8,Intra_16x16 |
| Intra_16x16 | Intra_16x16, Intra_8x8 |

Table 2 – For each reference direction of block in plane_0 (Such as 4MB:0 means direction 0 of Intra_4x4), two candidate directions sharing the same partition with the reference direction are denoted in bracket.

| |
|---|
| 4MB:0 (0,1); 4MB:1 (1,0); 4MB:2 (2,1);<br>4MB:3 (1,3); 4MB:4 (4,1); 4MB:5 (1,5);<br>4MB:6 (6,0); 4MB:7 (1,7); 4MB:8 (1,8); |
| 8MB:0 (0,1); 8MB:1 (1,0); 8MB:2 (1,2);<br>8MB:3 (3,2); 8MB:4 (4,1); 8MB:5 (5,1);<br>8MB:6 (6,1); 8MB:7 (7,1); 8MB:8 (8,1); |
| 16MB:0 (0,3); 16MB:1 (1,3);<br>16MB:2 (2,3); 16MB:3 (3,1); |

## 3. THE PROPOSED ALGORITHM

Depends on the prediction correlation analysis on the corresponding per-block in the two packed views, prediction type candidate sets including partition and direction have been established for plane_1's prediction. Generally, the proposed fast algorithm adds one consistency control module for plane_1 encoding. When it predicts which partition is the best mode for the MB in plane_1, firstly it performs the two candidate partitions prediction successively by retrieving the encoded reference MB's partition in plane_0 (referring to Table.1) through addressing the relative shift. Afterwards, on the direction prediction level, it just executes the two candidate directions prediction by retrieving the direction of the reference block in plane_0 (referring to Table.2). Compared with the original brute-force comparison applied to all the partition types and prediction directions, the proposed algorithm could achieve the target to reduce complexity of plane_1's prediction while trying to maintain the compression performance as much as possible.

The shift obtaining method executes only once before encoding the top-left MB in the first frame to retrieve the relative shift $\overline{\Delta x}$. During the encoding process of plane_0, each 4x4 block's partition type and direction information are recorded. Fig.6 shows the main flowchart of the proposed algorithm.

Due to the shift searching step is defined according to the recording unit 4x4 block, it is feasible for the plane_1 block to retrieve the reference block's direction. However, in the case when the shift is not an integer multiple of 16 pixels (MB size), to retrieve the partition of the reference MB exists an alternative decision. As a result, we assign that the partition type retrieved from plane_0 for $MB_i$ (in plane_1) is the partition type of the MB which has the biggest overlapped area in the corresponding $MB_i'$ in plane_0.

In addition, for the situation when the candidate partition in processing is different from the reference block's partition, the following comes as a discussion. From Table.2, in view of the situation when the partitions of the reference block and candidate block are Intra_4x4 to Intra_8x8 or Intra_8x8 to Intra_4x4. Because the candidate directions are similar to that of reference, in such situation, it is totally feasible to implement the proposed algorithm. Another situation is that when the reference block is in Intra_16x16 and the candidate partition for the corresponding block is Intra_8x8. In spite of the less number of directions in Intra_16x16 to that of Intra_8x8, all the candidate directions are included in the candidate partition according to the standard. So in such situation, it is also no problem to retrieve and predict as the candidate set, despite of the partitions of reference and candidate are different at times. However, under the high QP condition, a problem emerges as that the candidate Intra_16x16's 4 directions cannot cover the 9 directions of reference Intra_8x8. That is to say, if the direction of the reference block is direction 4-8 of Intra_8x8, the candidate directions of Intra_16x16 have to be 0-3, there is no mapping method. In such situation the prediction cannot retrieve any valid instruction. As referring to the statistic probability analysis on such situation, one solution comes as that: when the reference directions are numbered as 4-8 of Intra_8x8 and meanwhile the candidate partition in process is Intra_16x16, it performs direction 1 and 3 of Intra_16x16 prediction. Therefore, such solution qualifies the proposed algorithm implement smoothly according to the candidate sets for all the situations.

