Fast Motion Estimation Based on Perceptual-Aware for the Depth Map Coding in 3DVC*

PEI-JUN LEE, HSU-HSIEN HUANG AND CHUN-YUAN CHENG

Department of Electrical Engineering
National Chi Nan University
Nantou, 545 Taiwan
E-mail: pjlee@ncnu.edu.tw

This paper proposes a perceptual-aware mode decision algorithm to reduce the coding time of 3D video coding (3DVC) with 2D plus depth maps. The algorithm utilizes a factor called “interesting activity” to determine the interesting region and its candidate coding modes of the depth map. The interesting activity is defined based on the characteristic of visual perception for 3D video viewers, where the characteristics are presented by the existence of MBs with object feature in foreground, or with motion information. In interesting regions, the quality of constructed depth map is maintained; in the non-interesting regions, the computation time of the depth map coding is reduced. In the final step of the algorithm, we use the third dimensional search to improve the quality of the re-constructed depth map. Furthermore, the proposed algorithm is implemented in FPGA to achieve a hardware design. Experimental results show that the entire encoding time is reduced by at least 50.6% compared with the exhaustive mode decision approach, and the structural similarity (SSIM) index is increased by at least 0.0291 and the Peak Signal to Noise Ratio (PSNR) is increased by at least 4.67dB compared with the motion vector (MV) sharing approach. The logic utilization only occupies 8% of the total resource on the Altera DE3-260 FPGA demonstration board.

Keywords: 3D video coding, 3D perceptual attention region, interesting activity, third dimensional search, FPGA

1. INTRODUCTION

There are two critical issues for 3D display technique [1], 3D video quality improvement and 3D video data capacity reduction. Using two cameras in place of the left and right eyes of a human to capture the images is a conventional method for producing a 3D image. However, two or more color bitstream images are required for the transmission and data storage. Figs. 1 (a) and (b) show the 3D video (3DV) codec flowchart [2] for advance stereoscopic processing and auto-stereoscopic display, respectively. In the codec flowchart recommended by 3D video coding (3DVC) standard, the depth estimation is necessary for the encoder as shown in Fig. 1 (a), and then virtual views for users can be synthesized from the color image and its corresponding depth map as shown in Fig. 1 (b). However, the high quality 3D virtual views which are generated from 3DVC standard rely on the high quality of the depth map. Therefore, the standard recommends that the color image and depth map are encoded using the H.264/AVC [3] or HEVC [4] standard, separately.

For compatibility consideration, some literature [2], [3] suggested encoding the color

Received August 28, 2014; revised December 22, 2014; accepted.
Communicated by Jar-Ferr Yang, C.-C. Jay Kuo, Hsueh-Ming Hang , Ming Ouhyoung and Chu-Song Chen.
* This work was supported by the National Science Council of Taiwan and the Ministry of Economic Affairs of Taiwan under Contract No. NSC 101-2221-E-260-032-MY3 and 101-EC-17-A-02-S1-201, respectively.
image and depth map, separately using H.264/AVC. However, H.264 has very heavy computational complexity when using variable block sizes to achieve higher coding efficiency [5]. Therefore, if 3D video coding adopts H.264, the complexity of the bitstream will be increased at least two times more than 2D video coding. To reduce the coding time in the depth map, the MV sharing method [6] takes the MVs and modes from the color images to obtain the motion estimation. However, the characteristics of the depth map are not similar enough to those of the color image thus depth values predicted using the color image MV and coding mode are not precise enough. Fig. 2 shows an illustration where the error is generated in depth map by utilizing MVs and modes from color image. Figs. 2 (a) and (b) are the original image and reconstructed color image by using 3D video codec of “interview”, respectively. Figs. 2 (c) and (d) present the original depth map and reconstructed depth map by using MVs and modes from color image of “interview”, respectively. In Fig. 2 (d), we found an re-motion estimation is needed to improve motion vectors accurately for depth map coding.

