Adaptive Depth Map Filter for Blocking Artifacts Removal and Edge Preserving

Wei Hu, Oscar C. Au, Lin Sun, Wenxiu Sun, Lingfeng Xu, Yujun Li
Department of Electrical and Computer Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
Email: {huwei, eeau, lsunece, eeshine, lingfengxu, liyujun }@ust.hk

***Abstract*—In depth map coding for 3D video coding systems, coding errors in edges can severely affect the synthesis quality. Edge errors mainly compose of two parts: one is blurring and ringing artifact around sharp edge and the other is fake edge caused by blocking artifact. In this paper, we propose an adaptive depth map filter to remove blocking artifacts while preserving depth edges. The proposed filter is designed based on bilateral filter, in which the range kernel parameter is changed adaptively considering the strength of edges and blocking artifacts. Experimental results demonstrate that the proposed depth map filter can achieve up to 0.41 dB gain on the synthesis quality compared to the deblocking filter in MVC at low bit rate.**

I. Introduction

Depth map represents a relative distance of an object from a camera or the origin of 3D space, and has been used to provide 3D scene information in many multimedia applications, such as Kinect [1] by Microsoft. In recent years, the Moving Pictures Experts Group (MPEG) utilizes depth map to define a multi-view plus depth (MVD) data format [2], with the technique of Depth Image Based Rendering (DIBR) to synthesize virtual views for some applications such as free viewpoint television [3].

In 3D video coding systems, after generated by stereo matching algorithm or captured by depth cameras, depth map will be compressed and transmitted. There exist many depth map compression schemes, but in order to keep compatible with conventional video coding standards, such as H.264/AVC, traditional block-based video coding technique is preferred. However, coarse quantization of the block-based discrete cosine transform (DCT) coefficients in traditional video coding will lead to some errors in depth map, especially around edges. Since depth map is not used for display but for view synthesis, and human eyes are sensitive to edges, the accuracy of edges in depth map plays a very important role in view synthesis. Thus, it is crucial to preserve the sharp edges as well as remove fake edges caused by blocking artifacts in depth map.

Although H.264/AVC provides in-loop deblocking filter [4] to remove blocking artifacts, it does not work well in some cases, especially at low bit rate. Oh et al. [5] proposed a depth boundary reconstruction filter considering occurrence frequency, similarity and closeness of pixels. However, a bilateral filter [6] has to be applied after the proposed filtering in order to eliminate some remaining Gaussian noise. Also, they located their proposed filter right after the in-loop deblocking filter in H.264/AVC instead of replacing it, which is a little complicated. Liu et al. [7] proposed a trilateral filter to filter depth map utilizing the proximity of pixel positions, the similarity among depth samples as well as the similarity among the collocated pixels in the corresponding video frame. This is based on the assumption that structure between depth map and corresponding video is similar, which fails when colors on two sides of the boundaries are similar in the corresponding video. Also, neither of the above methods considered blocking artifacts of depth map. An improved loop filter based on H.264/AVC deblocking filter for depth map coding was proposed in [8], in which a 4-tap filter algorithm replaced the one in H.264/AVC, but it does not perform well at low bit rate. In this paper, we develop an adaptive depth map filter, which takes both blocking artifacts removal and sharp edges preserving into account. The proposed filter is simple and effective, especially at low bit rate.

The rest of the paper is organized as follows. In Section II, we describe the proposed depth map filter. Section III presents and analyzes the experimental results. Conclusions are drawn in Section IV.

II. Proposed Adaptive Depth Map Filter

Our proposed algorithm consists of four steps: (1) for a given depth map, compute its corresponding gradient using first order derivative of Gaussian (referred to as Gaussian gradient), which indicates the strength of edges; (2) build a blocking map to indicate the strength of blocking artifacts of the depth map; (3) for every pixel in the depth map, the range kernel parameter of bilateral filter is set according to the blocking map and Gaussian gradient map. If the blocking map indicates that there is blocking artifact at a pixel, then the range kernel parameter at that pixel is set related to corresponding blocking map for blocking artifacts removal. Otherwise, it is related to corresponding Gaussian gradient map for edge preserving; (4) bilateral filter with adaptive range kernel parameter is applied to the depth map, thus realizing the proposed adaptive depth map filter. The remaining four parts of this section explains basic bilateral filter, detection of blocking artifacts in depth map, measure of edge strength and adaptive range kernel parameter respectively.

