Post-processing of 3D Video Extension of H.264/AVC for a Quality Enhancement of Synthesized View Sequences

Gun Bang, Namho Hur, and Seong-Whan Lee

Since July of 2012, the 3D video extension of H.264/AVC has been under development to support the multi-view video plus depth format. In 3D video applications such as multi-view and free-view point applications, synthesized views are generated using coded texture video and coded depth video. Such synthesized views can be distorted by quantization noise and inaccuracy of 3D wrapping positions, thus it is important to improve their quality where possible. To achieve this, the relationship among the depth video, texture video, and synthesized view is investigated herein. Based on this investigation, an edge noise suppression filtering process to preserve the edges of the depth video and a method based on a total variation approach to maximum a posteriori probability estimates for reducing the quantization noise of the coded texture video. The experiment results show that the proposed methods improve the peak signal-to-noise ratio and visual quality of a synthesized view compared to a synthesized view without post processing methods.

Keywords: 3D video coding, view synthesis, post-processing, denoising, MAP, total variation.

Manuscript received Sept. 22, 2013; revised Dec. 13, 2013; accepted Dec. 30, 2013.

This work was supported by MSIP (Ministry of Science, ICT & Future Planning), Korea (11-921-02-001, ICT R&D Program).

Gun Bang (phone: +82 42 860 6791, gbang@etri.re.kr) is with the Broadcasting & Telecommunications Media Research Laboratory, ETRI, Daejeon and with the Department of Computer Science and Engineering, Korea University, Seoul, Rep. of Korea.

Namho Hur (namho@etri.re.kr) is with the Broadcasting & Telecommunications Media Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Seong-Whan Lee (swlee@image.korea.ac.kr) is with the Department of Brain and Cognitive Engineering, Korea University, Seoul, Rep. of Korea.

I. Introduction

Owing to the growing need for realistic content, industries related to stereoscopic video are rapidly developing. To provide a more immersive experience to users, multi-view or free-view point video applications, as well as stereoscopic video applications, are becoming popular.

In general, the depth video is used as an aid to generate the synthesized views to be rendered for stereoscopic, multi-view, and free-view point applications. Since a depth video has to be processed with a texture video to generate a synthesized view, the video-plus-depth format is a well-known format for 3D video representation. Furthermore, it can be extended to the multi-view video-plus-depth (MVD) format consisting of multiple texture video and the corresponding depth video to provide 3D video applications such as a free-view point application.

The Joint Collaborative Team on 3D Video Coding Extension (JCT-3V) of both the ITU-T SG16 WP3 Video Coding Experts Group and the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group is developing a standard for an efficient compression of the MVD format. For compression of the MVD format, JCT-3V has extended the existing video codec such as H.264/AVC [1] for MVD coding, and thus the depth video is compressed with the 3D video extension of H.264/AVC [2] as well as the texture video. However, it should be noted that the existing video codec is optimized to encode multi-view video sequences to be viewed by users while the corresponding depth video is not supposed to be viewed by users but is used for generating the synthesized views.

© 2014 ETRI

Therefore, if the 3D video extension of H.264/AVC is used to compress a depth video, the compression artifacts on the depth video generate distortions in the synthesized views.

To solve this problem, two kinds of approaches are generally used: compression techniques that can efficiently encode depth video and post-processing techniques applied to the depth video, that is, compression and reconstruction using an existing video coding technology.

In the first approach, new compression techniques have been developed, including a platelet-based depth video coding algorithm [3], silhouette-based algorithms [4], and a modification of the existing video codec, such as object-based coding of depth video [5], 3D-motion-estimation-based methods [6], and a lossless-compression technique for a depth video [7].

In the second approach, many researchers have proposed suppressing undesirable artifacts seen on the edges of depth videos. To smooth the depth image, a symmetric Gaussian filter [8] and an asymmetric Gaussian Filter [9] have been proposed. Joint bilateral filtering was used to align the depth image with its corresponding texture image [10]. In [11], edge, motion, and depth range information were used to improve the depth estimation in MVD.

Most existing methods in the second approach are focused on reducing noises on the edges, which are object boundaries in the depth video since it is assumed that the accuracy of such edges plays a very important role in view synthesis.

On the other hand, since a synthesized view is generated by the texture and depth videos, the quality of the texture video can affect the synthesized view. When the texture video is coded using existing hybrid video coding algorithms, for example, H.264/AVC, employing a discrete cosine transform (DCT) and quantization methods, a quantization error always occurs, generating artifacts, resulting from a quantization of the DCT coefficients. To improve the quality of the coded texture video, it is therefore important to reduce such quantization errors.

There are two approaches to reducing a quantization error: a spatial-domain approach and a frequency-domain approach. For the spatial-domain approach, a deblocking filter specified in H.264/AVC or HEVC [12], and methods utilizing a Wiener filter [13], [14] as in, in-loop or post-filter, are used. For the frequency-domain approach, statistically-based models of the quantization error have been proposed as a part of different post-processing schemes [15], [16].

In this paper, for a better understanding of the relationships among the depth video, texture video, and synthesized view, the most recently developed compression standard, the 3D video extension of H.264/AVC for MVD, is investigated. Based on this analysis, two post-processing methods; an edge noise suppression filtering process to preserve the edges of the depth video, and a method based on a total variation (TV) approach for a maximum a posteriori (MAP) estimate for reducing quantization noise of the coded texture video, are then proposed.

The remainder of this paper is organized as follows. In section II, the 3D video extension of H.264/AVC for MVD is investigated to analyze the relationship among the depth video, texture video, and synthesized view. Sections III and IV present the proposed post-processing method for enhancing the quality of a synthesized view, and the experiment results, respectively. Finally, some concluding remarks are given in section V.

II. Investigation of 3D Video Extension of H.264/AVC for Multi-view Video Plus Depth

In this section, the encoder structure of the 3D-AVC Test Model (3D-ATM), which is reference software of the 3D video extension of H.264/AVC for MVD, and the coding order for MVD are described. In addition, the relationships among the depth video, texture video, and synthesized view quality are investigated by utilizing 3D-ATM.