Fig. 2. (a) Original color image; (b) Reconstructed color image by using 3D video codec; (c) Original depth map; (d) The reconstructed depth map by using MVs and modes from color image.
The papers [7, 8] worked on accelerating the mode decision process in the depth map using H.264 video coding. They not only estimate the MV and coding mode on the 2D plane (i.e. x and y directions), but also refine the quality of the estimated depth map on the 3D plane (i.e. z direction). However, [7, 8] only utilized the fixed 16×16 block size to perform motion estimation. In our previous research [9, 10], we utilized the variable block size to perform motion estimation in the color image, and found that the depth map should be refined after motion compensation using the MVs and the coding modes of the color image. The paper [9] proposed a third dimensional search to enhance the quality of a depth map with variable block size. However, the accuracy of MVs from the color image is not precise, because the characteristics of the color image are not necessarily the same as those of the depth map. If the predicted MVs from the color image are incorrect, then these methods will produce only limited refinement. In the paper [10], we used the motion activity to determine the types of MB and their candidate Modes. The paper [11] determined the motion activity by utilizing the maximum motion vector length. The motion activity is used to determine whether or not the coding block in the depth map needs to perform re-motion estimation. The paper [12] presents a pixel-based motion estimation scheme assisted with the coded texture video for depth inter-prediction, in view of motion similarity between depth and texture video.

To reduce the coding time of the re-motion estimation with variable block size, the proposed algorithm achieves a fast mode determination to select a coding mode based on interesting activity of each coding macro-block (MB). The proposed algorithm determines the MBs which need to perform re-motion estimation (re-ME) in the depth map using the 3D perceptual attention region determination. The results of the experiment show that the SSIM index is increased at least 0.0291 and the PSNR is increased at least 4.67dB compared with the MV sharing approach, because the proposed 3D perceptual attention region determination can successfully detect whether the coding MBs need to use an MV from the corresponding MV in the color image or to perform the re-motion estimation in the depth map. The entire encoding time is reduced at least 50.6% compared with the exhaustive search approach; because our fast mode determination does not perform an exhaustive search in every MB, and some macro-blocks reuse MVs and modes from color image. The proposed algorithm is implemented in FPGA. The logic utilization only occupies 8% of the total resource on the Altera DE3-260 FPGA demonstration board [16]. The proposed algorithm is suitable for the hardware design.

2. THE PROPOSED FAST MOTION ESTIMATION METHOD FOR DEPTH CODING

To reduce the bit-rate on 3D video transmission and to preserve the depth map quality with low computation time are of most benefit to the 3D encoding engine. This paper proposes a fast mode determination algorithm based on 3D perceptual-aware mode decision to determine the MBs which need to perform re-ME in the depth map coding. To refine the quality of the depth map, this paper applies a third dimensional search (z-search) to the reconstructed depth map. Fig. 3 shows the flowchart of the proposed method. The proposed algorithm utilizes H.264/AVC to encode the color image. On the other hand, in the depth map, the paper proposes a 3D perceptual attention detection to
determine the MBs which need to perform re-ME in the depth map. Then the coding mode candidate generated by the interesting activity selects the coding mode for the re-ME MBs with variable block size, and refines the quality of the estimated depth map in the z direction.

In Visual Perception Systems (VPS), both object features in foreground and motion feature are interesting to human vision. For instance, let the original frame shown in Fig. 4 (a) be considered. The motionless region as shown in Fig. 4 (b) is a low interesting region because it is ignorable to viewers. The region with a low motion and interesting objects in foreground as shown in Fig. 4 (c) is middle interesting region to viewers, the high motion region as shown in Fig. 4 (d) is the most interesting to viewers. This study utilizes the characteristics of visual perception of 3D video viewers to determine the interesting regions in all view videos in order to preserve high video quality in these regions. The categorization of interesting regions depends upon the existence of MBs object feature in foreground, or motion information.