A. Bilateral Filter

Bilateral filter [6] is a combination of domain filter and range filter, which is an extension to traditional domain filter. It considers not only the geometric closeness between pixels but also the photometric similarity, which is good for edge preserving. Bilateral filter applied to an image $f(x)$ produces an output image defined as follows:

$$h(x) = \frac{1}{k(x)} \int\int f(\xi) c(\xi, x) s(f(\xi), f(x)) d\xi, \qquad (1)$$

with the normalization

$$k(x) = \int\int c(\xi, x) s(f(\xi), f(x)) d\xi, \qquad (2)$$

where $c(\xi, x)$ is the domain kernel which measures the geometric closeness between the center pixel x and a neighboring pixel ξ, and $s(f(\xi), f(x))$ is the range kernel which measures the photometric similarity between x and ξ. An important case of bilateral filtering is shift-invariant Gaussian filtering in both the domain kernel and the range kernel, which can be described as follows:

$$c(\xi, x) = e^{-\frac{|\xi - x|^2}{2\sigma_d^2}}, \qquad (3)$$

$$s(f(\xi), f(x)) = e^{-\frac{|f(\xi)-f(x)|^2}{2\sigma_r^2}}, \quad (4)$$

where σ_d and σ_r is the standard deviation of the domain kernel and the range kernel respectively. σ_d and σ_r play an important role in the effect of filtering. A large σ_d blurs more, because it combines values from more distant locations, while a large σ_r also leads to more blurring for it combines values from more diverse pixel values. In homogeneous regions or around blocking artifacts, large parameter value is appropriate, while in texture regions or around edges, the two parameters should be carefully selected so as not to blur the details.

B. Detection of Blocking Artifacts in Depth Map

At the encoder in H.264/AVC, an integer transform is applied to prediction residues. Then the transform coefficients are quantized to reduce bit rate, resulting in signal loss. In general, the coding errors are larger around the block boundaries than in the middle of the block [4], which leads to blocking artifacts.

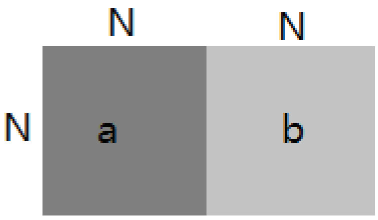

Fig. 1. Two neighboring N by N constant blocks. Pixel intensities within each block are the same and equal to a and b respectively.

Since depth map is nearly piecewise constant, that is, depth map typically consists of constant areas separated by edges due to depth discontinuity, and blocking artifacts look obvious in smooth regions, it is reasonable to assume that obvious blocking artifacts exist at the boundary of two constant blocks, as shown in Fig. 1. This assumption is appropriate especially at low bit rate due to large quantization parameter (QP). If QP is small, this assumption may not hold. However, it can be found by experiments that blocking artifacts are so trivial that they can be neglected when QP is equal to or less than 22. So we can consider blocking artifacts only when QP is larger than 22. We treat the two blocks in Fig. 1 as a whole and take N by 2N DCT transform for it:

$$F(\mu, \nu) = c \sum_{m=0}^{N-1} \sum_{n=0}^{2N-1} f(m, n) \cos[\frac{(2m+1)\mu\pi}{2N}] \cos[\frac{(2n+1)\nu\pi}{4N}], \quad (5)$$

where μ = 0,1,...,N-1, ν = 0,1,...,2N-1, and c is a constant. The result is as follows:

$$F(\mu, \nu) = \begin{cases} \frac{N(a+b)}{\sqrt{2}}, & \mu = \nu = 0 \\ \frac{(-1)^{\frac{\nu-1}{2}}(a-b)}{2\sin(\frac{\nu\pi}{4N})}, & \mu = 0, \nu = 1, 3, ..., (2N-1) \\ & k = 0, 1, ..., N-1 \\ 0, & otherwise \end{cases} \quad (6)$$

Equation 6 shows four properties of DCT coefficients of the combined N by 2N block in Fig. 1 as follows.

1) Non-zero transform coefficients in $F(\mu, \nu)$ matrix only lie in the DC component and the odd columns of the first row.

2) The absolute values of non-zero AC component of DCT coefficients are proportional to the intensity difference $|a - b|$.