1. 3D-ATM Encoder Structure and Coding Order for MVD

The 3D-ATM encoder structure for the enhancement views is shown in Fig. 1. Since 3D-ATM supports backward compatibility to H.264/AVC, texture and depth videos of the base view are encoded by utilizing H.264/AVC coding tools so that they can be decoded using the H.264/AVC decoder [17]. For texture video of the enhancement view, depth-based motion vector prediction (DMVP) and block-based view

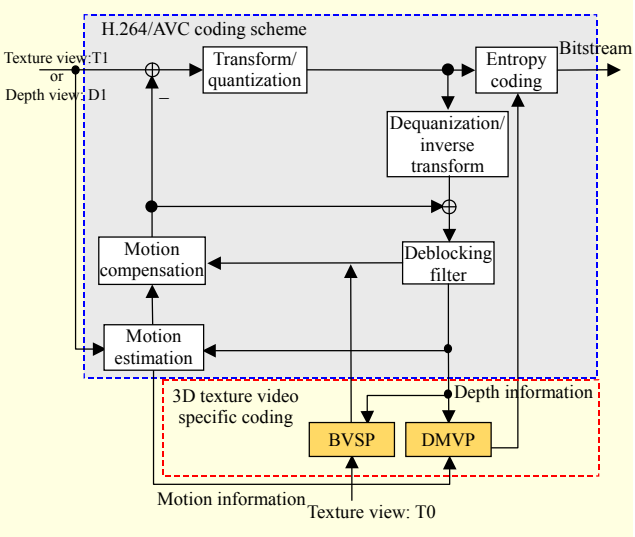


Fig. 1. Encoder structure of enhancement views in 3D-ATM.

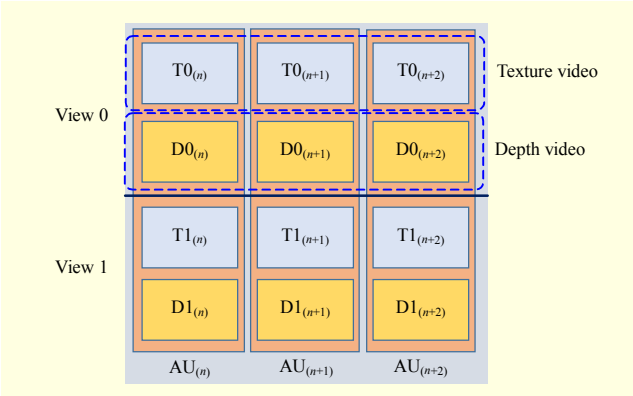


Fig. 2. Definition and coding order of access units.

Table 1. Examples of coding order for AU.

Coding order	Example	Compatibility
Texture first coding order	T0, T1, D0, D1,…	AVC/MVC compatible texture views: T0, T1
Texture first coding order	T0, D0, T1, D1,…	AVC/MVC compatible texture views: T0, T1
Depth first coding order	T0, D0, D1, T1,…	AVC compatible texture view: T0

synthesis prediction (VSP) are employed to improve the coding efficiency. For depth video of the enhancement view, inter-view prediction, which is employed in the Multi-view Video Coding (MVC) extension of H.264/AVC, is used without introducing any new coding tool.

Based on the encoder structure in 3D-ATM, there are various types of coding order. Figure 2 shows an example of the coding order when there are two views. As shown in Fig. 2, an access unit (AU) consists of all components associated with the same output time. Each view includes texture and depth videos. The base view (View 0) consists of the texture video T0, and the depth video D0; and the enhancement view (View 1) consists of the texture video T1, and the depth video D1. In addition, an AU for time n, AU(n), consists of T0(n), D0(n), T1(n), and D1(n). In this example, the applicable coding orders for an AU are specified in Table 1 when considering the provisioning of backward compatibility to H.264/AVC. "Depth-first coding order" is defined as a decoding order, whereby the depth view component is followed by the texture view component for each enhancement view; although the texture view component is followed by the depth view component in the base view. It is well known that the depth-first coding order, in comparison to the other coding orders, provides the better coding performance. This is due to the coding efficiency tools DMVP and block-based VSP, which are used for the texture video coding of the enhancement view [18]. Since DMVP is a coding tool that utilizes information from the depth video to encode the texture video, it can be assumed that the high quality of the depth video, results in the high quality of the texture video when using the depth-first coding order.

2. Relationship among Depth Video, Texture Video, and Synthesized View

In this subsection, the relationship among the depth video, texture video, and synthesized view is investigated using the 3D-ATM with the depth-first coding order. Two experiments are performed. First, the relationship between the depth and texture videos in terms of the coding performance is analyzed. Second, the effect of the coded depth video and coded texture video on the synthesized view is analyzed.

To analyze the relationship between the depth and texture videos in terms of coding performance, the following experiments are performed. Four quantization parameters (QPs) (26, 31, 36, and 41) are used for the texture video coding, and four different QPs (41, 36, 31, and 26) for the depth video coding, each of which are tested for each QP value of the texture video coding, as specified in Table 2. Except for the QP settings, the experiments were performed under the common test conditions of JCT-3V [19], and for the test sequences, Kendo and Balloons, three-view coding was performed. In addition, for a fair comparison, the full-resolution depth video was used as an input of the encoder instead of the half-resolution.

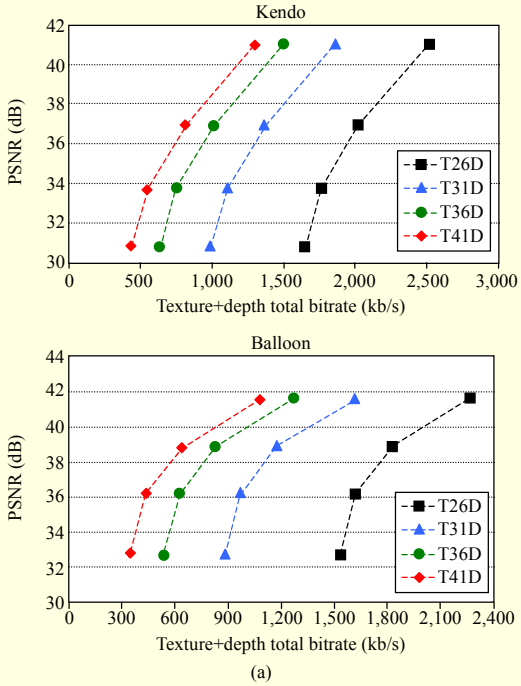
To better visualize the coding performance of the depth video, Fig. 3(a) presents the peak signal-to-noise ratio (PSNR) of the depth video for the total bitrate, which is the sum of the texture video coding bitrate and depth video coding bitrate. For example, the PSNR of depth video increases as the total bitrate increases for the texture video coded with QP 26 of Table 2.

