A Low-Power Surveillance Video Coding System with Early Background Subtraction and Adaptive Frame Memory Compression

Hyun Kim and Hyuk-Jae Lee, *Member, IEEE*

***Abstract*—Various commercial video surveillance systems have become popular in recent years. The increasing demand for video surveillance systems which support high resolutions and operate with battery power makes it important to reduce the power consumption and to prolong battery lifetimes. H.264/AVC video coding standard and HEVC standard are widely used in many video surveillance systems due to their high compression efficiency. However, the high computation complexity of video coding standards requires a considerable amount of power consumption. This paper proposes a novel low-power surveillance video coding system that reduces both the internal power consumption inside the video encoder chip and the external power consumption by the external memory outside the chip. The proposed algorithm attempts to segment the input video into the foreground and background (BG) according to integer motion estimation (IME) information and adaptively controls the complexity of video compression modules according to the segmentation results. In order to reduce the external power consumption further, adaptive frame memory compression (FMC) using the segmentation results from the BG subtraction is proposed. The adaptive selection of the compression ratios and processing units for the FMC significantly improves the compression efficiency of the FMC. The proposed system reduces power consumption by up to 61.6% compared to a conventional video surveillance system.**

***Index Terms*— Adaptive Frame memory compression, Early background subtraction, Low power design, Video compression, Video surveillance system**

I. Introduction

Due to the increasing level of demand for security enhancements, the number of video surveillance systems has increased rapidly. Video surveillance systems have long been deployed in places where security is needed, such as state institutions, airports, subway stations, and highways. Recently, they have also found use in houses and quotidian places such as parking areas, markets, and banks in order to ensure safety. This trend is driven by the recent development of technology such as smart phones, car dashboard cameras, and other mobile devices as well as analog closed-circuit television systems which are also utilized for video surveillance systems. As a result, numerous commercial video surveillance systems are widely used in many places. Several researchers have studied effective security management [1]-[3] with video surveillance systems. As the resolution of video frames increases, the size of raw video data increases as well. Consequently, video compression techniques should be integrated into a video surveillance system to reduce the memory space for video storage. In general, state-of-the-art compression techniques, i.e., the H.264/AVC video coding standard [4] and the High Efficiency Video Coding (HEVC) standard [5], are widely used in many video surveillance systems recently due to their high compression efficiency [6]-[9]. H.264/AVC and HEVC offer remarkable compression efficiency but also require high computational complexity that results in significant power consumption. Furthermore, as battery-powered surveillance systems without power lines have come into wider use recently [10], it is becoming increasingly important to reduce power the consumption during video compression in video surveillance systems in order to prolong battery lifetimes [11].

Video surveillance systems can be classified into two categories: moving devices and stationary devices. A stationary device, the target of this paper, uses a static camera and therefore has a background (BG) region which is always fixed. As a result, only moving foreground (FG) objects are changed in the images captured by this device, whereas the regions in static BG do not contain any objects of interest [12]. Extensive research has been undertaken to enhance the performance and to reduce the power requirements of a stationary system. In two studies [13] and [14], high-quality long-term reference frames are used to find BG regions; consequently, the rate-distortion (R-D) performance of the stationary systems is enhanced. In [15], the complexity reduction method for uncovered BG regions which utilize a static BG model has been proposed. However, these schemes require a complete change of the system architecture for an additional transmission of the reference BG model. In [16], the application of computer vision to surveillance systems was studied, but the target is limited to software level; thus, hardware logic analysis is not included.

Manuscript received September 28, 2017; accepted November 15, 2017. Date of publication December 19, 2017. This work was supported by the Brain Korea 21 Plus Project in 2017. *(Corresponding author: H.-J. Lee.)*

H. Kim is with the Inter-university Semiconductor Research Center, Department of Electrical and Computer Engineering, Seoul National University, Seoul, 08826, Korea. (e-mail: snusbkh0@capp.snu.ac.kr).