From the 3D visualization, viewers’ attention will be attracted to the object feature in foreground regions which are close to the viewers and have a large depth value. Furthermore, viewers will concentrate their attention on the edge of objects which have large difference in depth value. These edges are referred to as perceptual attention regions.
Therefore, the sum of the depth value and the gradient between the current and previous MB are considered to determine the 3D perceptual-aware regions. The sum of the depth value of a MB \( (S_d) \) is defined in Eq. (1) to determine whether the coding MB is close to the viewer or not, and \( D_i(x, y) \) is the depth value at \((x, y)\) in \(i\)th MB. The distributions of the numbers of MB corresponding to \( S_d/\text{(max } S_d) \) for sequences “Kendo”, “Breakdancer” and “Ballet” are shown in Figs. 5 (a)-(c), respectively. From the distribution figures, \( T_h_1 \) for the (1) is set as \( 0.3 \times \text{(max } S_d) \), where the \( \text{max } S_d \) is defined as \( 16 \times 16 \times 255 \). Then we can find there are 60\% MBs larger than \( T_h_1 \), it means we consider those MBs to be close to the viewer. Eq. (2) defines the gradient of the depth value \( G_d \) in the MB, which is the difference between the depth value of the current MB \( (D_i(x, y)) \) and the previous MB \( (D_{i-1}(x, y)) \). Figs. 5 (d)-(f) are the distributions of the numbers of MB corresponding to \( G_d/\text{(max } G_d) \) for sequences “Kendo”, “Breakdancer” and “Ballet”, respectively. From the distribution figures, \( T_h_2 \) for (2) is set as \( 0.3 \times \text{(max gradient value)} \) where the max gradient value is defined as \( 16 \times 16 \times 255 \). Then we can find there are 30\% MBs less than \( T_h_2 \). If \( G_d(MBi) \) is smaller than \( 0.3 \times \text{(max gradient value)} \), it means the \( MB_i \) belongs to the background region. Combining the above two conditions, we can reach the conclusion that if \( S_d(MBi) > T_h_1 \) or \( G_d(MBi) > T_h_2 \), then the \( MB_i \) belongs to the interesting region.

\[
S_d(MBi) = \sum_{x,y=1}^{16} D_i(x, y) \tag{1}
\]
\[
G_d(MBi) = \sum_{x,y=1}^{16} |D_i(x, y) - D_{i-1}(x, y)| \tag{2}
\]

In additional, the viewer will pay more attention to objects which have fast motion. Therefore, we check the feasible motion vectors (MVs) for depth map coding from the color image by \( SAD \) as defined in (3).

\[
SAD(MBi) = \sum |f_i(x, y) - f_{i-1}(x+s,y+t)| \tag{3}
\]

where \( f_i(x, y) \) and \( f_{i-1}(x, y) \) are the intensity values of the current MB in color image, and the reference MB, respectively. \((s, t)\) is the MV of the color image. If (3) is larger than the threshold \( T_h_3 \), then the MV of the color image might be unreliable for the depth map coding. \( T_h_3 \) is set as \( 0.1 \times \text{(max } SAD \text{ value)} \), where the max \( SAD \) value is defined as \( 16 \times 16 \times 255 \). The distributions of the numbers of MB corresponding to \( SAD/\text{(max } SAD \text{ value)} \) for sequences “Kendo”, “Breakdancer” and “Ballet” are shown in Figs. 5 (g)-(i), respectively. From the distribution figures, we can find there are 30\% MBs larger than \( T_h_3 \). Therefore, the MVs of the color image are unsuitable for use in the depth map coding, and it is necessary to perform re-ME in this region. Otherwise, the MB belongs to non-interesting region, and then MVs and coding modes are share from 3D video coded stream of color image to do ME in depth map.

