Adaptive Scanning Method for Fine Granularity Scalable Video Coding

Gwang Hoon Park and Kyuheon Kim

One of the recent and most significant technical properties can be expressed as "digital convergence," which is helping lead the technical paradigm into a ubiquitous environment. As an initial trial of realizing a ubiquitous environment, the convergence between broadcasting and telecommunication fields is now on the way, where it is required to develop a scalable video coding scheme for one-source and multi-use media. Traditional scalable video coding schemes have, however, limitations for higher stable picture quality especially on the region of interest. Therefore, this paper introduces an adaptive scanning method especially designed for a higher regional-stable picture quality under a ubiquitous video coding environment, which can improve the subjective quality of the decoded video by most-preferentially encoding, transmitting, and decoding the top-priority image information of the region of interest. Thus, the video can be more clearly visible to users. From various simulation results, the proposed scanning method in this paper can achieve an improved subjective picture quality far better than the widely used raster scan order in conventional video coding schemes, especially on the region of interest, and without a significant loss of quality in the left-over region.

Keywords: Water ring scan method, ubiquitous, fine granularity scalability (FGS).

I. Introduction

The conversion of media from analogue to digital has led to explosive technical developments and provided higher quality services in the individual IT field over the last decade. On the basis of this digital conversion, the IT field faces another new technical trend called "digital convergence" which includes the convergence of diverse digital home appliances and broadcasting and telecommunications. Under this type of paradigm, technology becomes virtually invisible and provides more convenient services to users. Weiser and Tennenhouse describe this paradigm as a ubiquitous environment, where all the services are "user-centric," and thus the techniques intelligently and invisibly adopt and optimize the delivered services for user environments [1], [2]. This characteristic of the ubiquitous environment can be explained as the following 5 C's and 5 Any's: Computing, Communication, Connectivity, Contents, Calm and Anytime, Anywhere, Any-network, Any-device, Any-service [2]. Therefore, the ubiquitous environment allows a user access to any service preferable with anybody at anytime, anywhere, and at any terminal.

One of the possible service scenarios for the realization of the ubiquitous environment is the convergence of broadcasting and telecommunication schemes. In this scenario, a user can watch TV programs through HD or SD TV sets and through PDA or mobile phones no matter where he or she is located. One of the major constraints for realizing this ubiquitous environment is that the conditions of both network bandwidth and terminal capabilities are diverse. For example, higher quality multimedia services can be enjoyed with a powerful computer through ATM or LAN, rather than with a PDA terminal through wireless communication channels. Thus, we need to develop a proper multimedia coding scheme that can transmit

---

Manuscript received Apr. 8, 2003; revised Apr. 29, 2004.

Gwang Hoon Park (phone: +82 31 201 2469, email: ghpark@khu.ac.kr) is with the Department of Computer Engineering, Kyunghee University, Seoul, Korea.

Kyuheon Kim (phone: +82 42 860 5525, email: kkim@etri.re.kr) is with Digital Broadcasting Research Division, ETRI, Daejeon, Korea.

multimedia contents suitable to the individual conditions of terminals and network bandwidths, for example, one-source content that can be efficiently applicable both for mobile telecommunications and digital broadcasting services simultaneously. In order to realize a video coding scheme for one-source and multi-use media, Moving Pictures Experts Group (MPEG) has provided scalable video coding schemes to accommodate the various levels of picture quality depending on the transmission environments and the performances of the user-level terminals [3], and has also put its considerations in a new advanced video coding standard with ITU-T for broadcasting and internet multimedia services [4].

Conventional scalable video coding schemes consist of both a base and enhancement layers. The basic quality of a video is guaranteed by the transmitted bitstream at the base layer, and an improved quality can be obtained by adding the enhancement layer bitstreams to the base layer bitstream. However, the conventional scalable video coding schemes cannot gradually keep a stable picture quality. This is because the step-wise-based conventional scalable video coding scheme cannot diligently keep up with a wide range of bandwidth variations. The conventional schemes, such as both MPEG-2 and MPEG-4 signal-to-noise ratio/spatial/temporal, scalable coding schemes can only restore low-quality image frames in cases where all of the bitstreams from the enhancement layer are not received by the decoder due to wide bandwidth fluctuations. In these cases, the decoder should request the encoder for a retransmission, give up performing image reconstruction until all the bitstreams are received, or carry out an error concealment by using the previous image frames [3], [5]-[12].

In order to avoid the limitation of using step-wise-based enhancement layers for wide bandwidth variations, the MPEG-4 fine granularity scalability (FGS) coding scheme has thus been designed to more robustly distribute the enhanced video bitstreams suitable for the wide variation of a network. The MPEG-4 FGS decoder can accept a truncated version of the enhancement bitstream and reconstruct the quality-improved video sequences by using the partially received bitstream [3], [5]-[10]. According to MPEG verification test results, the performance of FGS has been more satisfying than the conventional scalable coding schemes under certain bandwidth variation conditions [13]-[15]. Although the MPEG-4 FGS is more suitable than conventional scalable video schemes for bandwidth variation, it can be further improved to increase its visual quality in selected regions of interest.

Under a ubiquitous environment, the quality of video sequences is still the important factor; that is, a user expects the same quality of video sequences no matter what terminals are used. This is why various scalable coding schemes including MPEG-4 FGS have been developed. However, it is very difficult to achieve the same quality of video sequences with all the different terminals through various bandwidth conditions. The quality of video sequences can be judged in terms of human preferences rather than in terms of peak signal-to-noise ratio (PSNR) aspects. This is because human beings cannot consider all the pixels in an image frame of a video sequence with the same priority, and thus human beings recognize the quality of a video sequence in terms of its subjective quality rather than of its whole-frame picture quality. It is thus worthy of improving coding efficiency in terms of subjective quality and obtaining visually higher quality video sequences for a viewer. Therefore, this paper introduces a new adaptive scanning method for improving the robustness of enhancement bitstreams in inner video frames in terms of subjective quality, which can thus be considered as a potential technology suitable to the applications for a ubiquitous environment [13]-[15].

This paper consists of the following. Section II addresses an explanation of an MPEG-4 FGS video coding scheme that is a potential technology suitable to a ubiquitous environment. Section III introduces the adaptive scanning method for the improvement of subjective quality, and section IV shows various simulation results, where the current MPEG-4 Advanced Simple Profile (ASP) video coding scheme has been adopted for a base layer, in order to show the adaptation capabilities of the proposed adaptive scanning method. On the basis of these simulation results, the effectiveness of the proposed scanning method is finally indicated in section V.

II. FGS Coding Scheme

A conventional scalable coding scheme has a limitation in that a decoder can safely manage only a video sequence for a base layer when not being able to receive a full enhancement layer bitstream. It is thus difficult to stably guarantee higher picture qualities under a wide network-bandwidth variation. This is why the initial motivation of FGS is to efficiently reconstruct enhanced video sequences even if a decoder can receive only a partial bitstream of the enhancement layers.