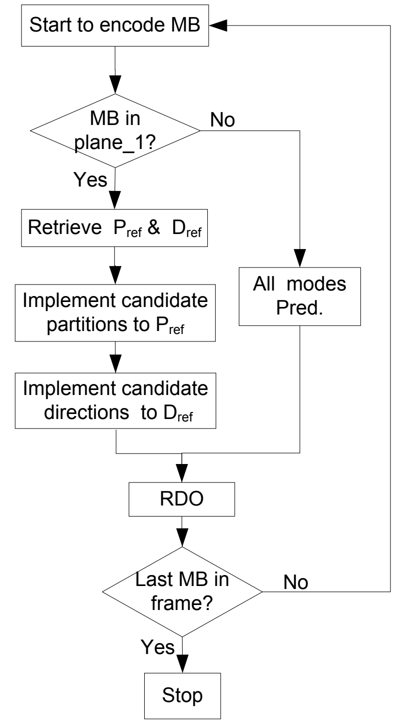



Figure 6 –The flowchart of proposed fast algorithm. $P_{ref}$ means the partition of reference MB in plane_0, $D_{ref}$ refers to the direction of reference block in plane_0.

## 4. EXPERIMENT RESULTS

The proposed consistency control algorithm has been integrated into H.264/AVC reference software JM16.2 [10] for performance evaluation. Four 1024x768 frame-compatible format sequences (Book_arrival's view 1 view 2 (B_1&2) [8], newspaper's view 5 and 6 (NP_5&6) [9], Doorflower's view 7 and 8 (DF_7&8) [8], Alt_Moabit's view 3 and 4 (AM_3&4) [8]), and two 640x480 frame-compatible format sequences (flamenco2's view 3 and 4 (F2_3&4) [11], race1's view 6 and view 7 (R1_6&7) [11]) are generated by packing the neighbouring views in the top-bottom arrangement. 100 frames of each sequence are Intra coded at four QP values (22, 27, 32 and 37) by using High Profile. The proposed algorithm is compared with JM16.2, both are with high complexity rate-distortion-optimization (RDO). The objective compression efficiency is measured by BD-PSNR [12] and the computational complexity reduction is measured by $\Delta Time$, which is given by Eq. (5).

$$\Delta Time = \frac{Time_{JM} - Time_{proposed}}{Time_{JM}} \times 100\% \qquad (5)$$

In Eq.5 $Time_{JM}$ and $Time_{proposed}$ are the total running time which encoded by the original JM16.2 and the proposed algorithm application, respectively. The shift obtaining method implementation time is included in $Time_{proposed}$. In the similar way, $\Delta Time_{p1}$ is short for the time reduction of plane_1's encoding performed by the proposed algorithm compared with that of original JM. $\mathrm{Avg.}\Delta\mathrm{Time}$ and $\mathrm{Avg.}\Delta\mathrm{Time}_{\mathrm{p1}}$ in Table.3 are retrieved and averaged over four QP values for each sequence encoding.

Table.3 shows the experiment results of the proposed algorithm compared with JM16.2 as a benchmark.

Table 3 – Objective performance of the proposed algorithm compared with JM16.2

| Sequence | BD-PSNR (dB) | BD-BR (%) | Avg. ΔTime(%) | Avg. $\Delta\mathrm{Time}_{p1}$(%) |
|---|---|---|---|---|
| B_1&2 | -0.17 | 3.71 | 36.44 | 69.99 |
| NP_5&6 | -0.15 | 2.64 | 32.62 | 66.13 |
| DF_7&8 | -0.22 | 4.65 | 36.16 | 72.09 |
| AM 3&4 | -0.18 | 3.11 | 33.42 | 71.67 |
| F2_3&4 | -0.23 | 3.48 | 39.77 | 86.37 |
| R1_6&7 | -0.12 | 2.01 | 38.98 | 87.90 |
| Average | -0.18 | 3.27 | 36.23 | 75.69 |