To reduce the re-ME coding time, this paper proposes a fast mode determination algorithm based on the interesting activity. We establish the MV set \{\( mv_1, mv_2, mv_3, mv_4 \)\} in \{\( MB_1, MB_2, MB_3, MB_4 \)\} for the current macro-block \( MB_0 \) from the adjacent region, where \( mv_i = (MV_{xi}, MV_{yi}) \) is the motion vectors of the corresponding \( MB_i \) for \( i = 1 \) (left), 2 (upper left), 3 (above), and 4 (upper right), respectively. If a MB has multiple MVs for
different partition coding mode sizes, a single MV of this MB is obtained using the bottom-up merging procedure reported in [8]. Let the variance of the motion vector length, \( L \), (i.e. interesting activity) be defined as (4).

\[
L = \sum_{i=1}^{4} \frac{(l(mv_i) - \bar{l})^2}{4}
\]  

(4)

where \( l(mv_i) \) is defined as the length of each MV as in (5). \( \bar{l} \) is the mean value of the length of motion vector set \( \{MB_1, MB_2, MB_3, MB_4\} \)

\[
l(mv_i) = |MV_{x_i}| + |MV_{y_i}|, i = 1, 2, 3, 4
\]  

(5)

Then we will classify the coding MBs into the low, middle, and high interesting activities according to \( L \) in (4). The interesting activity classification and the candidates of the coding mode are shown in Table 1. The coding mode 16×16 is the only one candidate when the MB type belongs to the low interesting activity for \( L \leq 2 \). If \( 2 < L \leq 5 \), the MB type belongs to middle interesting activity, then the candidate coding modes are 16×8 or 8×16. If \( L > 5 \), the MB type belongs to high interesting activity, then the candidate coding mode is P8×8 (i.e., 8×8, 4×8, 4×4).
Table 1. The interesting activity classification and coding mode candidates.

<table>
<thead>
<tr>
<th>$L$</th>
<th>MB Types</th>
<th>Candidate Modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L \leq 2$</td>
<td>Low interesting activity</td>
<td>16x16</td>
</tr>
<tr>
<td>$2 &lt; L \leq 5$</td>
<td>Middle interesting activity</td>
<td>16x8, 8x16</td>
</tr>
<tr>
<td>$L &gt; 5$</td>
<td>High interesting activity</td>
<td>P8x8(i.e., 8x8, 8x4, 4x8, 4x4)</td>
</tr>
</tbody>
</table>

The coding mode and MVs are obtained from the interesting activity classification for the depth map coding. To improve the quality of the reconstructed depth map and to reduce residual information in the bitstream, a third dimensional search method is implemented on the depth map. The main concept of the third dimensional search is shown in Fig. 6.

In Fig. 6, $\text{depth}_{\text{MC}}$ is the compensated depth map via MVs with perceptual based two dimensional motion estimation. The $Z$ directional estimation is the depth value compensation for depth refinement. Similar to the search window on the two dimension motion estimation, there must be a search range on the $Z$ dimension. Here, the search range for $Z$ depth ($\text{SR}_{Z}$) is defined as an interval $[-z, z]$. Each MB with variable block size on the depth map can find the best candidate within $[-z, z]$ which will be added to $\text{depth}_{\text{MC}}$ to minimize the distortion of the depth value in the current frame. Here, $z$ is set as 15. Consequently, fifteen is the third dimensional vector which should be encoded into the bitstream in this case. The three dimensional vector is represented in (6) by $\text{MV}_d(s, t, z')$.

$$\begin{align*}
\text{MSE}(z) &= \frac{1}{16 \times 16} \sum_{x=0}^{15} \sum_{y=0}^{15} (D_i(x, y) - [D_{i-1}(x+s, y+t)+z])^2 \\
\text{z'} &= \arg \min_{z\in[-15,15]} \{ \text{MSE}(z) \}
\end{align*}$$

(6)

where the best candidate is based on the minimum of Mean Square Error (MSE) value. However, there is some difference between (1) and (2). The two terms $D_i(x, y)$ and $D_{i-1}(x+s, y+t)$ represent the depth values on the current block and the compensated block with two dimension vector $\text{MV}_d(s, t)$ in depth map, respectively. $z$ is added to the compensated value to find all the possible MSE values in the search range. Therefore, the third dimensional vector $z'$ can be found with a minimal MSE value. Since the coding mode with variable block size in the color image is also inherited by the depth map. There is more than one $\text{MV}_d(s, t, z')$ for one MB if it is divided into sub-blocks.
In depth coding, we can encode a MB by following by the flowchart in Fig. 7.