In order to reconstruct a video sequence with a partial enhancement layer bitstream, the FGS coding scheme uses the bit-plane-based coding methodology. When transmitting an enhancement layer bitstream from an encoder to decoder, the information on the most significant bit (MSB) bit-plane is most preferentially encoded and transmitted to the decoder, and then a second MSB bit-plane is encoded and sent, and so on, until the last bit-plane, the least significant bit (LSB), is completed for transmission. At the decoder side, the transmitted enhancement layer bitstream is received in the order of bit-planes sent. To transmit this information, an encoder should first of all notify the decoder how many bit-planes would be sent. The transmission is then carried out from the MSB to the LSB on a bit-by-bit basis. If no more bitstream can be transmitted to the decoder due to the bandwidth limitation of the delivery layer, the decoder can reconstruct a similar number of originally encoded information by using the partially decoded bitstream [3], [5]-[10]. Figure 1 shows the encoder structure of the FGS coding methodology based on MPEG-4 Part 2 ASP. As shown in this figure, the encoder structure in the base layer is the same with that of MPEG-4 Part 2 without any change [3].

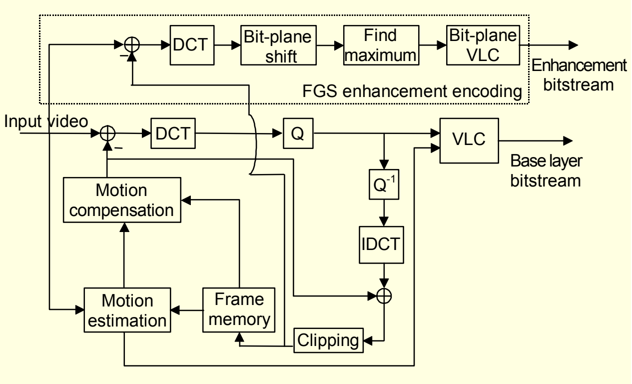


Fig. 1. Encoder structure of the FGS coding methodology based on MPEG-4 Part 2 ASP.

The encoder of the FGS enhancement layer takes both the input image frame and the reconstructed image frame of the base layer as inputs, and then produces the FGS enhancement bitstream as an output. Also, the reconstructed image frame of the base layer is subtracted from the input image frame in order to obtain an image domain residue. The image domain residue is then transformed into discrete cosine transform (DCT) coefficients on 8×8 blocks. The DCT coefficients may be optionally weighted according to the selective enhancement. After collecting all of the weighted DCT coefficients, the maximum value of the coefficients would be found to determine the maximum number of bit-planes. Finally, the bit-plane-based variable length coding (VLC) procedure is then carried out for each bit-plane of each block of the DCT residues. The DCT coefficients obtained on a block-basis according to each bit-plane are inserted in a matrix in a zigzag scan order, and each matrix is run-length encoded. According to the raster scan order, symbols are formed and encoded based on VLC to generate the FGS bitstream. To reconstruct the enhanced image frame at the FGS decoder, the enhancement bitstream is first decoded using bit-plane variable length decoding according to raster scan order. The decoded bit-planes are then shifted based on the selective enhancement shifting factors. The output of a bit-plane shift is the residue of the DCT coefficients. After the inverse discrete cosine transform, the image domain residue is reconstructed and added to the base layer's reconstructed clipped pixels to generate the enhanced image frame [3], [5]-[7], [9], [10].

III. Adaptive Water Ring Method

The current FGS coding scheme uses the raster scan order to encode the enhancement bitstream. Raster scan order is the sequential coding order that starts from the macroblock (MB) located in the upper-left part of the picture down to the MB located in the lower-right part of the picture, and is usually used for moving picture coding schemes. When the network condition guarantees to receive every MB, it shows the improved quality of the entire frame. When it only allows for obtaining a partial bitstream as shown in Fig. 2, however, the picture quality improvements would be achieved only in the upper region of the image frame due to the raster scan order. The quality-improved region might therefore not be the one a user wants to have even if the decoding is successfully carried out with the received MBs.

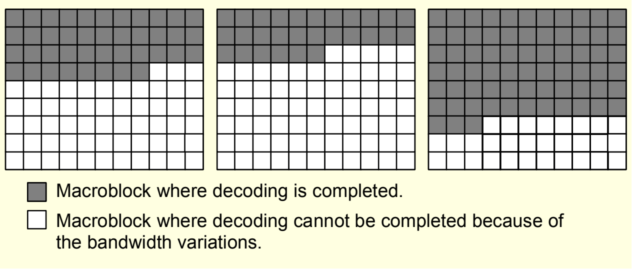
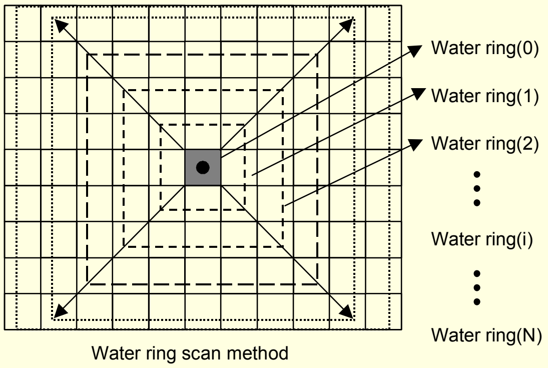
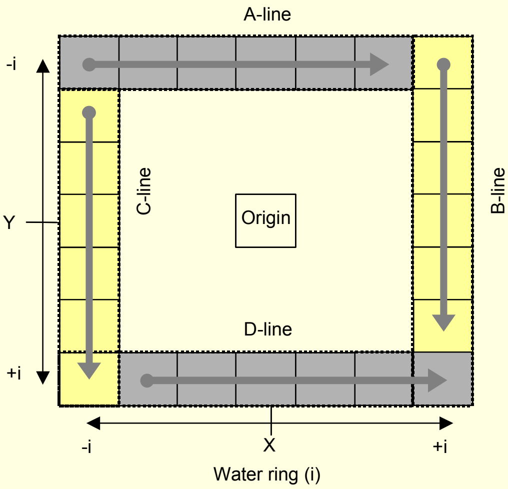


Fig. 2. Macroblocks where decoding is completed when raster scan order is applied in FGS coding.