3) Non-zero AC component of DCT coefficients are approximately inversely proportional to the frequency ν when $\frac{\nu\pi}{4N}$ is small.

4) Non-zero AC component of DCT coefficients are alternately positive and negative.

This suggests that we can take advantage of the DCT coefficients to represent the strength of blocking artifacts. Property 2 and 3 indicates that the absolute value of the first non-zero AC component of DCT coefficients, i.e., $|F(0, 1)|$, is a good measure of the strength of blocking artifacts, since it is the largest absolute value of all AC components and is proportional to the intensity difference $|a - b|$.

Note that the DCT coefficients of a combined N by 2N block with a sharp edge along vertical direction in the middle also have the above properties. We can distinguish depth edge from blocking artifact by the intensity difference along the boundary between the two blocks. This is because if the abrupt change along the boundary between two N by N blocks is large, then it is very likely to be a sharp edge. Therefore, the boundary won't be detected as blocking artifact when the abrupt change satisfies the following condition (in the case of N = 4).

$$max|block1(Col4) - block2(Col1)| > 0.5 * (2^{\frac{QP-4}{6}}) * \frac{maxI}{255.0}, \quad (7)$$

where block1(Col4) represents pixel intensities in the fourth column of the left 4 by 4 block while block2(Col1) represents pixel intensities in the first column of the right 4 by 4 block. $2^{\frac{QP-4}{6}}$ represents quantization step size and $maxI$ is the maximum intensity value in the depth map. The threshold is set based on the idea that the strength of blocking artifacts should be related to quantization step size as well as the intensity range of the whole depth map. On one hand, the larger QP is, the more severe blocking artifacts are, the larger the threshold should be. On the other hand, the smaller the intensity range is, the less intensity difference along blocking artifacts, the smaller the threshold should be. As to the factor 0.5 in the threshold, it is set according to empirical study.

The above analysis is for blocking artifacts along vertical direction. As to those along horizontal direction, it is analogous.

In conclusion, the detection of blocking artifacts consists of three steps.

1) Take two 4 by 4 neighboring blocks and verify if equation 7 is satisfied or not. If not, go to step 2. Otherwise, we assume the block boundary is not caused by blocking artifacts.

2) Take DCT of the 4 by 8 block combined by the two 4 by 4 neighboring blocks in step 1. If the DCT coefficients satisfy the aforementioned four properties, then the block boundary between the two blocks is detected as blocking artifact.

3) Initialize a blocking map, which is the same size as the depth map, with intensity of all the pixels set as zero. If the block boundary is detected as blocking artifact, then the intensity of pixels in the two columns along the block boundary in the blocking map is assigned to be $|F(0, 1)|$.

C. Analysis of Edge Strength in Depth Map

It is demonstrated in [5] that most coding errors are concentrated in high frequency region such as object boundaries, whereas even small error around object boundaries in depth map may severely degrade the synthesis quality. Therefore, it is very important to preserve edges in depth map.

Inspired by the work in [9], we select Gaussian gradient to measure edge strength in depth map, that is, we firstly apply Gaussian smoothing to depth map and then take the first-order derivative of the smoothed signal. Gaussian gradient is robust to noise because of Gaussian smoothing.

D. Adaptive range kernel parameter

Although bilateral filter owns the property of preserving edges, fixed range parameter σ_r and domain parameter σ_d cannot adapt to the local characteristics of depth map, since texture and edge regions need small σ_r and σ_d so as not to lose details while large σ_r and σ_d is appropriate for homogeneous area and blocking artifacts. To solve this problem, we relate the parameter values to the measurement of blocking artifacts and edge strength analyzed above. In texture and edge regions, since σ_r should be very small so that the range kernel is close to zero, and bilateral filter is the nonlinear combination of domain kernel and range kernel, smoothing barely occurs no matter how σ_d is selected. That is, σ_r dominates the effect of bilateral filter in texture and edge regions. Therefore, we fix σ_d to 1.0 for simplicity.

Fig. 2 (c) and (d) shows the 3-D shaded surface of blocking map and Gaussian gradient map of a 16 by 16 block from a depth map. It can be seen from the blocking map surface that there exists a blocking boundary along one direction, while the Gaussian gradient map surface indicates two edges in perpendicular directions. Comparing Fig. 2(b) with Fig. 2(a), we can see that the marked block contains two obvious edges, of which the horizontal edge is real depth discontinuity while the vertical one is caused by blocking artifacts. That is to say, one edge detected by Gaussian gradient is actually blocking boundary. According to the blocking map, we can distinguish blocking artifacts from real edges, thus setting appropriate σ_r.