Figure 3(b) presents the ratios of the texture and depth bitrate to the total bitrate when the QP of the texture video coding is 26 and the QP of the depth video coding is 26, 31, 36, and 41.

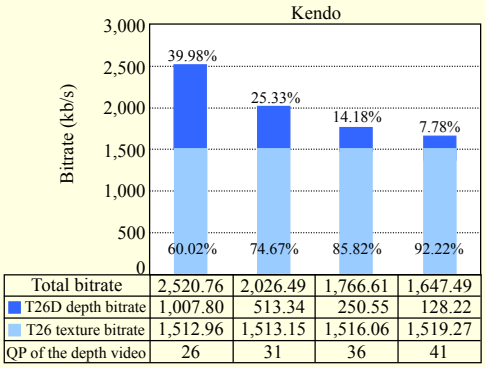
Table 2. QP settings of texture and depth coding.

Coding order	Texture video		Depth video	
	Seq. ID	QP	Seq. ID	QP
Depth first coding order (T0, D0, D1, T1, D2, T2)	T26	26	T26D	41, 36, 31, 26
	T31	31	T31D	41, 36, 31, 26
	T36	36	T36D	41, 36, 31, 26
	T41	41	T41D	41, 36, 31, 26

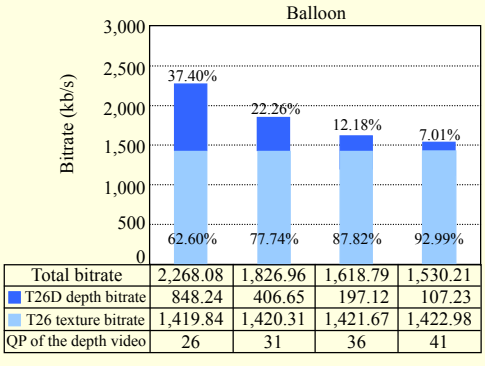
ETRI Journal, Volume 36, Number 2, April 2014
http://dx.doi.org/10.4218/etrij.14.2113.0082



(a)



Total bitrate	2,520.76	2,026.49	1,766.61	1,647.49
T26D depth bitrate	1,007.80	513.34	250.55	128.22
T26 texture bitrate	1,512.96	1,513.15	1,516.06	1,519.27
QP of the depth video	26	31	36	41

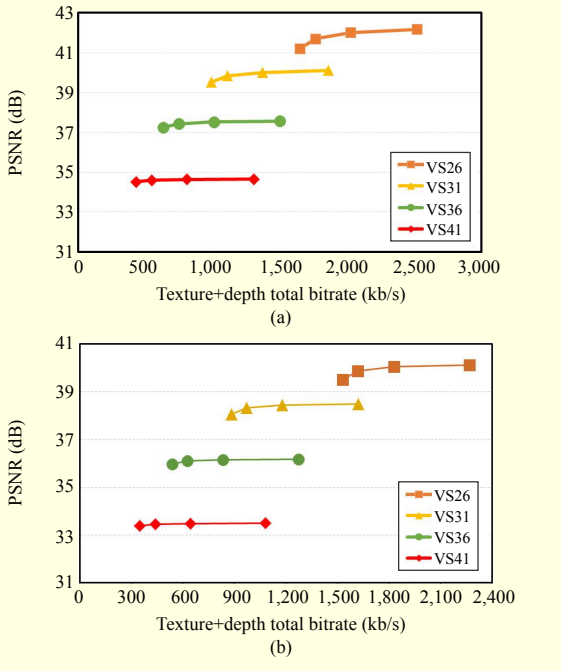


Total bitrate	2,268.08	1,826.96	1,618.79	1,530.21
T26D depth bitrate	848.24	406.65	197.12	107.23
T26 texture bitrate	1,419.84	1,420.31	1,421.67	1,422.98
QP of the depth video	26	31	36	41

(b)

Fig. 3. Experiment results regarding relationship between texture and depth coding: (a) PSNR comparison of depth video vs. total bitrate for four texture QPs and (b) ratios of texture and depth bitrate to total bitrate when QP of texture video coding is 26 and QP of depth video coding is 26, 31, 36, and 41.

According to Fig. 3(b), the depth bitrate increases by up to



(a)

(b)

Fig. 4. Experiment results regarding average PSNR of synthesized views vs. total bitrate: (a) Kendo and (b) Balloons.

39.98% of the total bitrate when the QP value for the depth video coding is 26. In addition, the texture bitrate marginally decreases as the total bitrate increases. The reason for this phenomenon is the accurate depth video coding at a high bitrate. At a higher bitrate, since the depth video can be encoded more accurately, the coding efficiency of the texture video coding increases owing to DMVP and block-based VSP, which utilize the coded depth video. In the first experiment, the visual quality enhancement of the synthesized view could be improved with the higher bitrate depth video in the same QP for texture video coding.

As a second experiment for analyzing the effect of the depth and texture videos on the synthesized view, the six synthesized views were generated using the view synthesis reference software (VSRS) [20] developed by JCT-3V. For the synthesized views, the input sequences are encoded using the QP sets in Table 2 and the average PSNR plots are shown in Fig. 4. In Fig. 4, each solid line represents one of the four average PSNRs of the six synthesized views, which are generated by using the texture video with a fixed QP and the depth video with different QPs (41, 36, 31, and 26). VS26, VS31, VS36, and VS41 represent QPs of the texture video coding of 26, 31, 36, and 41, respectively. According to the experiment results, Fig. 4 shows that the PSNR of the synthesized view is almost constant with a marginal decrease as the QP of the depth video coding increases at a high bitrate. For example, the PSNRs of the VS26 marginally decrease