H.-J. Lee is with the Inter-university Semiconductor Research Center, Department of Electrical and Computer Engineering, Seoul National University, Seoul, 08826, Korea. (e-mail: hyuk_jae_lee@capp.snu.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCE.2017.015073

Several other researchers have investigated BG subtraction in order to concentrate on video compression for FG objects [9], [12], [17]-[20]. Previous works attempted to segment out FG objects from a BG by utilizing various components. In [9], content differences in consecutive input frames are detected by analyzing the color and moving correlations. The scheme in [12] proposes segmentation of the sampled Y pixels as either FG regions or BG regions by using a speeded up Gaussian mixture model algorithm. Razif et al. [17] attempts the block segmentation by comparing the sum of absolute difference (SAD) to a pre-defined threshold. Qin et al. [18] presents the segmentation of BG regions by the complex basis matrix construction and basis matrix update processes. In two studies [19] and [20], they segment out BG regions by finding the difference of pixel luminance values between the current and previous frames and comparing the current frame to reconstructed G-picture which is generated as the long-term reference frame, respectively. These methods exploit suitable encoding conditions according to different object types. However, there are three main problems in these works related to BG subtraction. First, additional modules for BG subtraction are added to video surveillance systems, which increase the gate counts and computational requirements. Second, the research range is limited to the internal computational complexity. Although power consumption can be greatly reduced by applying optimized operation conditions [21], the amount of the power consumed by data transfers to external memory is significant [22]. Third, the criteria which determine the division of BG and FG regions are too simple because they only exploit one or few factors for BG subtraction. As a result, the accuracy of BG subtraction is not satisfactory.

In order to address these problems, this paper focuses on video compression in the video surveillance system and proposes a low-power video surveillance system with a pipelined IME-based BG subtraction (PIBS) and an adaptive frame memory compression (AFMC), which reduces both internal and external power consumptions. The proposed system performs BG subtraction without an additional module by utilizing the information from integer motion estimation (IME) of a pipelined architecture [23]. Furthermore, the PIBS improves the accuracy of BG subtraction by simultaneously considering various computation results from IME. Based on the segmentation results from the accurate PIBS, a macroblock (MB) which belongs to BG is determined as the skip mode provided by H.264/AVC and HEVC standards and a co-located MB in the consecutive frame is encoded with a relatively small search range (SR) and a simple block mode (BM) because BG regions are nearly unchanged from the previous frame. This significantly reduces the internal power consumption. The proposed PIBS scheme can classify complex regions such as an uncovered BG and an object boundary with reference to the segmentation results of both previous and current frames. In addition, the segmentation results from the PIBS can help to determine appropriate operation constraints for the frame memory compression (FMC), which contributes to external power savings. Adaptive selection of the compression ratios and processing units for the FMC significantly improves the compression efficiency. In order to support the efficient memory address table for various compression ratios, this paper proposes novel address calculation schemes which facilitate real-time application of the address calculation with a small increase in internal memory usage. Experimental results show that the proposed scheme reduces the internal power consumption by up to 51.2% with only a minor degree of quality degradation compared to a conventional system that includes only video coding standard without applying any schemes. Furthermore, additional utilization of the AFMC improves power savings by up to 61.6%. As the power consumption caused by the video compression module contributes around a half of the total power in a video surveillance system [9], it is possible to prolong the device usage time up to 1.45 times by applying the proposed scheme to battery-powered video surveillance systems.

The remainder of this paper is organized as follows. Section II explains the hardware platform for the proposed surveillance system and the adaptive encoding configurations according to the results of the PIBS. Section III proposes the AFMC scheme and the memory address generation scheme to support various compression ratios in the AFMC. In Section IV, experimental results are presented, and the study is summarized in Section V.