If an interesting region is most-preferentially encoded, transmitted, and decoded, a user can have a quality-improved video sequence of the interesting region even if all of the MBs have not safely arrived due to truncation caused by bandwidth limitation. The water ring scan method proposed in this paper can simply improve the picture quality of the region of the interest (ROI) in the decoded scalable streaming video. The principle idea of the proposed scanning method is on the basis of water rings expanded from the location where a piece of gravel has fallen on the water surface when it is thrown into a lake. Fig. 3(a) shows the property of the water ring in the lake when applied to an image frame domain. In this figure, each block may represent a pixel, a block, an MB, or a region. When the 'water ring in the lake' principal is adopted to the enhancement layer coding of the FGS, both the encoding and



(a) Basic idea of the water ring scan method



(b) Illustration of the MB coordinate in the *i*-th-generated water ring

Fig. 3. Concept of water ring method.

decoding processes can be started at the location designated by a user, such as when water rings are generated consecutively outward. That is, when a particular origin MB, as shown as the water ring (0) in Fig. 3(a), is decided on as the most preferable, that MB would be first encoded, transmitted, and decoded at the terminal. Next, the nearest neighbor MB's, water ring (1) in Fig. 3(a) surrounded by the origin MB are encoded and transmitted, and the MBs at water ring (i) are consequently encoded and transmitted until every MB in the picture is reached and coded. It is straightforward and very simple to construct. As illustrated in Fig. 3(b), a possible structure of the *i*-th-generated water ring, water ring (i), is composed of the MBs along the A-line (top-horizontal line), B-line (right-vertical line), C-line (left-vertical line), and D-line (bottom-horizontal line). All the MBs in the A-line are those located between –i and +i–1 on the x-axis with –i on the y-axis from a user-decided origin (X, Y). Also, all the MBs in the B-line are those placed in the coordinate between (X+i, Y–i) and (X+i, Y+i–1).

One possible mechanism to implement the concept of the water ring scan method can be described as follows:

```
nRing: # of n-th Water Ring (Water Ring (n)) (initial value = 1)
nMB: # of MB's in each line (A, B, C, D) of Water Ring (n)
OriginX, OriginY: Origin coordinate (x, y)
CurrX, CurrY: Start (upper-left) coordinates of Water Ring (n)
CodedMB(x, y): Encoding or decoding function for designated
    MB(x, y), in which VLC or VLD for each bit-plane are performed
InBoundary(x, y): A checking function to distinguish whether the
    designated MB(x, y) is located at the outside of the picture
Step 1: Coding of the 'Origin'
  if (InBoundary(OriginX, OriginY)) CodedMB (OriginX, OriginY);
Step 2: Coding 'A-line' of Water Ring (n)
  CurrX = OriginX – nRing;
  CurrY = OriginY – nRing;
  nMB = 2 * nRing;
  for(j=0; j< nMB; j++)
      if(InBoundary(CurrX+j, CurrY)) CodedMB(CurrX+j, CurrY);
Step 3: Coding 'B-line' of Water Ring (n)
  x = CurrX + nMB;
  for(j=0; j < nMB; j++)
      if(InBoundary(x, CurrY+j)) CodedMB(x, CurrY+j);
Step 4: Coding 'C-line' of Water Ring (n)
  for(j=1; j < nMB; j++)
      if(InBoundary(CurrX, CurrY+j)) CodedMB(CurrX, CurrY+j);
Step 5: Coding 'D-line' of Water Ring (n)
  y = CurrY + nMB;
  for(j=1; j < nMB; j++)
      if(InBoundary(CurrX+j, y)) CodedMB(CurrX + j, y);
Step 6: Check finish
      if (not Frame Coded) {nRing++; goto Step 2;}
      else   Stop;
```

It does not have to be transmitted to a decoder if the location of the origin is known as a fixed location, such as a center of the picture (which we call, the fixed-centered water ring), by both the encoder and decoder. In addition, the origin of the water ring method can be flexibly determined by a user (which we call, the arbitrarily-centered water ring). To use the arbitrarily-centered water ring scan method, the encoded bitstream header should include an additional bit flag for each video object layer (VOL) header to notify the usage of the arbitrary water ring scan order (arbitrary_center_enable flag), and another 14 bits for each FGS VOP header to specify the start point of scanning, 7 bits for both horizontal (arbitrary_center_x flag) and vertical (arbitrary_center_y flag) MB positions (in cases where the picture size is 2000 × 2000 pixels at maximum). Then, the decoder checks the arbitrary_center_enable flag in the VOL header to identify the usage of the arbitrarily-centered water ring scan method. If that flag is set to 1, then the decoder reads both the horizontal and vertical coordinates of the MB position of the origin from the FGS VOP header and then starts the bit-plane-based decoding from that position. If that value of the flag is 0, then the MB position of the origin is set to be the center of the image frame, thus followed by usage of the fixed-centered water ring scan method. The corresponding syntax changes are shown in Table 1. Various methodologies may be used to determine the MB position of the origin of the water ring. In this paper, the mouse tracking is simply used while displaying the original input image in the encoding process. The encoding operator just traces the important object using the mouse while viewing the video sequences. Then, the nearest MB position of the designated pixel is used as the origin of a water ring scan, and that MB position is sent to the decoder to synchronize the starting MB position. An example of the observed variations of the origins while encoding for a Stefan CIF-format (total 22 × 18 MBs) sequence is illustrated in Fig. 4.

To obtain a quality-improved region with high priority, the current MPEG-4 FGS uses the selective enhancement (SE) method [6] to more accurately transmit an important ROI area according to the raster scan order. As shown in the left diagram in Fig. 5, if the bitstream is truncated due to network bandwidth variations, only the MBs in the upper region of the image frame can be enhanced, and thus the enhanced quality for the macroblocks in the whole SE area may not be obtained. This phenomenon may be overcome by using the water ring scan method instead of using the raster scan order. Figure 5 shows how the water ring scan method can improve the picture quality of the SE area. As shown in this figure, the MBs located in the SE area can be most preferentially encoded and decoded with a higher priority in each bit-plane. As a result, we found that the water ring scan method inherently makes the SE method more efficient. The brilliant concept of the SE method may have problems in the actual implementation of the encoding process. In the encoding process, the SE area should be independently decided for each image frame. In addition,

Table 1. Syntax change for arbitrarily-centered water ring scan method.

Layer	Syntax	No. of bits
Video object layer	arbitrary_center_enable	1
FGS video object plane	arbitrary_center_x	7
	arbitrary_center_y	7

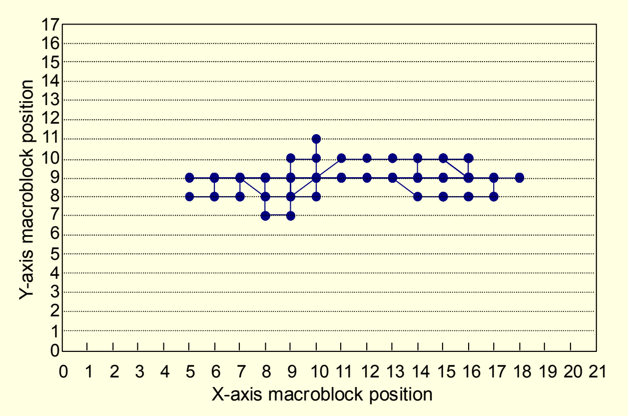


Fig. 4. Example of observed MB positions of water ring origin for Stefan CIF-format sequence.