Assuming that the computational time to predict the best mode of one MB is the same, the limit ΔTime can achieve at 47.77%, by taking one 1024x768 frame-compatible format sequence with one MB shift as an example. Along with the relative shift increases, the percentage of assumed maximum reduction of time will decrease gradually. Of course, the practical encoding structure not only includes the prediction process, the realistic possible maximum reduction of complexity would be less than what we speculated. As the proposed fast algorithm is mainly designed for plane_1's encoding, the significant reduction of encoding plane_1 shown in the result proves that the proposed algorithm is effective to reduce computational complexity for frame-compatible format coding. In addition, the encoding of plane_0 can certainly absorb other practical Intra fast algorithms, getting integrated to reduce the computational complexity of frame-compatible format coding. As it is out of the scope of this paper, it will not go into details here.

## 5. CONCLUSION

Although frame-compatible format is becoming more and more pervasive, there is little effort focusing on this promising solution for 3D service promotion recently. In this paper, a fast encoder is proposed to reduce the computational complexity of frame-compatible format Intra coding relied on the content similarity. Based on the prediction correlation between the MBs which locate in the two packed views, the prediction of plane_1's MB only performs the candidate Intra modes and candidate directions of each partition according to the reference MB in plane_0. By applying the proposed algorithm solely on half of the packed frame, an averaged 75.69% complexity reduction for plane_1's encoding can be achieved, without any requirement to change the decoder hardware. Statistic analysis is discussed to specify the correlation referring to partition mode and prediction direction between the corresponding per-MB in the two packed views. Furthermore, a matching measurement to obtain the relative shift between the packed views is introduced, instead of seeking for the camera parameters and depth image. The proposed algorithm can be integrated with other fast algorithms, working together to reduce the computational complexity of frame-compatible format coding. And also, this proposal is potentially suitable for Inter coding, which will be investigated further.

## ACKNOWLEDGEMENT

This work is sponsored by a grant from the Core Research for Evolution Science and Technology (CREST), and Japan Science and Technology Agency (JST), JAPAN

## REFERENCES

[1] Anthony Vetro, "Frame Compatible Formats for 3D Video Distribution," in *Proc*. IEEE ICIP 2010, Hong Kong, China, September 26-29, pp. 2405-2408.

[2] Generic Coding of Moving Pictures and Associated Audio Inform.–Part 2: Video, ITU-T and ISO/IEC JTC1 ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2), Nov. 1994.

[3] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC), Mar. 2003.

[4] "Use cases and requirements for the enhancement of MPEG frame-compatible stereoscopic 3D video," ISO/IEC JTC1/SC29/WG11 MPEG 2010/N11681, Oct.2010, Guangzhou, China.

[5] HDMI Licensing, LLC., "HDMI Specification 1.4," 2009.

[6] Gary J. Sullivan, "Draft AVC amendment text to specify Constrained Baseline profile, Stereo High profile, and frame packing SEI message," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, 31st meeting, London, UK, June 18-July 3, 2009.

[7] Dong Tian, Po lin Lai, "View Synthesis Techniques for 3D video," in *Proc*. SPIE 2009, Vol.7443, 74430T.

[8] Ingo Feldmann, Marcus Mueller, Frederik Zilly, "HHI Test Material for 3D Video," ISO/IEC JTC1/SC29/WG11, MPEG2008/M15413, Archamps, France, April 2008.

[9] "Draft Call for Proposals on 3D Video Coding Technology," ISO/IEC JTC1/SC29/WG11, MPEG2010/N11679, Guangzhou, China, Oct, 2010.

[10] JM reference software 16.2, download at http://iphome.hhi.de/suehring/.

[11] "Common Test Conditions for Multiview View Coding," ISO/IEC JT1/SC29/WG11 and ITU-T SG16 Q.6, 20th meeting, Klagenfurt, Austria, July 15-21, 2006.

[12] S.Pateux, J.Jung, "Improvements of Excel macro for BD-gain computation," ISO/IEC JTC1/SC29/WG11, MPEG2009/M16894, Xian, China, Oct.2009.