II. Early Background Decision and Adaption for Low-Power Video Surveillance System

A. Hardware Platform for the Proposed System

Fig. 1 shows the block diagram of the hardware platform for a low-power video surveillance system. A video surveillance system receives input video from a camera, compresses it using the video encoder, and stores the compressed bitstream in long-term memory for video indexing or searching. OpenRISC undertakes system configuration. In this paper, a hardware based H.264/AVC encoder with a widely used three-stage pipelined architecture is utilized for video compression standard [24]. Fig. 2 presents a three-stage pipelining architecture of the H.264/AVC encoder for the low-power surveillance system. In this architecture, IME is executed in the first stage. An important difference with regard to the original pipelining architecture in [23] is the insertion of BG subtraction into the first stage after the IME operation, which is possible because BG subtraction is performed by only using the computation results from IME. A detailed description of the PIBS technique is explained in Section II-B. Then, fractional motion estimation (FME), intra prediction (IP), and discrete cosine transform and quantization (DCTQ) are executed in the second stage. In the third stage, context-based adaptive variable length coding (CAVLC) and adaptive deblocking filter (ADF) are executed. It should be noted that the PIBS technique can utilize the existing pipeline structure of video compression standard by inserting only a few additional operations since all of these processes in Fig. 2 can be designed with a single pass based on video coding standard. In order to store the FG/BG marker which indicates whether each MB belongs to FG or BG, an additional SRAM module, denoted as “Marker SRAM” in Fig. 1, is embedded in the hardware platform. It should be noted

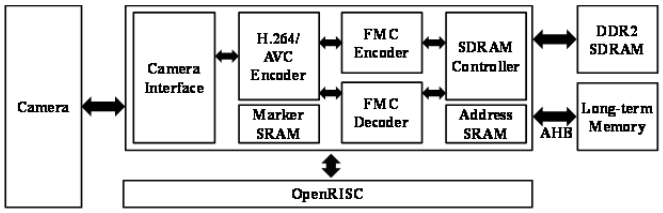


Fig. 1. The proposed hardware platform.

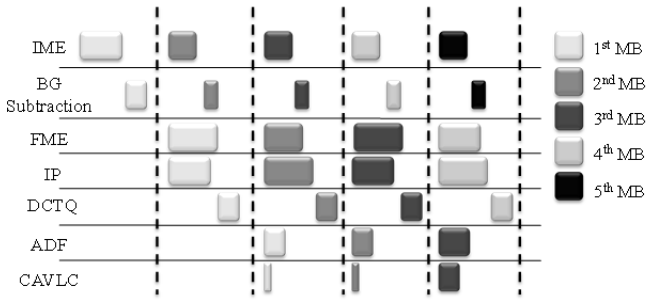


Fig. 2. Pipelined operation for the low-power surveillance system with the H.264/AVC encoder.

that it is possible to apply the proposed design to the HEVC standard because the pipelined architecture in the HEVC standard is very similar to that in the H.264/AVC encoder [25].

For the FMC encoder and decoder, one-dimension (1-D) three-level synthesis lifting discrete wavelet transform (DWT) coefficients that achieve high compression efficiency with low computational complexity are compressed by a set partitioning in hierarchical trees (SPIHT) coding [26], [27]. There are three main reasons why the combination of the 1-D DWT and SPIHT is selected for the FMC. First, it generates a bitstream that precisely meets the target compression ratio. This property enables a simple address calculation to access the reference frame data in the external memory. Second, the 1-D structure requires fewer hardware resources and thus less power. Therefore, it can minimize the power consumption with additional FMC modules. Third, the combination of the 1-D DWT and SPIHT can support various coding granularities and compression ratios. This property facilitates the adaptive application of the appropriate coding granularity and compression ratio to each MB according to the FG/BG marker.

The SDRAM controller that operates with the H.264/AVC encoder [24] is revised to support the PIBS and AFMC. Specifically, for the AFMC scheme, an additional SRAM denoted as "Address SRAM" in Fig. 1 is embedded in the hardware platform. With the help of "Address SRAM," the SDRAM controller can calculate the address of the reference frame data which are compressed with various compression ratios in the AFMC.