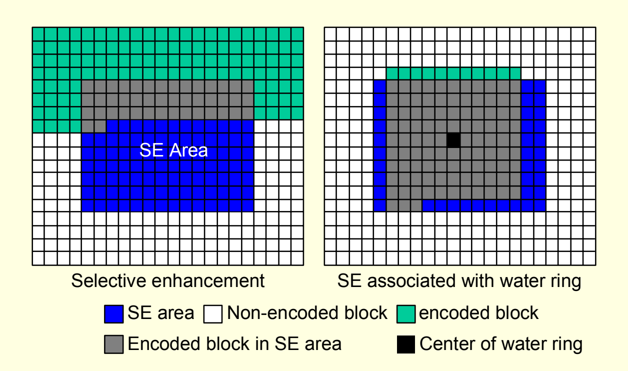
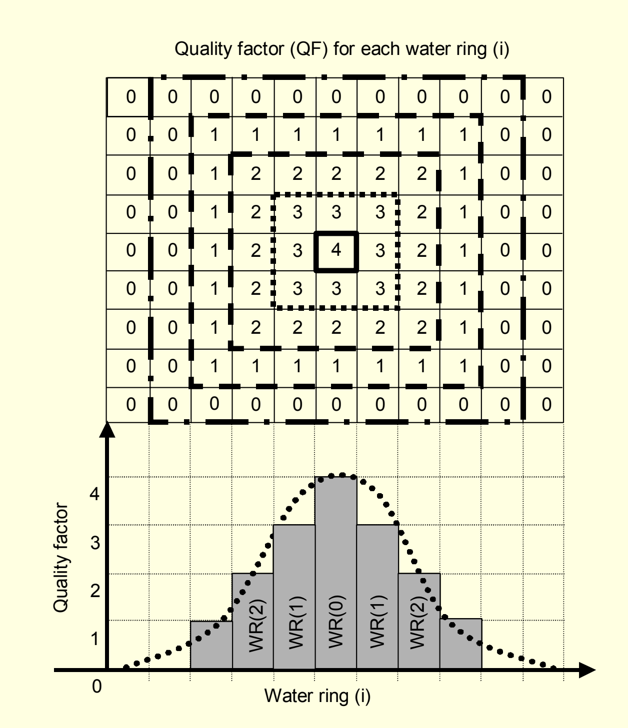


Fig. 5. The relationship between SE area and the water ring scan method.



0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	0	0
0	0	1	2	2	2	2	2	1	0	0
0	0	1	2	3	3	3	2	1	0	0
0	0	1	2	3	4	3	2	1	0	0
0	0	1	2	3	3	3	2	1	0	0
0	0	1	2	2	2	2	2	1	0	0
0	0	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0

Fig. 6. Adaptive water ring method associated with quality factors.

the shift factors of the MBs inside the SE area should be simultaneously determined. The assigned bits used to transmit the shift factor for each MB may be a very large amount if the SE area is relatively large, therefore those bits may degrade the decoded picture quality because the actual bits for reconstructing the image information may be reduced after sending the data for the MB-based shift factors. Also, the required manual workload of the encoding operator may be very high, and the encoding procedure will be refined by taking the following into consideration:

i) The top priority ROI regions that should be clearly visible to users must be determined.
ii) Top-priority ROI regions can be decided by using an SE-type methodology. This is accomplished using the bit-plane shift to the left of the corresponding pixel information.
iii) It is very effective for the actually decoded image if the information of the top-priority ROI region is sent on the preferential basis.
iv) The water ring scan method is the optimal solution for sending the ROI region preferentially.

A design for a simplified self-operating method that can carry out the items i) to iv) is strongly recommended. The adaptive water ring method, as illustrated in Fig. 6, is proposed to satisfy the contents of i) to iv). The quality factor (QF) for each water ring, QF (i) for water ring (i), is introduced in this adaptive method. Figure 6 describes the relationship between the water ring and its corresponding QF values that is based on the significance of the ROI region, which we call the QF-based water ring. QF values can be effectively used for priority-based video coding methodologies. They may be used as the weighting factors for adaptive quantization for ROI regions [22], [23]. In this paper, the QF value is used in exactly the same way as the shift factor used in the SE method. As shown in this figure, the maximum QF value is given as 4, as mentioned in the MPEG-4 FGS international standard [3], [5]-[7]. The data in the origin MB of the water ring is assumed to be the most significant, therefore the maximum QF value 4 is assigned to water ring (0), 3 to water ring (1), 2 to water ring (2), 1 to the QF of water ring (3), and 0 to the water ring beyond water ring (3). Thus, the image is encoded based on the given QF (i) along each water ring (i). Decoding can also be performed based on the QF values of each water ring, whose values may be predetermined (agreed upon between the encoder and decoder) or transmitted from the encoder to the decoder. Accordingly, the image data in the ROI area can be transmitted and received on a top priority using the water ring method, and the quality of the ROI area can be adaptively improved by the QF values assigned to the water rings. Thanks to the QF-based adaptive water ring method, the Human Visual System (HVS) based visual coding can be naturally embodied. In other words, we can easily realize that the ROI region can be sent with top-priority using a guaranteed preferential decoding, and that the picture quality of the outskirt of the image frame, whose regions are little noticed by the viewers, may be gradually degraded in cases where network bandwidth is not sufficient.

Six types of adaptive water ring methods can be implemented in the FGS coding scheme, as shown in Table 2.

Table 2. Various types of the adaptive water ring methods.

Type	Center position		Quality factor (QF)		
	Fixed	Varying	No	Fixed	Varying
1	O		O		
2	O			O	
3	O				O
4		O	O		
5		O		O	
6		O			O

Type 1: A water ring method that starts from the fixed MB origin at the center of the image frame without the use of QF values. This method is effective to code a relatively large object staying at the center region of the image and moving little by little. It is very simple to implement and doesn't require sending any corresponding information.

Type 2: A water ring method that starts from the fixed MB origin at the center of the image frame with the use of predetermined QF values that are engaged between the encoder and decoder. Still, additional information for the QF values does not need to be sent from the encoder to decoder. However, HVS-type picture quality improvement is possible.

Type 3: A fixed-centered water ring method based on constantly varying QF values. The QF value assigned to each water ring needs to be sent frame-by-frame.

Type 4: A water ring method based on an arbitrarily-determined origin without using the QF values. The varying position of the origin needs to be sent from the encoder to decoder in a frame basis. This is effective in cases when the relatively-small fast-moving object is interesting to the viewers.

Type 5: The arbitrarily-centered water ring with the use of predetermined QF values. HVS-type quality improvement is possible; nevertheless additional information for the QF value is not required.

Type 6: The arbitrarily-centered water ring based on constantly varying QF values along designated water rings. Mechanisms for sending both varied-center positions and the assigned QF values are required. Corresponding syntax changes are required in the VOL and FGS VOP header. A sophisticated encoding process for improved visual quality can be possible because HVS-type quality improvement for the object located in an arbitrary position can be accomplished.