![Flowchart for fast mode determination in depth map coding](image)

**Fig. 7.** The proposed fast mode determination for depth map coding.

### 3. THE ARCHITECTURE DESIGN OF THE PROPOSED FAST MOTION ESTIMATION METHOD FOR DEPTH CODING

The hardware architecture for the proposed algorithm is composed of three sub-systems, which are the fast mode determination with the interesting region determination scheme, the motion estimation scheme, and the depth map prediction scheme. There are five circuit modules in the fast mode determination with the perceptual interesting region determination scheme as shown in Fig. 8 which include the Sum_MB_Depth module, the Gradient_MB_Depth module, the SAD_Skip module, the Variance module and the Comparator module. There are eight inputs data (include current color image, current depth image, previous color image, sum of depth value in previous MB, and four neighborhood MVs) and three outputs data (includes sum of depth value for next MB calculation, MB type and candidate modes) in the circuit of this scheme.

In the interesting region determination scheme, it needs to calculate the sum of the depth value \(S_d(MB_i)\) in Sum_MB_Depth module and the gradient between the current and previous MB \(G_d(MB_i)\) in Gradient_MB_Depth module as (1) and (2). Then, the interesting region is determined. Next, the SAD_Skip module is designed to judge the reliability of the MVs of the color image for the depth map coding with SAD as (3). The variance module is designed based on the interesting activity value as (4). The interesting activity classification and the candidates for the coding mode are presented as Table 1. Finally, the perceptual-aware regions are determined by a comparator module and then we use the sum of the depth value, MB interesting activity level and coding mode candidate information to perform the motion estimation.
In the motion estimation scheme as shown in Fig. 9, there are two parts which are variable block size motion estimation (VBSME) module and MV decision module. The function of the VBSME module is to execute the variable block size motion estimation operation on the color image. In this paper, the propagated partial SAD architecture is used to construct the VBSME module. The SAD of all different block sizes is calculated in SAD combination circuit according to the MB type and candidate modes from interesting region determination scheme. This design is a low logic gate count, low latency, and high hardware utilization architecture. After VBSME operation, all SAD values of each of the sub-blocks can be collected into a comparator. Therefore, the coding mode and minima SAD can be obtained for depth map coding. In the variable MV decision module, MV combination circuit is designed to obtain the MV of all different coding mode. Then the MV for coding mode can be selected with variable block sizes.
When the coding modes and MVs (MV_d(s, t)) for the depth map are obtained, the third dimensional search module is utilized to refine the quality of the depth map in the first prediction result. The architecture design is shown in Fig. 10. The function of the third dimensional search module is to execute the refinement operation on the constructed depth using the best MV and coding mode. The proposed third dimensional search structure in this paper is a parallel processing architecture, which can be executed both on the positive direction (Summation of D_1(x+s, y+t) and z) and the negative direction (Subtraction of D_1(x+s, y+t) and z) at the same time to reduce the execution time. We compare the error values (error_add and error_sub) of all kinds of MV_d(z) candidates in each direction. The MV with minima error value is called z_1 and -z_2 in two directions, respectively. The third dimensional MV (z') is selected the one which is with minimum error value in z_1 and -z_2 (min error_add and min error_sub).