B. Background Subtraction from Integer Motion Estimation

To avoid an increase of the computational complexity for BG subtraction, the proposed PIBS scheme only utilizes the computational results from the IME operation. These results, such as the motion vector (MV), the predicted MV (PMV), the encoded BM, the SAD cost, and the neighbor information are naturally generated from the encoding process; thus, it is unnecessary to perform any supplementary computation for BG subtraction. As a result, the hardware architecture for the proposed PIBS is readily feasible on a conventional system because it does not incur an additional hardware cost except for a small number of arithmetic logic units.

The proposed PIBS classifies input frames into two categories according to the amount of ambient light because the noise in input frames is greatly affected by the camera exposure time determined by the amount of light. The classification according to the amount of ambient light achieves the video quality improvement and facilitates accurate BG subtraction even when ambient lighting conditions change rapidly. This classification is performed by ambient light sensors or by user-implemented modules. This study uses the user-defined module that compares the average pixel luminance values of the frame with pre-defined threshold which is determined by 80 through experiments. However, in commercial devices, it is possible to determine bright and dark conditions by using the ambient light sensors which are widely used in many mobile devices including smartphones [28]. Fig. 3 shows a flowchart of the proposed PIBS scheme for dividing FG objects and BG regions. The PIBS scheme operates at the MB level. In case of the bright condition, the noise is relatively low; as a result, relatively accurate motion information can be obtained. Therefore, various types of motion information, such as the MV and PMV, become the main determining factors in the segmentation process, whereas the BM and SAD costs serve to improve the segmentation accuracy. There are two conditions yielding BG candidates. First, if the MVs for both the horizontal and vertical directions (i.e., MV_X and MV_Y) are zero and the best SAD cost for the 16×16 BM in the IME operation (i.e., BC_{IME}) is under a certain threshold (i.e., $TH_{16\times16}$), this MB is assumed to be a BG candidate because zero MVs are assumed to mean a stationary BG. BC_{IME} is used to classify an object boundary region which has a relatively high SAD cost and to filter out MBs whose MVs are incorrectly determined as to be zero. $TH_{16\times16}$ is not decided as a fixed value, but is set to a variable value based on the average SAD cost of the 16×16 BM for FG MBs in the previous frame because the SAD cost depends on the encoding options such as the QP value and the characteristic of the input video. It should be noted that the difference between the average SAD cost in FG regions and that in BG regions is very large due to the MB complexity. Second, if both the MV_X and MV_Y have a value between -1 and 1, the PMVs for both the horizontal and vertical directions (i.e., PMV_X and PMV_Y) are zero, and BC_{IME} is smaller than $TH_{16\times16}$, this MB is assumed to be a BG candidate. On the other hand, in case of the dark condition, the noise is very high, which results in inaccurate motion information. Therefore, the SAD cost with the co-located MB in the previous frame (i.e., $SAD_{CO\text{-}LO}$) becomes the main factor for BG segmentation. There are two conditions yielding BG candidates. First, if the $SAD_{CO\text{-}LO}$ is smaller than $TH_{16\times16}$, the current MB is assumed to be a BG candidate because a small $SAD_{CO\text{-}LO}$ means that there are no changes between consecutive frames. Second, if all instances of MV_X, MV_Y, PMV_X, and PMV_Y are zero and the BC_{IME} is smaller than $TH_{16\times16}$, this MB is assumed to be a BG candidate. The MBs which have both zero MVs and zero PMVs are highly likely to be established as a BG MB.