If the adaptive water ring scan method can be possibly adopted into the current MPEG-4 FGS video standard, two conceptual changes will be required in terms of its functional components, in comparison with the conventional raster-scan-based FGS. These required changes can be explained using the proposed decoding structure shown in Fig. 7.

i) Use of the water ring scan method rather than raster scan method: according to the order decided by the water ring scan, the VLC symbols for each bit-plane of each MB in DCT residues are formed and encoded to produce an FGS bitstream, and the bitstream produced from the water ring scan-based FGS is decoded and reconstructed based on the water ring scan method by using the reverse procedure at the decoder side.

ii) Use of the QF values for each water ring as the shift factors in the bit-plane coding: the bit-plane shift is performed using quality factors assigned to each water ring rather than using the shift factors assigned to the MBs located in the ROI.

IV. Simulation Results

In this chapter, we will show the effectiveness in the subjective coding efficiency of the adaptive water ring scan method. The proposed method is associated with the FGS coding schemes whose base-layer coding scheme is based on the MPEG-4 Part 2 ASP coding scheme. The 'Stefan' and 'Foreman' CIF-format sequences are used as test image sequences in the simulation. More test results for various image sequences can be found in our contribution documents presented in MPEG meetings [16]-[21]. As shown in Fig. 8, three regions, the full-size frame regions, the object regions within the shape masks, and the ROI regions composed of 14 MBs × 10 MBs located at the center of the CIF-format image frames, are used for comparisons of the PSNRs in luminance (Y). The coding performances of the optimized MPEG-4 Part 2 ASP single-layer codec, the current raster-scan-based FGS codec, and the water-ring-scan-based FGS codecs (Type 1, Type 4 and Type 5, as shown in Table 2) are fully compared in the computer simulation.

We will now discuss the coding results for the 'Stefan' sequence. Among the various simulation results, an example of the performance comparison is illustrated in Fig. 9. We assume that the base layer of the test sequence is encoded and successfully decoded at 128 kbps, and only a partial bitstream of the enhancement layer at 256 kbps is successfully decoded.

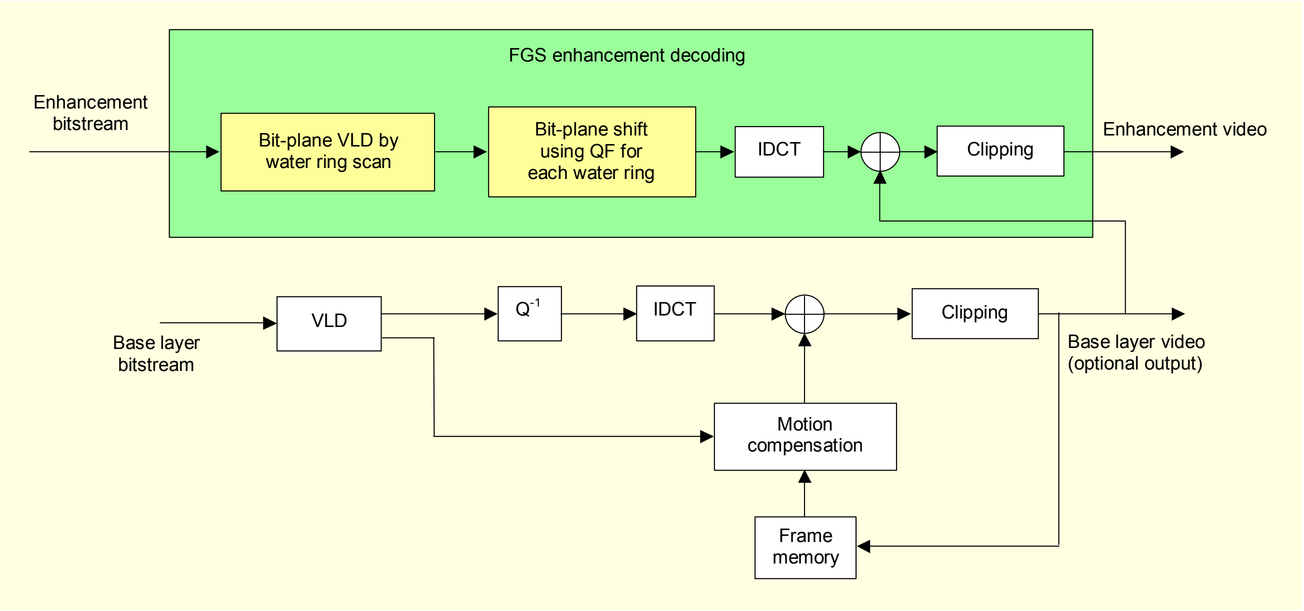
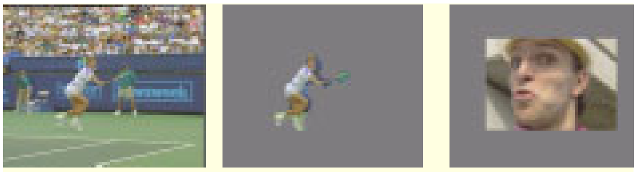


Fig. 7. Decoder structure of the MPEG-4 FGS based on the adaptive water ring scan method.



(a) Frame region (b) Object region (c) ROI region

Fig. 8. Three regions where PSNRs are obtained in the simulation.

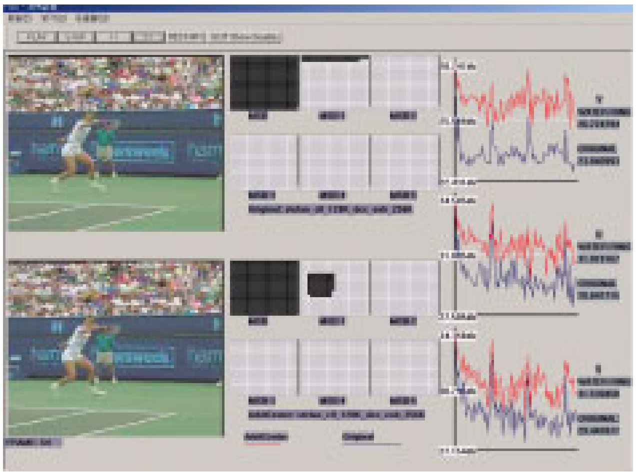


Fig. 9. An example of a performance comparison between MPEG-4 FGS and the same FGS with usage of arbitrarily-centered (Type 4) water ring scan order (Stefan, CIF, base layer: 128 kbps, enhancement layer: 256 kbps).