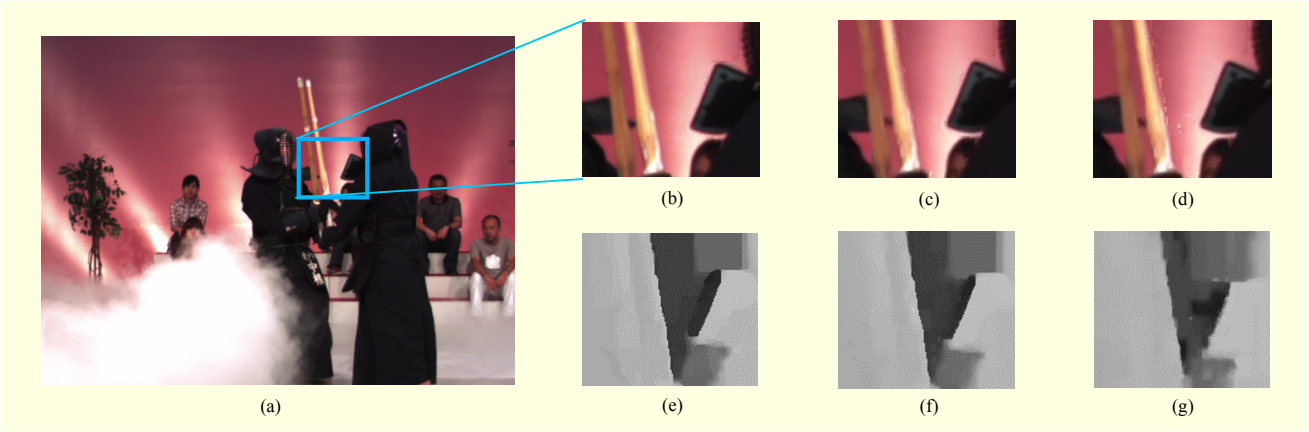


Fig. 5. Example of synthesized view using coded texture video with QP 26 and coded depth video with QP 26, 41: (a) reference synthesized view, (b) uncompressed texture and depth video, (c) texture video, QP=26 depth video, QP=26, (d) texture video, QP=26 depth video, QP=41, (e) uncompressed depth video, (f) depth video, QP=26, and (g) depth video, QP=41.

when the QPs of the depth video coding are increased from 26 to 41. On the other hand, the PSNR of the synthesized view increases as the QP of the texture video coding is decreased from 41 to 26. In addition, the PSNR increment of the synthesized view caused by decreasing the QP of the texture video coding outperforms caused by decreasing the QP of the depth video coding. Therefore, it can be said that the quality of the synthesized view is almost unaffected by the QP of the depth video coding, but is affected by the QP of the texture video coding.

Furthermore, since the PSNR does not always reflect the visual quality, the relationship between the visual quality of the synthesized view and the depth video needs to be further investigated. As a result of our experiment, Fig. 5 shows examples of the synthesized view and the depth video. Figure 5(a) shows the reference synthesized view generated using uncompressed texture video and uncompressed depth video. Figure 5(b) shows a rectangular area of the reference synthesized view specified in Fig. 5(a). The synthesized views of Fig 5(c) are generating using coded texture video with QP 26 and coded depth video with QP 26 and in Fig 5(d), they are generating using coded texture video with QP 26 and coded depth video with QP 41. The QPs used to encode the depth video in Figs. 5(c) and 5(d) are 26 and 41, respectively. Figures 5(e), 5(f), and 5(g), respectively show the depth video corresponding to Figs. 5(b), 5(c), and 5(d).

According to the experiment results, it should be noted that the visual quality of the synthesized view is affected by the distorted edge of the depth video. If there is a severely distorted edge in the coded depth video, as shown in Fig. 5(g), there will be artifacts in the synthesized view, as shown in Fig. 5(c).

Based on an analysis of the experiment results, it can be known that the quality of the texture video coding and distorted edge of the depth video affects the synthesized view. Therefore, to improve the quality of the synthesized view, it is necessary to improve the quality of the coded texture video and reduce the edge noise of the coded depth video.

III. Proposed Post-processing Method for Enhancing Quality of Synthesized View

In this section, based on an analysis of section II, two post-processing methods to enhance the quality of the synthesized view are proposed: an edge noise suppression filtering process of the coded depth video and a method based on a TV approach to a MAP estimate for the noise removal of the coded textual video.

1. Edge Noise Suppression Filtering Process for Coded Depth Video

According to the 3D video extension of H.264/AVC, the depth video is encoded using the tools specified in H.264/AVC without introducing any additional tool. Edge noises are therefore caused by a conventional encoding process. In this paper, to eliminate the edge noise of the coded depth video, an edge noise suppression filter using the coded texture video of the base view is proposed. The proposed method can reduce the edge noises efficiently since the edge of the coded texture video of the base view, is taken into account to accurately detect the edge of the coded depth video.

In Fig. 6, a flowchart of the edge noise suppression filtering process for the coded depth video is described. First, an edge map for the texture video of the base view and an edge map of the code depth video are detected. For the edge map of the texture video, it is always detected at the position of the base view, because it could be assumed that the coded texture of the base view contains the edge information for the view synthesis.

ETRI Journal, Volume 36, Number 2, April 2014
http://dx.doi.org/10.4218/etrij.14.2113.0082

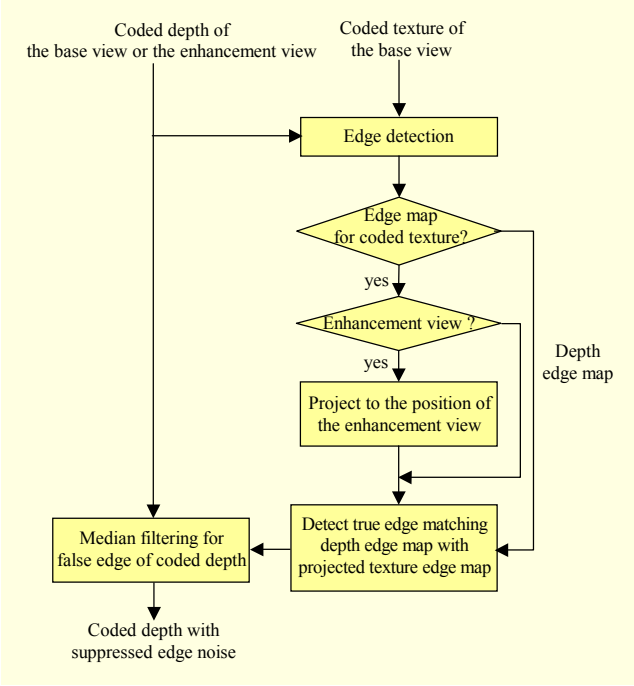


Fig. 6. Flowchart of edge noise suppression filtering process.