**Fig. 10.** The hardware architecture of the third dimensional search module.

4. SIMULATION RESULTS

In the experiments, we consider video-plus-depth sequences in the Advanced Three-dimensional Television System Technologies (ATTEST) [1], where the sequences are “Interview” (720×576, 25fps), “Orbi” (720×576, 25fps), “Ballet” (1024×768, 25fps), “break dancer” (1024×768, 25fps) and “Kendo” (1024×768, 25fps) with 100 frames. We add one more sequence “Dog” (736×528, 25fps), which was created in our laboratory. The proposed and compared algorithms are implemented on 3DV ATM version 6.0 reference software platform [13] based on H.264 coding [14]. For the parameters of video encoding are extended as the profile, the reference frame number is set as 1, and the rest of the parameters are based on the default definition.

The proposed perceptual-aware coding mode determination is compared with the mode decision algorithm by using 3DVC based on H.264 with exhaustive search in Figs. 11 and 12 for different test sequence. In Figs. 11 and 12, (a)-(c) are shown by the modes decision using 3DVC based on H.264 with exhaustive search. (d), (e), and (f) are shown by the proposed perceptual-aware fast mode determination. From the results, we found the proposed algorithm can use small size MB mode to do motion estimation in high
interesting region (foreground with edge texture and fast movement region). Therefore, the proposed algorithm can preserve high quality in this region. In non-interesting region (background), the proposed algorithm uses larger size MB mode such as 16×16 coding mode to do motion estimation and to save computation time on the motion estimation.

In Table 2, there are three comparisons in PSNR, SSIM index [15] and time saving, respectively for four algorithms (3DVC based on H.264 with exhaustive search (3DVC...
(exhaustive search)), exhaustive search with z-search [9], MV sharing [6] and proposed algorithm), where PSNR and SSIM index are used for the evaluation of the reconstructed image in depth map. Coding Timing is used to evaluate the coding time of the both color image and the depth map. The SSIM index is a method of measuring the similarity between two images and the index is more suitable than PSNR in the comparison of the depth quality.

Since the color image and the depth map are independently encoded using 3DVC with the exhaustive motion estimation needs to be performed twice, the results show that this method has the highest execution time. The results in Table 2 show that the proposed method reduces the coding time by at least 50.6% with exhaustive search comparison and achieves almost the same effectiveness of SSIM and PSNR.

The paper [6] proposed a MV sharing approach with a z search method, the advantage of which is fast execution, because it reuses the MVs and modes. However the disadvantage is a lack of precision when there is not a high level of correlation between the features of the color image and those of the depth map. The results in the Table 2 show that the proposed method achieves at least 0.0291 SSIM higher than that of the method [6] and 4.67dB increase in PSNR , but only with a small increase in coding time. The paper [11] proposes a region classification method by using macroblock motion ac-

<table>
<thead>
<tr>
<th><strong>Table 2. The comparisons between the proposed method and other methods.</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSNR Comparison (dB)</strong></td>
</tr>
<tr>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>3DVC (exhaustive search)</td>
</tr>
<tr>
<td>Proposed</td>
</tr>
<tr>
<td>[9]</td>
</tr>
<tr>
<td>MV sharing [6]</td>
</tr>
<tr>
<td>[11]</td>
</tr>
<tr>
<td><strong>SSIM index Comparison</strong></td>
</tr>
<tr>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>3DVC (exhaustive search)</td>
</tr>
<tr>
<td>Proposed</td>
</tr>
<tr>
<td>[9]</td>
</tr>
<tr>
<td>MV sharing [6]</td>
</tr>
<tr>
<td>[11]</td>
</tr>
<tr>
<td><strong>Coding Time Comparison (s/Frame)</strong></td>
</tr>
<tr>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>3DVC (exhaustive search)</td>
</tr>
<tr>
<td>Proposed</td>
</tr>
<tr>
<td>[9]</td>
</tr>
<tr>
<td>MV sharing [6]</td>
</tr>
<tr>
<td>[11]</td>
</tr>
</tbody>
</table>
tivity for a MB. Since [11] does not consider the depth features, the performance of the proposed method achieves at least 0.005 SSIM higher than the method [11] and 0.055dB increase in PSNR. On the other hand, the proposed algorithm can early determine the bypass mode for non-interesting region; therefore the coding time of the proposed algorithm is shorter than that of the method [11].