This figure contains two captured pictures at their 64th decoded frames, where the upper picture is a decoded image from the raster-scan-based FGS coding, and the lower picture is decoded by the FGS based on the arbitrarily-centered water ring scan method (Type 4) under the same network environment. The diagrams to the right of the pictures indicate the successfully decoded MBs (black: completely decoded, white: discarded) of both the raster-scan- and water-ring-scan-based FGS, respectively. It can be easily found that all MBs in both MSB bit-planes are completely decoded. However, for the MSB-1 bit-planes, some of the macroblocks in both cases are decoded, and the rest of them are discarded due to the network limitations. As shown in the upper diagram, it can be easily found that the raster-scan-based FGS always send the information on the MBs from the upper-left region in raster scan order and thus cannot decode the MBs corresponding to an important region such as a player area in cases where the bandwidth is limited. In comparison, the proposed water-ring-scan-based FGS coding (Type 4) preferentially encodes and decodes the macroblocks of the important region (a player area) as shown in the lower diagram. Therefore a viewer can feel more comfortable while looking at the video sequence, even if the decoder can not reconstruct the boundary part of the image frame due to the bandwidth limitation. Also, in terms of the PSNRs (luminance, chrominances) calculated on the object regions within the shape masks, the proposed water-ring-scan-based FGS produces a higher subjective quality than the raster-scan-based FGS as shown in the three graphs to the right of the diagrams, where the proposed water ring-based FGS algorithm is shown to produce about a 3-dB better image than the conventional raster-scan-based FGS. Figures 10 and 11 show the comparison results of both the object-region-based and frame-based PSNRs for the 'Stefan' sequence (CIF format, 10 Hz) with different bitrates, where the MPEG-4 single layer coding is marked as "single layer coding," the raster scan based FGS as "FGS coding", the raster-scan-based FGS with selective enhancement mode as "FGS coding SE ON," the FGS-based arbitrarily-centered water ring scan (Type 4) as "FGS water ring," and an FGS based both on the arbitrarily-centered water ring and quality factor mode (Type 5) as "FGS water ring (QF ON)." For the raster-scan-based FGS coding, the SE area is composed of 14 MBs × 10 MBs located at the center of the image frame and a 4-bit shift is used for quality enhancement. For the water-ring-scan-based FGS coding (Type 5), the QF value for each water ring is set as shown in Table 3, and those are assumed to be predetermined QF values that are engaged between the encoder and decoder. Therefore, no additional information for the QF value is required to be sent; however an HVS-type quality improvement can be possible. As mentioned earlier, the QF values are assigned as the shift factors in the coding process.

Table 3. QF values for each water ring in the simulation.

# of water ring	0	1	2	3	4	5	6
QF	3	3	2	2	1	1	1

For the FGS coding schemes, the base-layer bitstreams are transmitted at 128 kbps using a TM-5 MB-based rate control [1]. The FGS bitstreams are truncated by the FGS Server at 128 kbps, 256 kbps, 384 kbps, 512 kbps, 640 kbps, 768 kbps, and 896 kbps, and are added to the base-layer bitstream. Figure 10 shows the object-region-based PSNRs obtained from five compared coding schemes. The optimized "single layer coding" scheme achieves about a 3 to 6 dB better coding efficiency over the "FGS coding." However, the "FGS water ring" scheme (Type 4) can lead the "FGS coding" to achieve a significantly improved picture quality, about 2 to 5 dB better. The maximum gap between the "FGS water ring" scheme (Type 4) and the "single layer coding" over the various bitrates is only less than 1.5 dB and even better at 256 kbps, which means that the proposed scan method can considerably make up the limitation of the current FGS coding scheme. As expected, “FGS coding (SE ON)” has a better picture quality than the “single layer coding” because more bits are used for quality improvement in the object region. However the “FGS water ring (QF ON)” scheme (Type 5) has a much better picture quality (3 to 4 dB) than the “FGS coding (SE ON)” scheme, especially on the user-interested object region.

Figure 11 shows the overall frame-based PSNRs obtained from five compared coding schemes. As expected, the “single layer coding” scheme achieves the highest PSNRs for all bitrates. The “FGS coding” and “FGS water ring” schemes (Type 4) have very similar coding efficiencies and are about

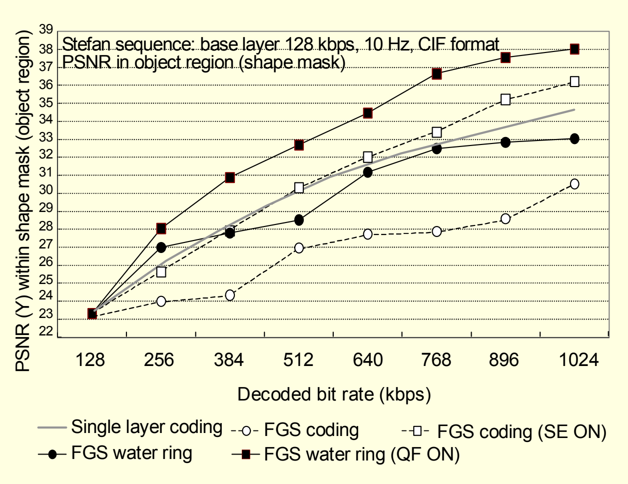


Fig. 10. Object region-based PSNR comparison results for the ‘Stefan’ sequence at 10 Hz, CIF resolution (base layer bitrates 128 kbps).

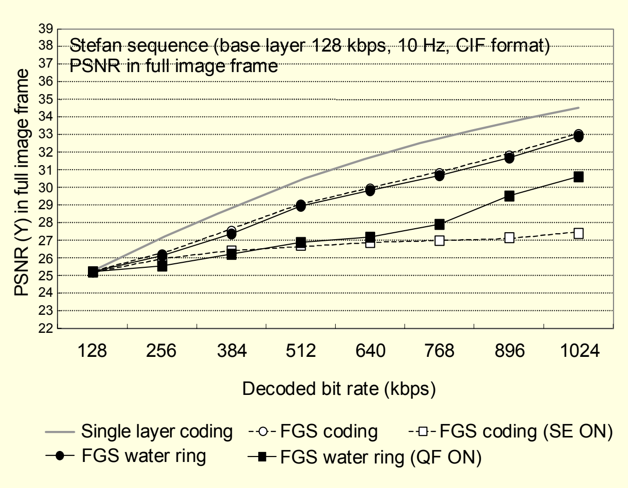


Fig. 11. Frame-based PSNR comparison results for the ‘Stefan’ sequence at 10 Hz, CIF resolution (base layer bitrates: 128 kbps).

(a) Single layer coding

(b) FGS coding

(c) FGS coding with water ring scan: Type 4

(d) FGS coding (SE ON)

(e) FGS coding with water ring scan (QF ON): Type 5

Fig. 12. Decoded image of Stefan sequence (frame rates: 10 Hz, base layer bitrates: 128 kbps, enhancement layer bitrates: 640 kbps, total bitrates: 768 kbps).