Based on the above analysis, σ_r should be larger if blocking map (referred to as blkMap) is larger and smaller if Gaussian gradient (referred to as ggMap) is larger. We found by experiments that below is a good model for σ_r at a pixel with coordinate (i,j):

$$\sigma_r = \frac{k}{gblk}, \tag{8}$$

where

$$gblk = \begin{cases} blkMap^{-2}, & blkMap > 0 \\ ggMap^2, & otherwise \end{cases} \tag{9}$$

and k is a constant, which is set as the maximum value of gblk to guarantee that σ_r is no less than 1.

According to Equation 8 and Equation 9, when blkMap at a pixel is larger than 0, which indicates the corresponding pixel in depth map suffers from blocking artifact, σ_r is proportional to squared blkMap, so stronger blocking artifact leads to larger σ_r, which makes bilateral filter smooth the blocking artifact more. On the other hand, if there is no blocking artifact detected, σ_r is inversely proportional to squared ggMap, so stronger edge leads to smaller σ_r, which prevents bilateral filter from smoothing the edge.

III. Experimental Results and Analysis

The proposed algorithm is evaluated over 5 sequences: Lovebird1, Cafe, Mobile, Newspaper and Balloons. They are firstly encoded using JMVC 8.5 in intra mode with QP 37, during which the in-loop deblocking filter is replaced by our proposed depth map filter, which is not a loop filter. After filtering, virtual views are synthesized by VSRS 3.5 (View Synthesis Reference Software) provided by MPEG. Since depth map is used for view synthesis instead of being displayed directly, the filtering quality is evaluated by synthesis quality employing PSNR metric. Table I shows experimental results of synthesis quality using depth map filtered with the proposed filter and the in-loop deblocking filter in JMVC 8.5 respectively.

TABLE I. Comparison of Synthesis Quality (dB)

Sequences	Left View	Right View	Virtual View	Deblocking Filter	Proposed Filter	PSNR Gain
Lovebird1	6	8	7	38.46	38.71	0.25
Cafe	3	5	4	41.88	42.29	0.41
Mobile	5	7	6	42.78	43.11	0.33
Newspaper	4	6	5	39.56	39.73	0.17
Balloons	3	5	4	34.69	34.81	0.12
Average PSNR Gain : 0.26dB						

Compared with the deblocking filter in MVC, our proposed filter can achieve PSNR gain of 0.25 dB on average and up to 0.4 dB