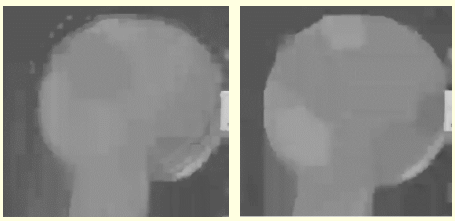


Fig. 7. Effect of proposed edge noise suppression filtering process: (a) coded depth image with QP 36 and (b) filtered depth image in "Balloons" sequence.

Herein, the Sobel operator is used for generating the edge map for texture and depth since it is well known for detecting the edge noise of a depth map. Second, the detected texture edge is projected onto the view position of the coded depth video using 3D-wraping [21], if the coded depth is positioned on enhancement view. Otherwise, the detected texture edge is not required to be projected onto the view position of the depth. Third, if the depth edge is matched with the texture edge of the same view position, it is claimed to be a true edge of the coded depth. If there is a non-matched depth edge, it is claimed to be a false edge. The false edges among the depth edges can be assumed to be noises. The median filter for noise suppression is therefore applied to the false edges.

Figures 7(a) and 7(b) present a coded depth image with QP 36 and a noise-suppressed depth image using the proposed method, respectively. In Fig. 7 (b), it can be seen that the background noise around the edge is suppressed by the proposed edge noise suppression filtering process.

2. Method Using MAP Estimation for Noise Removal of Coded Texture Video

When a texture video is compressed using a DCT transform and quantization methods, which most existing video codecs employ, quantization error/noise exists from a quantization of the DCT coefficients, causing blocking and ringing artifacts. Since these artifacts are one of the main reasons for blurred and fake edges in 3D video codec, they can severely affect the quality of the synthesized view. Therefore, in this paper, a post-processing method based on a TV approach to a MAP estimate is proposed to reduce quantization noise.

A coded texture image can generally be described using the following image formation model.

$$T_{\text{dec}} = T_{\text{pred}} + T_{\text{qresi}}, \tag{1}$$

where T_{dec} is a coded texture image, T_{pred} is a predicted texture image, and T_{qresi} is a quantized residual image. In 3D video codec, T_{pred} is derived from inter-view and collocated depth.

In the general coding process, quantization noise N occurs, which can be defined as follows:

$$N = T_{\text{qresi}} - T_{\text{resi}}. \tag{2}$$

Quantization noise N is a difference image, subtracting the quantized residual image from the residual image, which is the difference between the original texture image and the predicted texture image. The quantized residual image T_{qresi} comes from the process of mapping the value of the residual image onto the representative value in the DCT transform domain.

By (1) and (2), the coded texture image can be described as

$$T_{\text{dec}} = T_{\text{pred}} + T_{\text{resi}} + N = T_{\text{org}} + N, \tag{3}$$

where T_{org} is the original texture image, and is the sum of T_{pred} and T_{resi}.

To apply a MAP estimate for the original texture image when there is noise N in the coded texture image, a Bayesian framework is employed and realized by

$$\begin{aligned} \hat{T}_{\text{orgMAP}} &= \arg\max_T P(T_{\text{org}} | T_{\text{dec}}) \\ &= \arg\max_T \frac{P(T_{\text{dec}} | T_{\text{org}}) P(T_{\text{org}})}{P(T_{\text{dec}})}. \end{aligned} \tag{4}$$

In the MAP estimate of the original texture image, $P(\cdot)$ is a probability distribution. Taking the logarithms and recognizing that the optimization is independent of $P(T_{\text{dec}})$, the problem can be realized by

$$\hat{T}_{\text{orgMAP}} = \arg\min_T \left\{ \log P(T_{\text{dec}} | T_{\text{org}}) + \log P(T_{\text{org}}) \right\}. \tag{5}$$

To solve the optimization problem, a prior for the original image should be restricted in the form of a probability density function; however, it is difficult to select an appropriate prior

for an unknown image. Equation (5) shows that two probability density functions need to be constructed.

The first term is the data fidelity model, which provides a measure of the conformance of the estimated image $\hat{T}_{org}$ for the original texture image to the coded texture image T_{dec} according to the image observation model described in (3). The second term is determined by the probability density function of the original texture image.

Therefore, it is defined that distribution $P(T_{org})$ describes the prior model for the original texture image, and distribution $P(T_{dec}|T_{org})$ indicates the data fidelity model of the coded texture image with quantization noise.

A. Data Fidelity Model for Quantization Noise

Distribution $P(T_{dec}|T_{org})$ can be replaced by distribution $P(T_{qresi}-T_{resi}) = P(N)$, and thus the distribution of the quantization noise is defined by

$$P(N) = \frac{1}{\sqrt{2\pi \cdot |K_N|}} \exp\left\{-\frac{1}{2} N^t K_N^{-1} N\right\}, \tag{6}$$

as a zero-mean Gaussian distribution with auto-covariance matrix K_N.

For auto-covariance matrix K_N, since quantization noise is an error caused by the quantized DCT coefficients in the frequency domain, it can be expressed by a quantized DCT coefficient error. The quantized DCT coefficient can be represented by

$$k_{q(i)}(m,n) = Q_i[k(m,n)] = Q_{\text{Step}(i)} round\left(\frac{k(m,n)}{Q_{\text{Step}(i)}}\right), \tag{7}$$

where, $k(m,n) = \sum_{x=0}^{M-1}\sum_{y=0}^{M-1} H \cdot T_{org}(x,y)$.

The pixel value of the original texture image $T_{org}(x, y)$ is transformed into DCT coefficient $k(m,n)$ by DCT matrix H before quantization. The DCT coefficient $k(m,n)$ is divided by the i-th scale factor $Q_{\text{Step}(i)}$, and the operation round(·) maps the scaled DCT coefficient $k(m,n)$ to the nearest integer.

Therefore, the auto-covariance matrix K_N of (6) can be defined as

$$K_N = H^t Z_N H, \tag{8}$$

where the covariance matrix $Z_N = E[(k - k_q)(k - k_q)^t | k_q]$ and the quantized DCT error variances are δ^2. Applying to an estimated DCT coefficient for k, the covariance matrix Z_N can be calculated with the variance of the quantization DCT coefficient error [22]. In this paper, the uniform distribution is assumed for the model of the quantized DCT coefficient error. The variance of the quantized DCT coefficient error is defined as

$$\delta^2 = \frac{Q_{\text{Step}(i)}^2}{12}, \tag{9}$$

for $-\frac{Q_{\text{Step}(i)}}{2} < k(m,n) - k_q(m,n) < \frac{Q_{\text{Step}(i)}}{2}$.