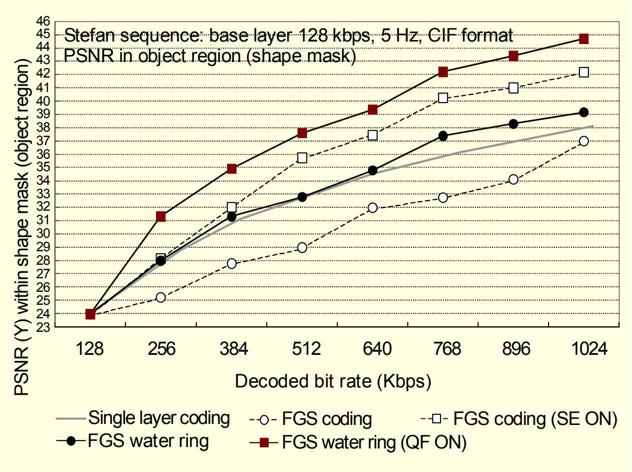


Fig. 13. Object-region-based PSNR comparison results for the ‘Stefan’ sequence at 5 Hz, CIF resolution (base layer bitrates: 128 kbps).

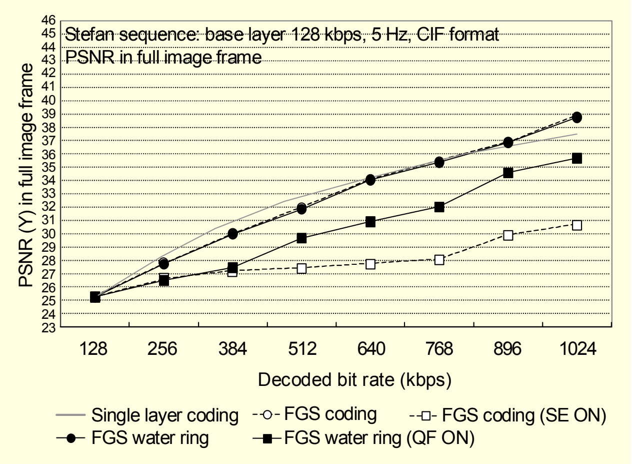


Fig. 14. Frame-based PSNR comparison results for the ‘Stefan’ sequence at 5 Hz, CIF resolution (base layer bitrates: 128 kbps).

2 to 3 dB lower than the “single layer coding” scheme. The “FGS coding (SE ON)” scheme obtained the lowest performances, about 3 to 8 dB lower than “single layer coding” and about 2 to 6 dB lower than both the “FGS coding” and “FGS water ring” schemes (Type 4). This is because most of the bits are assigned to the SE areas, therefore less bits are used for the outside of the interesting region, thus followed by a poor overall picture quality in terms of the visual aspect. The performance of the “FGS water ring (QF ON)” scheme (Type 5) obtained up to a 4-dB better overall picture quality than the “FGS coding (SE ON)” scheme. This means that the QF-based water ring scan method (Type 5) can inherently improve the viewing quality of the user-interested region much better than the FGS coding based on an MB-based SE, while maintaining a better overall picture quality.

Subjective picture quality comparisons can be also found in Fig. 12, where we can easily recognize that the water ring scan method can definitely improve the subjective picture quality of the raster-scan-based FGS coding scheme. This figure shows the decoded image snapshots of the ‘Stefan’ sequences while the base-layer bitstreams are maintained at 128 kbps and the enhancement-layer bitrates are set to 640 kbps with a 10 Hz frame rate (total bitrates are 768 kbps). As expected, “single layer coding” gives the best subjective picture quality. Many defects irritatingly appeared, especially on the object area in the “FGS coding” scheme. The “FGS water ring” scheme (Type 4) can obtain a subjectively better picture quality than the raster-scan-based FGS coding. “FGS coding (SE ON)” scheme can get a subjectively better picture quality in the SE area. However, a severe quality degradation in the outside area is greatly visible. It is like a crystal-clear window in a foggy picture. Similar effects also shown in the results of the "FGS water ring (QF ON)" coding scheme (Type 5) can be investigated. However, that effect is more subjectively natural to viewers.

Figures 13 and 14 also show the comparison results of both the object-region-based and frame-based PSNRs for the 'Stefan' sequence (CIF format, 5 Hz) with different bitrates. Five coding schemes, the "Single layer Coding" scheme, "FGS coding" scheme, "FGS coding SE ON" scheme, "FGS Water Ring" scheme (Type 4), and the "FGS water ring (QF ON)" scheme (Type 5), are also fully compared. The same phenomenon as shown in Figs. 10 through 12 was also observed.

We will now discuss the coding results for the 'Foreman' sequence (CIF format, 10 Hz), as shown in Fig. 15. The ROI PSNR is used for comparison because the object in this sequence is large and the face of the worker is the interesting region to the viewers. That face is located in the ROI region as shown in Fig. 8. The single layer coding scheme achieves about a 2- to 4-dB better coding efficiency over the raster-scan-based FGS coding scheme ("FGS coding"). However, the FGS based on a fixed-center (Type 1: "FGS water ring") produces about a 0.5- to 3-dB better PSNR than the raster-scan-based FGS, and the gaps between the "single layer" and the water-ring-based FGS (Type 1) are drastically reduced to about 0.9 to 1.9 dB.

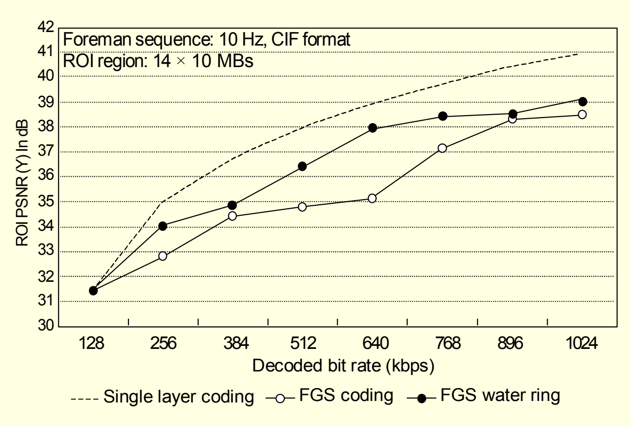


Fig. 15. ROI PSNR comparison results for the 'Foreman' sequence at 10 Hz, CIF resolution.

V. Conclusion

Under a ubiquitous environment, one source needs to be transmitted through various bandwidth conditions to different potential terminals, and thus conventional scalable video coding schemes and the MPEG-4 FGS have been designed for improving the usage efficiency of enhancement bitstreams in terms of video frames. However, those video coding schemes cannot efficiently guarantee picture quality in terms of a user's point of view due to the lack of robust bitstream delivery in the enhancement layers. This paper therefore introduces the adaptive scanning method for network-adaptive video coding methodologies, which is more suitable to a ubiquitous environment. In the proposed adaptive scanning method, the macroblocks dependent on the priority of the region in an image frame are transmitted, and thus a video sequence can be more faithfully reconstructed for a viewer. From the various simulation results, we conclude that the proposed scanning method significantly improves the subjective picture quality by about 0.5 to 6 dB over the conventional FGS coding methodologies, while maintaining the same overall frame picture quality. It also significantly reduces the visual quality gaps of the FGS in comparison with a non-scalable optimized single-layer codec, especially on the region that users are interested in. Therefore, the proposed scan method would be more suitable for video sequence transmission under fluctuated ubiquitous environments.