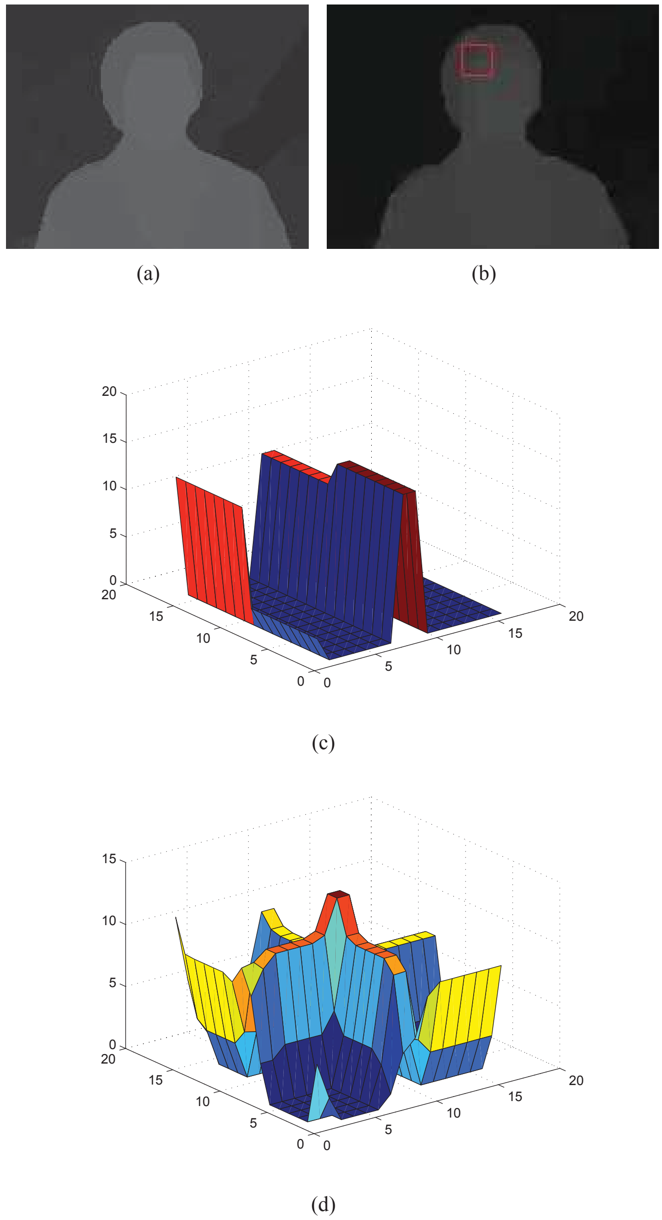

Fig. 2. (a) Part of original depth map of Lovebird1 sequence (view 6, frame 1). (b) Coded result of (a) with QP 37 and deblocking filter off. The region marked by red square is a 16 by 16 block in the depth map. (c) The 3-D shaded surface of blocking map for the marked block in (b). (d) The 3-D shaded surface of Gaussian gradient map for the marked block in (b).

in synthesis quality. Also, we test the proposed filter over different quantization parameter values for Lovebird1 sequence to evaluate the effect of our algorithm under various bit rate. The experimental results are shown in Table II.

Since our method is mainly for highly compressed depth map, the proposed filter is turned off when QP is equal to or less than 22. With higher QP, the proposed filter achieves up to 0.31 dB gain for Lovebird1 sequence.

In spite of the PSNR gain, the subjective synthesis quality with our proposed method is also improved compared to the deblocking filter in MVC, as shown in Fig. 3 and Fig. 4. In Fig. 3(f), we can see that part of the lady's head is eroded, which is the consequence of ringing artifacts around edges in the reconstructed depth maps. This erosion

TABLE II. Synthesis Quality with Different QP (dB)

Lovebird1			
QP	Deblocking Filter	Proposed Filter	PSNR Gain
27	40.03	40.08	0.05
32	39.08	39.39	0.31
37	38.46	38.70	0.24

artifact is avoided in the synthesis image with our proposed method, as shown in Fig. 3(i). Also, depth map filtered by our algorithm is sharper and cleaner than that using the deblocking filter in MVC, leading to cleaner synthesis result, as shown in Fig. 3(i). In Fig. 4, only the left depth map is shown because the white chair is occluded in the right depth map. We can see that the boundary of the white chair has much less noise with our proposed filter.

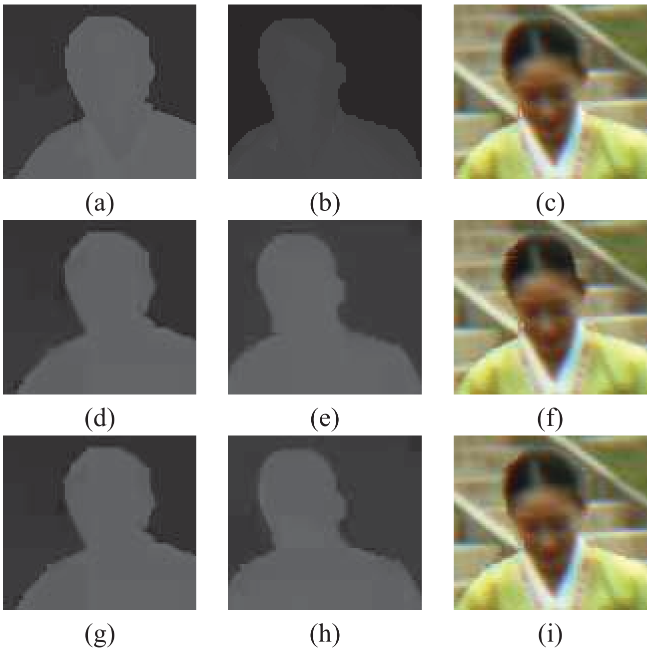

Fig. 3. Original depth maps, reconstructed depth maps coded with QP 37 and synthesis result for Lovebird1 sequence (first frame). (a) and (b) are original left and right reference depth maps. (c) Synthesized image from (a) and (b). (d) and (e) are reconstructed left and right depth maps with deblocking filter in MVC on. (f) Synthesized image from (d) and (e). (g) and (h) are reconstructed left and right depth maps with our proposed filter. (i) Synthesized image from (g) and (h).