B. Total Variation Approach for Image Prior

TV was developed to overcome image denoising and restoration, particularly denoising images with piecewise constant features while preserving the edges [23]. The TV approach to MAP estimation uses the geometric property of the texture image and avoids a subjective selection of a prior distribution, $P(T_{org})$, for a MAP estimation.

For the texture image prior, the distribution $P(T_{org})$ is defined as (10) for the considered TV function.

$$P(T_{org}) = \frac{1}{F_{TV}} \exp\left[-f(T_{org})\right], \tag{10}$$

with $F_{TV} = \int \exp[-f(T_{org})]$. Based on Rudin, Osher and Fatemi's model [24], the total variation of image T is defined as

$$f(T) = \int_{\Omega} \sqrt{\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2}\, dxdy. \tag{11}$$

Finally, it can be said that solving the problem of (6) becomes identical to finding the optimized solution of the energy function described in

$$E_{\min}(T) = \min\{\alpha(T) + \lambda f(T)\}, \tag{12}$$

where $\alpha(T) = \frac{1}{2} N^t K_N^{-1} N$ is defined in (6), and parameter λ is a normalizing constant.

IV. Experiment Results

The proposed post-processing methods were implemented in 3D-ATM version 8.0, revision 3. For the experiments, the three test sequences, Balloons, Kendo, and Newspaper with a resolution of 1,024×768 provided in JCT-3V are used, and for each test sequence, three views (left, center, and right views) are encoded with the QP values specified in Table 3. For the experiments, the common test condition described in JCT-3V [19] is applied. For the proposed edge noise suppression of the coded depth video, the size of the median filter is set to 3×3, and for the proposed noise removal of the texture video, the parameter $\lambda = 0.03$ is used for each test sequence.

To measure the performance of the proposed methods, the quality of the synthesized views is used. A total of six synthesized views are generated with three texture images and three depth images, using the VSRS developed by JCT-3V.

ETRI Journal, Volume 36, Number 2, April 2014
http://dx.doi.org/10.4218/etrij.14.2113.0082

Table 3. Comparison of average PSNR: Anchor vs. Proposed.

Sequences	QP for texture and depth coding		Average PSNR for the synthesized views	
	Texture	Depth	Anchor	Proposed (λ=0.03)
Kendo	26	26	42.1723	42.1834
	31	31	40.0082	40.0223
	36	36	37.4259	37.4423
	41	41	34.5161	34.5225
Balloons	26	26	40.1052	40.0679
	31	31	38.4284	38.4247
	36	36	36.1003	36.1182
	41	41	33.3811	33.4010
Newspaper	26	26	38.0475	38.0459
	31	31	36.4464	36.4451
	36	36	34.4328	34.4349
	41	41	32.0619	32.0782
Total average PSNR			36.9272	36.9322

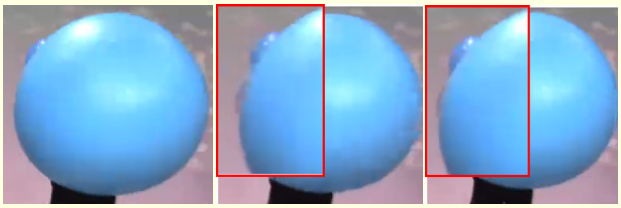


Fig. 8. Example of effect of edge noise suppression filtering in (a)-(c) Balloons sequence: (a) coded texture image with QP 36, (b) synthesized image using coded depth, and (c) synthesized image using noise removal depth.

For an objective comparison, the average PSNRs of the synthesized views are computed with respect to the reference synthesized views generated using the reference texture and depth video. Table 3 shows the average PSNR comparisons between the synthesized views without the proposed post-processing methods, called "Anchor," and the synthesized views with the proposed post-processing methods, called "Proposed." The total average PSNR gain of Proposed is 0.005 dB compared with Anchor. In the Balloons and Newspaper sequences, the PSNR of Proposed is shown to be marginally lower than Anchor. In high QP 26 and 31, the PSNR of the reference has better gain than proposed because the proposed method removes the quantization noise of the reference texture video and sharp edge simultaneously. Therefore, the reference texture video of high QP is made smoother by the proposed method.

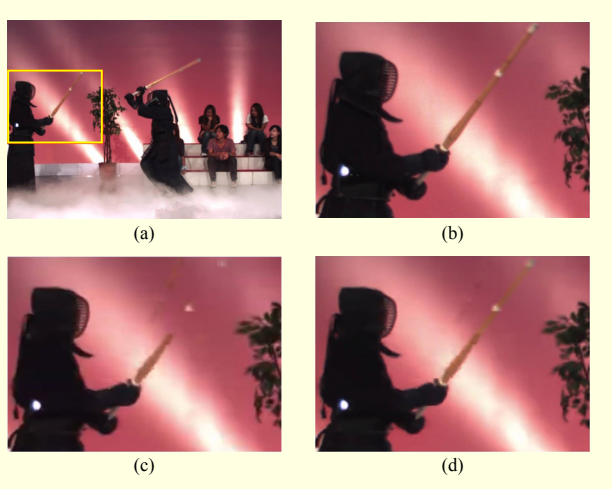


Fig. 9. Visual comparisons of synthesized view in the "Kendo" sequence: (a) reference synthesized view (entire frame), (b) rectangle area of reference synthesized view, (c) synthesized view using texture and depth coded with QP 36, and (d) synthesized view applying proposed method for noise removal.

In addition to the improvement of the average PSNR of the synthesized views with the proposed post-processing methods, it can be seen that the subject visual quality of the synthesized views using the proposed methods is noticeably improved.