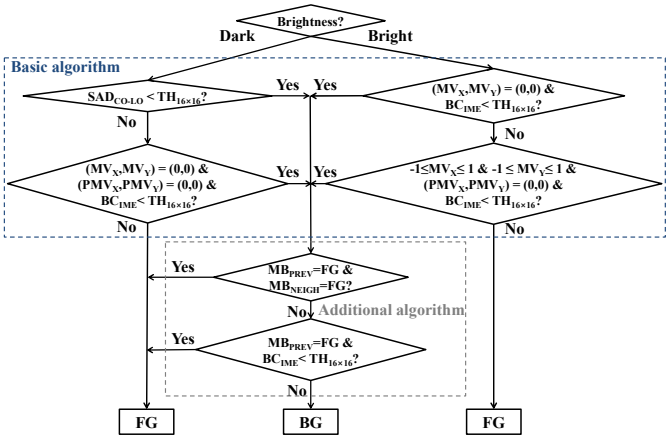
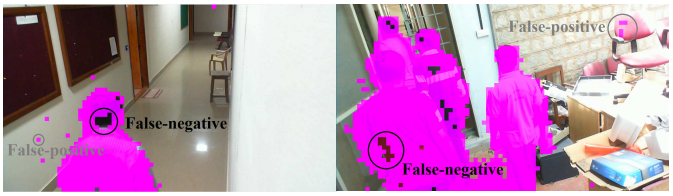


Fig. 3. The flow chart of the PIBS scheme.

However, BG candidates are not sufficient to be determined as a BG MB because "false-negative" errors, common errors during BG segmentation as shown in Figs. 4 (a) and (b) with the black circles, can occur when moving FG objects stop for a while or the regions inside slowly moving FG objects have very similar pixel values with the co-located MB in the previous frame. These errors would result in serious problems because important data can be lost in the objects of interest. In order to address this possibility, two additional operations are performed. First, in order to detect the middle regions of FG objects, the FG/BG markers of the co-located MB in the previous frame (i.e., MB_{PREV}) and four neighboring MBs in the current frame (i.e., MB_{NEIGH} for the left, upper, upper-left, and upper-right MBs) are utilized. If all of them are determined as FG, the current MB is assumed to be FG MB. Second, to detect stationary FG objects, the MB_{PREV} and the BC_{IME} are utilized. If the MB_{PREV} is determined to be FG and the BC_{IME} is smaller than $TH_{16\times16}$, this is assumed to indicate stationary FG objects and the current MB is assumed to be FG MB. There are other common errors in BG segmentation, with one such example being "false-positive" errors, as shown in Figs. 4 (a) and (b) with the gray circles. Especially, these errors mainly occur when BG regions or ambient lighting conditions change. However, in contrast to serious false-negative errors, these are not crucial because they cause false inclusions of the BG regions into FG regions. Accordingly, the computation and corresponding power consumption levels slightly increase without loss of important data in the objects of interest. Furthermore, owing to the precise PIBS scheme, the proportion of such errors is not critical. As a result, regardless of the amount of light, the PIBS significantly enhances the accuracy



(a) (b)

Fig. 4. False-negative and false-positive errors in the background subtraction: (a) "Access Door" sequence. (b) "Entrance" sequence.

of BG subtraction by simultaneously considering various computation results from IME.

C. Various Encoding Options according to FG/BG Markers

In this paper, the segmentation results from the PIBS are stored in the embedded memory for the FG/BG marker and are utilized for the low-power encoding operation. Table I shows the encoding options for the H.264/AVC encoder according to the FG/BG markers. For optimized encoding process for the video surveillance system, the segmentation results of the previous frame (MK_P in the second column) as well as those of the current frame (MK_C in the first column) are used to determine the encoding option. Utilizing the segmentation results of both the current and previous frames facilitates classification of complex regions, such as an uncovered BG and an object boundary. Such markers for both the current and previous frames require additional internal memory; however, the additional memory size for an HD (1280×720) frame is only 900 (2×1280×720/(16×16)/8) bytes because each MB requires only two bits for representing the FG/BG status. There are four cases depending on the FG/BG markers and the classification results are denoted in the third column. The fourth and fifth columns represent the encoding options for the current MB (MB_{CURR}) and the IME options (SR and BM) for the co-located MB in the next frame (MB_{NCO-LO}), respectively.