References

[1] M. Weiser, "The Computer for the 21st Century," *Scientific American*, vol. 256, no. 3, Sept. 1991, pp. 94-104.

[2] D. Tennenhouse, "Proactive Computing," *Communications to the ACM*, vol. 43, no. 5, 2000, pp. 43-50.

[3] W. Li, J. Ohm, M.V. Schaar, H. Jiang, and S. Li, MPEG-4 Video Verification Model Version 18.0, ISO/ IEC/ JTC1/ SC29/ WG11/ N3908, Jan. 2001.

[4] Editor's Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Geneva Modification draft 37, Joint Video Team (JVT) of ISO/ IEC MPEG and ITU-T VCEG, JVT-E146d37, Geneva, Oct. 2002.

[5] Streaming Video Profile for FGS PDAM, ISO/ IEC/ JTC1/ SC29/ WG11/ M5673, 2000.

[6] H. M. Radha, M.V. Schaar, and Y. Chen, "The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, Mar. 2001, pp. 53-68.

[7] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, Mar. 2001, pp. 301-317.

[8] H. Radha, Y. Chen, K. Parthasarathy, and R. Cohen, "Scalable Internet Video Using MPEG-4," *Signal Processing: Image Communication*, vol. 15, 1999, pp. 95-126.

[9] MPEG-4 Fine Granular Scalability Verification Model Version 4.0, ISO/ IEC/ JTC1/ SC29/ WG11/ N3317, 2000.

[10] MPEG-4 Fine Granular Scalability Verification Model Version 3.0, ISO/ IEC/ JTC1/ SC29/ WG11/ N3097, 1999.

[11] W. Tan and A. Zakhor, "Real-Time Internet Video Using Error Resilient Scalable Compression and TCP-Friendly Transport

Protocol," *IEEE Trans. Multimedia*, vol. 1, 1999, pp. 172-186.

[12] S. McCanne, M. Vetterli, and V. Jacobson, "Low-Complexity Video Coding for Receiver-Driven Layered Multicast," *IEEE J. Select. Areas Commun.*, vol. 16, 1997, pp. 983-1001.

[13] Report on MPEG-4 Visual Fine Granularity Scalability Tools Verification Test, ISO/ IEC/ JTC1/ SC29/ WG11/ N4791, Fairfax, May 2002.

[14] Report on MPEG-4 Visual Fine Granularity Scalability Tools Verification Test, ISO/ IEC/ JTC1/ SC29/ WG11/ M8331, Fairfax, May 2002.

[15] Report on MPEG-4 Visual Fine Granularity Scalability Tools Verification Test, ISO/ IEC/ JTC1/ SC29/ WG11/ M8002, Jan. 2002.

[16] G.H. Park, Y.K. Lim, Y.J. Lee, S.T. Kim, J.H. Yoon, M.H. Lee, and C. Ahn, "Water Ring Scan Order for MPEG-4 Fine Granular Scalable Coding," ISO/ IEC JTC1/ SC29/ WG11/ MPEG2000/ M6183, Beijing, July 2000.

[17] W.-S. Cheong, K. Kim, G.H. Park, Y.K. Lim, Y.J. Lee, and J. Kim, "FGS Coding Scheme with Arbitrary Water Ring Scan Order," ISO/ IEC JTC1/ SC29/ WG11/ MPEG2001/ M7442, Sydney, July 2001.

[18] G.H. Park, W.-S. Cheong, K. Kim, Y.J. Lee, Y.K. Lim, and J. Kim, "Performance of the Water Ring Scan Method in MPEG-4 FGS Methodology," ISO/ IEC JTC1/ SC29/ WG11/ MPEG2001/ M7715, Pattaya, Dec. 2001.

[19] G.H. Park, Y.J. Lee, W.-S. Cheong, K. Kim, J. Kim, and Y.K. Lim, "Water Ring Scan Method for H.26L Based FGS," ITU-T SG16 Q.6 JVT-B094, Geneva, Jan. 2002.

[20] G.H. Park, W.S. Cheong, K. Kim, Y.J. Lee, Y.K. Lim, and J. Kim, "Water Ring Scan Method for MPEG-4 and H.26L Based FGS Methodologies," ISO/ IEC/ JTC1/ SC29 WG11/ MPEG/ M 8023, Jeju, Korea, Mar. 2002.

[21] G.H. Park, W.S. Cheong, K. Kim, Y.J. Lee, W.H. Yoo, S.H. Son, Y.K. Lim, and J. Kim, "Water Ring Scan Method for Advanced Fine Granular Scalability," ISO/IEC/ JTC1/ SC29/ WG11/ MPEG/ M8593, Klagenfurt, July 2002.

[22] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 Rate Control for Multiple Video Objects," *IEEE Trans. Circuits Syst. Video Techonol.*, vol. 9, no. 1, 1999, pp. 44-58.

[23] D. Chai, K.N. Ngan, and A. Bouzerdoum, "Foreground/ Background Bit Allocation for Region-of-Interest Coding," *Proc. IEEE Int'l Conf. Image Processing (ICIP 2000)*, vol. 2, 2000, pp. 923-926.

Gwang Hoon Park received the BS and MS degrees in the Department of Electronic Engineering from Yonsei University, Seoul, Korea in 1985 and 1987, and received the MS and PhD degrees in the Department of Electrical Engineering and Applied Physics from Case Western Reserve University, Cleveland, OH, USA in 1991 and 1995. He was a Principal Researcher at the Information and Telecommunication R&D Center, Hyundai Electronics Industries Co., Ltd., Korea (1995 to 1997). He was an Associate Professor of the Department of Computer Science, Yonsei University, Wonju, Korea (1997 to 2001). Since 2001, He has been an Associate Professor of the Department of Computer Engineering, Kyunghee University, Korea (2001 to present). He is the co-inventor of 15 international registered patents and 37 domestic registered patents. His current research interests are moving picture coding, video signal processing, multimedia systems, pattern recognition, and computational intelligence.

Kyuheon Kim received his BSc in electronics engineering in 1989 from Hanyang University, Seoul, Korea, and PhD in electrical & electronic engineering from University of Newcastle upon Tyne, UK, in 1996. From 1996 to 1997, he worked as a research fellow for University of Sheffield, UK. Since 1997, he has been a Member of Research Staff in Electronics and Telecommunications Research Institute, Korea, and is currently a Head of Interactive Media Processing Research Team, Broadcasting Media Technology Department. Also, he has been a Head of Korean Delegates to MPEG international standard organization. His interests are on digital signal and video processing and interactive multimedia broadcasting systems.