Figure 8 demonstrates the effect of the synthesized views using the proposed noise suppression filtering process for the coded depth video. Figure 8(a) is the Balloons coded texture image and Fig 8(b) is the synthesized view image generated by the texture and depth videos coded with QP 36. Although no artifacts appear in the coded texture of Fig 8(a), it can be seen that there is an artifact on the edge boundary of the synthesized view image. The edge noise of the coded depth is represented in Fig 7. Therefore, the artifacts result from the edge noise of the coded depth. As a result of edge noise suppression filtering, the synthesized image is shown in Fig 8(c). The artifacts on the boundary of the Balloon are noticeably reduced owing to the proposed edge noise suppression filtering process. Figure 8 presents an effect of the synthesized views using the proposed noise suppression filtering process for the coded depth video. Figure 8(a) is the Balloons coded texture image and Fig 8(b) is the synthesized view image generated by the texture and depth videos coded with QP 36. Although no artifacts appear in the coded texture of Fig 8(a), it can be seen that there is an artifact on the edge boundary of the synthesized view image. The edge noise of the coded depth is presented in Fig 7. Therefore, the artifacts result from the edge noise of the coded depth. As a result of edge noise suppression filtering, the synthesized image is shown in Fig 8(c). The artifacts on the boundary of the Balloon are noticeably reduced owing to the proposed edge noise suppression filtering process.

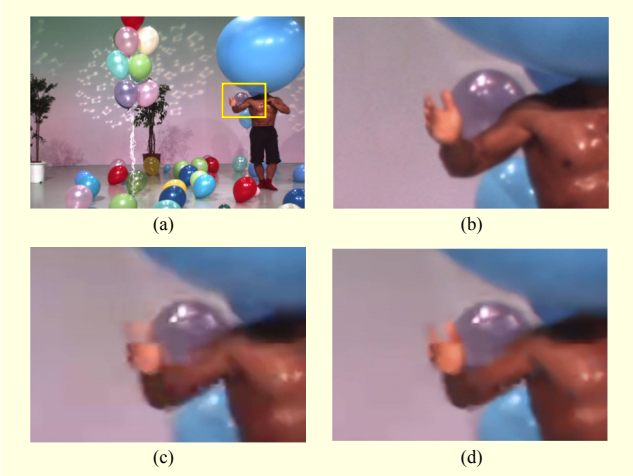


Fig. 10. Visual comparisons of synthesized view in "Balloons" sequence: (a) reference synthesized view (entire frame), (b) rectangle area of reference synthesized view, (c) synthesized view using texture and depth coded with QP 41, and (d) synthesized view applying proposed method for noise removal.

Figures 9 and 10 show examples of the synthesized views using the proposed quantization noise removal method for texture video coded with QP 36 and 41 for Kendo and Balloons, respectively. Figures 9(a) and 10(a) show the reference synthesized views, and Figs. 9(b) and 10(b) are rectangular areas of Figs. 9(a) and 10(a), respectively. Figures 9(c) and 10(c) are the synthesized views generated by the texture and depth videos coded with QP 36 and QP 41, respectively. Figures 9(d) and 10(d) are the synthesized views generated by the coded depth video used in Figs. 9(c) and 10(c), and the coded texture video applying the proposed quantization noise removal method to the coded texture used in Figs. 9(c) and 10(c), respectively. In Figs. 9(d) and 10(d), it can be seen that there are improvements in the visual quality of the synthesized views owing to the proposed quantization noise removal method, while there are severe distortions in the area of the sword and hands of Figs. 9(c) and 10(c), respectively.

V. Conclusion

In this paper, two post-processing methods were proposed to improve the quality of the synthesized view generated by the coded texture and depth videos: an edge noise suppression filtering process for the coded depth video and a method for noise removal of the coded texture video. In the proposed edge noise suppression filtering process, the edge of the coded texture video of the base view is taken into account to accurately detect the edge of the coded depth video. This accurately detected edge of the coded depth video helps to improve the quality of the synthesized video in the edge area when this edge is projected onto the current view position. The proposed method for noise removal of the coded texture video is based on the TV approach to a MAP estimate. The proposed method can effectively reduce the quantization noise of the coded texture video. According to the experiment results, the proposed methods noticeably improve the visual quality as well as the PSNR of the synthesized view compared to the synthesized view without the proposed post-processing methods.

References

[1] ITU-T Rec. H.264 and ISO/IEC 14496-10 *AVC: Advanced Video Coding for Generic Audio-Visual Services*, ITU-T and ISO/IEC JTC 1, May 2003 (and subsequent editions).

[2] M.M. Hannuksela et al., Eds., 3D-AVC Draft Text 7, no. JCT3V-E1002, ITU-T/ISO/IEC JCT-3V, Sept. 2013, work in progress.

[3] Y. Morvan, D. Farin, and P.H.N. de With, "Depth-Image Compression Based on an R-D Optimized Quadtree Decomposition for the Transmission of Multiview Images," *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, USA, vol. 5, Sept. 16 - Oct. 19, 2007, pp. 105-108.

[4] S. Milani and G. Calvagno, "A Depth Image Coder Based on Progressive Silhouettes," *IEEE Signal Process. Lett.*, vol. 17, no. 8, Aug. 2010, pp. 711–714.

[5] D.V.S.X. De Silva, W.A.C. Fernando, and S.L.P. Yasakethu, "Object Based Coding of Depth Maps for 3D Video Coding," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, Aug. 2009, pp. 1699-1706.

[6] B. Kamolrat et al., "3D Motion Estimation for Depth Image Coding in 3D Video Coding," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, May 2009, pp. 824-830.

[7] J. Heo and Y.-S. Ho, "Improved Context-Based Adaptive Binary Arithmetic Coding over H.264/AVC for Lossless Depth Map Coding," *IEEE Signal Process. Lett.*, vol. 17, no. 10, Oct. 2010, pp. 835-838.

[8] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression, and Transmission for a New Approach on 3D-TV," *Proc. SPIE Conf. Stereoscopic Display Virtual Reality Syst. XI*, vol. 5291, San Jose, CA, USA, Jan. 2004, pp. 93-104.

[9] L. Zhang and W.J. Tam, "Stereoscopic Image Generation Based on Depth Images for 3D TV," *IEEE Trans. Broadcast.*, vol. 51, no. 2, June 2005, pp. 191-199.

[10] O.P. Gangwal and R.-P. Berretty, "Depth Map Post-Processing for 3D-TV," *Dig. Int. Conf. Consum. Electron.*, Las Vegas, NV, USA, Jan. 10-14, 2009, pp. 1-2.