IV. Conclusion

In this paper, we proposed an adaptive depth map filter for blocking artifacts removal and sharp edge preserving. The proposed filter relates the parameter value of bilateral filter with the strength of blocking artifacts and edges, which are measured by DCT coefficients and Gaussian gradient respectively, thus making the filter adaptive to local characteristics of depth map. The proposed filter is effective especially at comparatively low bit rate. Experimental results demonstrate that our proposed depth map filter can improve the synthesis quality by 0.26 dB gain on average compared to the deblocking filter in MVC at low bit rate, and the subjective synthesis quality is also better.

V. Acknowledgement

This work has been supported in part by the Research Grants Council (GRF Project no. 610210) and the Hong Kong Applied

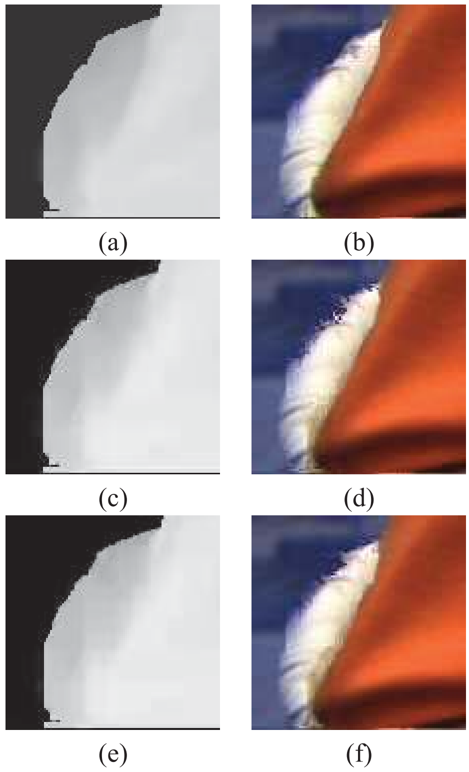

Fig. 4. Original depth maps, reconstructed depth maps coded with QP 37 and synthesis result for Cafe sequence (first frame). (a) Original left reference depth map. (b) Synthesized image from (a). (c) Reconstructed left depth map with deblocking filter in MVC on. (d) Synthesized image from (c). (e) Reconstructed left depth map with our proposed filter. (f) Synthesized image from (e).

Science and Technology Research Institute Project (ART/093CP).

References

[1] Microsoft Corp. Redmond WA. Kinect for Xbox360. 1297,1298.
[2] ISO/IEC JTC 1/SC 29/WG 11, "Committee Draft of ISO/IEC 23002-3 Auxiliary Video Data Representations. Doc. N8038", Montreux, Switzerland, Apr. 2006.
[3] M. Tanimoto, "Overview of Free Viewpoint Television", *Signal Process. Image Commun.*, vol. 21, pp. 454-461, Jul. 2006.
[4] P. List, A. Joch, J. Lainema, G. Bjntegaard, and M. Karczewicz, "Adaptive Deblocking Filter", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 614-619, Jul. 2003.
[5] K.-J. Oh, A. Vetro and Y.-S. Ho, "Depth Coding Using a Boundary Reconstruction Filter for 3-D Video System",*IEEE Trans. Circuits and Systems for Video Technology*, Vol. 21, No. 3, pp. 350-359, Mar. 2011.
[6] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images", *Proc. Sixth Int'l Conf. Computer Vision*, pp. 839-846, Jan. 1998.
[7] S. Liu, P. Lai, D. Tian and C. Chen, "New Depth Coding Techniques With Utilization of Corresponding Video",*IEEE Trans. Broadcasting*, Vol. 57, No. 2, pp. 551-561, Jun. 2011.
[8] N. Zhang and S. Ma, "H.264/AVC-Based Depth Map Sequence Coding Using Improved Loop-filter", *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 312-315, Sept. 2009.
[9] L. Sun and O. Au, "Adaptive Bilateral Filter Considering Local Characteristics", *Seventh International Conference on Image and Graphics*, Aug. 2011.