![Fig. 13](image1.png)

Fig. 13. A visual effect comparison from the Interview test sequence; (a) an original color image; (b) an original depth image; (c) A depth image reconstructed using H.264; (d) a depth image reconstructed using [9]; (e) A depth image reconstructed using [6]; (f) depth image reconstructed using the proposed algorithm.

![Fig. 14](image2.png)

Fig. 14. A visual effect comparison from the Breukdancer test sequence; (a) An original color image; (b) An original depth image, (c) a depth image reconstructed using H.264; (d) A depth image reconstructed using [9]; (e) A depth image reconstructed using [6]; (f) A depth image reconstructed using the proposed algorithm.
Figs. 13 and 14 show two sets of experiment results from the Interview sequence and the Breakdancer sequences, respectively. The original depth map, the H.264 compensated depth map, the algorithms proposed by papers [9], and [6], and the proposed algorithm are shown in the order (b), (c), (d), (e), and (f), respectively. From the experiment result, Figs. 13 and 14 show that the quality of the reconstructed depth map in [9] and [6] cause block effect. Consequently, the distortion between original depth map and compensated depth map is large. Better visual depth quality can be perceived by the proposed method.

In this paper, we present some simulation results in order to demonstrate the accuracy of the circuit functions. The wave simulation was carried out using the Quartus II simulation software and modelsim simulation software which were developed by Altera Corporation. The logic utilization only occupies 8% of the total resource on the Altera DE3-260 FPGA demonstration board [16] as shown in Fig. 15. Therefore, the proposed algorithm is suitable for the hardware design.

![Flow Status](image)

Fig. 15. The proposed hardware architecture of resource usage on the Altera DE3-260 FPGA demonstration board.

5. CONCLUSIONS

This paper has proposed a new solution and achieved hardware implementation to speedup 3D video coding by utilizing perceptual-aware fast mode decision method. The method determines the interesting regions with interesting activity based on the characteristics of visual perception for 3D video viewers. The characteristics depend on the existence of MBs with object feature in foreground, or with motion information. This paper uses the depth and color image information to define those characteristics. The proposed perceptual-aware mode decision algorithm preserves the quality of constructed depth map in interesting regions using small size coding modes. In non-interesting regions, the proposed algorithm uses large size coding mode to save computation time in depth map coding. Finally, the third dimensional search is used to improve the quality of the depth map. In order to evaluation this idea in hardware design, the proposed algo-
rithm has been implemented in Altera DE3-260 FPGA demonstration board. From the simulation results, the proposed algorithm is suitable for the hardware design.

REFERENCES


**Pei-Jun Lee** (李佩君) received the Ph.D. degree from the Department of Electrical Engineering, National Taiwan University in 2004. Since August 2015, she has been Distinguished Professor of the Department of Electrical Engineering, National Chi Nan University, Taiwan. She became a Senior Member of IEEE in 2014. In 2008, she was the recipient of FIRST RUNNER-UP and outstanding advisor of the Macronix Golden Silicon Awards Semiconductor Design and Application Competition from MXIC Co., Ltd. Her research interests include 2D/3D image conversion, 3D/Multi-view video coding, video communication, video compression, image processing, robot fish and FPGA system application.

**Hsu-Hsien Huang** (黃旭賢) received the MS degree from Department of Electronic Engineering National Chi Nan University in 2011. His research interests include 3D video coding, 3D motion estimation, and depth map compression. Now, he is an engineer of Castles Technology, Inc.

**Chun-Yuan Cheng** (鄭鈞元) received the MS degree from Department of Electronic Engineering National Chi Nan University in 2012. His research interests include 3D video coding, 3D motion estimation, and depth map compression. Now, he is an engineer of AU Optronics Corp.