[11] E. Ekmekcioglu et al., "Utilisation of Edge Adaptive Upsampling in Compression of Depth Map Videos for Enhanced Free-Viewpoint Rendering," *Proc. 16th IEEE Int. Conf. Image Process.*, Cairo, Egypt, Nov. 7-10, 2009, pp. 733-736.

[12] ITU-T Rec. H.265 and ISO/IEC 23008-2 *HEVC: High Efficiency Video Coding*, ITU-T and ISO/IEC JTC 1, June 2013.
[13] S. Wittmann and T. Wedi, "Transmission of Post-Filter Hints for Video Coding Schemes," *IEEE Int. Conf. Image Process*., San Antonio, TX, USA, vol. 1, Sept. 16 - Oct. 19, 2007, pp. 81-84.
[14] W.-J. Chien and M. Karczewicz, Adaptive Filter Based on Combination of Sum-Modified Laplacian Filter Indexing and Quadtree Partitioning, ITU-T Q6/SG16, doc. VCEG-AL27, London, UK/Geneva, Switzerland, July 2009. Accessed Mar. 2014. http://wftp3.itu.int/av-arch/video-site/0906_LG/VCEG-AL27.zip
[15] H. Paek, R.-C. Kim, and S.-U. Lee, "On the POCS-Based Postprocessing Technique to Reduce the Blocking Artifacts in Transform Coded Images," *IEEE Trans. Circuits Syst. Video Technol*., vol. 8, no. 3, June 1998, pp. 358-367.
[16] D. Sun and W.-K. Cham, "Postprocessing of Low Bit-Rate Block DCT Coded Images Based on a Fields of Experts Prior," *IEEE Trans. Image Process*., vol. 16, no. 11, Nov. 2007, pp. 2743-2751.
[17] D. Rusanovskyy et al., 3D-AVC Test Model 6, JCT-3V document D1003, Incheon, Rep. of Korea, Apr. 2013.
[18] Y. Chen et al., CE2.a related: MB-level NBDV for 3D-AVC, JCT-3V document D0185, Incheon, Rep. of Korea, Apr. 2013.
[19] D. Rusanovskyy, K. Müller, and A. Vetro, Common Test Conditions of 3DV Core Experiments, JCT-3V document D1100, Incheon, Rep. of Korea, Apr. 2013.
[20] G. Tech et al., 3D-HEVC Test Model 4, JCT-3V document D1005, Incheon, Rep. of Korea, Apr. 2013.
[21] G. Tech, K. Müller, and T. Wiegand, "Evaluation of View Synthesis Algorithms for Mobile 3DTV," *Proc. 3DTV Conf*., Antalya, Turkey, May16-18, 2011, pp.1-4.
[22] M.A. Robertson and R.L. Stevenson, "DCT Quantization Noise in Compressed Images," *IEEE Trans. Circuits Syst. Video Technol*., vol. 15, no. 1, Jan. 2005, pp. 27-38.
[23] A.B. Hamza and H. Krim, "A Variational Approach to Maximum a Posteriori Estimation for Image Denoising," *Proc. 3rd Int. Workshop Energy Minimization Methods Comput. Vision Pattern Recognition*, Sophia Antipolis, France, vol. 2134, Sept. 3-5, 2001, pp. 19-34.
[24] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D*, vol. 60, no. 1-4, Nov. 1, 1992, pp. 259-268.

Gun Bang is a senior member of the engineering staff in the Broadcasting and Telecommunication Media Research Laboratory at ETRI, Daejeon, Rep. of Korea. He received his MS degree in computer engineering from Hallym University, Chuncheon, Rep. of Korea, in 1997. He is currently pursuing his PhD degree in the department of computer science and engineering, Korea University, Seoul, Rep. of Korea. He has been an active participant in the ATSC T3/S2 Advanced Common Application Platform Specialist Group since 2002. He was also the secretary of the Telecommunications Technology Association (TTA) PG8062 Working Group, responsible for the standardization of the 3D broadcasting safety guideline from 2010 to 2011. In 2012, he was a visiting scientist in the Advanced Telecommunications and Signal Processing (ATPS) Group of RLE at MIT. He is active in the work conducted by the ITU-T/ISO/IEC Joint Collaborative Team on 3D Video Coding (JCT-3V). His current research interests include video coding, image processing, and computer vision.

Namho Hur received his BS, MS, and PhD degrees in electrical and electronics engineering from Pohang University of Science and Technology, Pohang, Rep. of Korea, in 1992, 1994, and 2000, respectively. He is currently with the Broadcasting and Telecommunications Media Research Laboratory, ETRI, Daejeon, Rep. of Korea. He is the managing director of the Broadcasting System Research Department at ETRI. He is currently a member of the steering committee of the Association of Realistic Media Industry, Seoul, Rep. of Korea. In addition, he has been an adjunct professor with the department of mobile communications and digital broadcasting, University of Science and Technology, Daejeon, Rep. of Korea, since September 2005. For collaborative research in the area of multi-view video synthesis and the effect of object motion and disparity on visual comfort, he was with the Communications Research Centre, Ottawa, ON, Canada from 2003 to 2004. His main research interests are in the field of next-generation digital broadcasting systems, such as the terrestrial UHD broadcasting system, the UHD digital cable broadcasting system, the mobile HD broadcasting system, and backward-compatible 3D-TV broadcasting systems for mobile, portable, and fixed 3D audio-visual services.

Seong-Whan Lee is the Hyundai-Kia Motor chair professor at Korea University, Seoul, Rep. of Korea, where he is the head of the department of brain and cognitive engineering. He received his BS degree in computer science and statistics from Seoul National University, Seoul, Rep. of Korea, in 1984 and his MS and

PhD degrees in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 1986 and 1989, respectively. From February 1989 to February 1995, he was an assistant professor in the department of computer science at Chungbuk National University, Cheongju, Rep. of Korea. In March 1995, he joined the faculty of the department of computer science and engineering at Korea University, Seoul, Rep. of Korea, and is now a tenured full professor. In 2001, he worked as a visiting professor with the department of brain and cognitive sciences, MIT MA, USA. A fellow of the IEEE, IAPR, and Korean Academy of Science and Technology, he has served several professional societies as chairman or governing board member. His research interests include pattern recognition, computer vision, and brain engineering. He has authored more than 300 journal articles and 10 books.

ETRI Journal, Volume 36, Number 2, April 2014
http://dx.doi.org/10.4218/etrij.14.2113.0082