TABLE I
H.264/AVC ENCODING OPTION ACCORDING TO A FG/BG MARKER

(MK_C, MK_P)	Classification	Coding option for MB_{CURR}	IME option for MB_{NCO-LO}
(FG, FG)	Strong FG (M3)	Regular (FME & IP)	Regular SR & Regular BM
(FG,BG)	Object Boundary (M2)	Only IP	Regular SR & Regular BM
(BG,FG)	Uncovered BG (M1)	Only IP	Regular SR & Regular BM
(BG,BG)	Strong BG (M0)	SKIP mode (Skip FME & IP)	Small SR & Simple BM

First, the MB whose markers for both the current and previous frames are FG certainly belongs to FG regions (denoted by M3). In an M3 region, the video quality is crucial. Therefore, the MB_{CURR} is encoded with regular operation options without low-power schemes. Furthermore, the M3 region is more likely to be a FG region in the next frame as well. Hence, regular SR and BM for IME are processed for the MB_{NCO_LO}. Second, the MB for which the marker of the current frame has a FG value but that of the previous frame has a BG value (denoted by M2) may represent the neighborhood of an object boundary. In an M2 region, it is difficult to find similar blocks in the previous frame with the current MB; accordingly, the best coding mode for the MB_{CURR} is likely to be determined as an intra mode and thus, only the IP operation is executed by omitting the FME operation. Third, the MB that the current marker denotes as BG but the previous marker denotes as FG (denoted by M1) may be an uncovered BG region. In an M1 region, it is also difficult to find similar blocks in the previous frame because the blocks in the uncovered BG region appear to be new; thus, the best coding mode for the MB_{CURR} is also likely to be determined as an intra mode. Therefore, the FME

operation is skipped and only the IP operation is processed to reduce unnecessary computation. Lastly, the MB whose markers for both the current and previous frames are BG (denoted by M0) certainly belongs to a BG region. Because the current MB and the co-located MB in the previous frame are nearly identical and small residual errors in BG regions have little impact on the video quality, both FME and IP are skipped and the MB_{CURR} is determined as the skip mode. It should be noted that even though the FME operation is not performed in BG regions, video quality degradation is not large unlike FG regions because most FME results in BG regions indicate inter-pixels rather than sub-pixels and consequently, the effect of the FME operation is minimal. Furthermore, an M0 region is more likely to be a BG region in the next frame; thus, the $MB_{NCO\text{-}LO}$ is encoded with a very small SR and very simple BM for further reducing the computation complexity and the external memory bandwidth (BW). From some statistical experiments for analyzing the distribution of the MVs, the range of the SR for IME is limited from -1 to 1 (i.e., -1, 0, 1) in both the horizontal and vertical directions since most MVs in BG regions are determined within these ranges. As a result, the computation complexity and the external BW for the IME operation in the proposed scheme is reduced by 98.48% compared to those in a 32×32 full search and by 59.38% compared to those in one earlier approach [29] which efficiently reduces the SR of IME. In the case of the BM for IME, only 16×16 BM is processed, as it is mostly selected for the best mode in unchanged MBs and is used for a criterion of the skip mode decision in H.264 encoding [4]. The computation complexity for the IME operation in the proposed scheme is reduced by 88.6% compared to that in H.264 compression standard [30]. The proposed operation options with the PIBS achieve remarkable power savings and the magnitude of the power savings increases as the proportion of the M3 region decreases. The power savings by the PIBS become even larger when the HEVC encoder is utilized for the main compression scheme instead of the H.264/AVC encoder because the HEVC encoder performs much more complicated computation for high compression efficiency.

III. Adaptive Frame Memory Compression

External memory accesses account for a considerable portion of the total power consumption. The FMC can achieve a further reduction in the power requirement by decreasing the number of external memory accesses. However, this causes significant video quality degradation. In order to avoid this problem, this section presents an AFMC that reduces the external power consumption in common with the previous non-adaptive FMC [31] while retaining the video quality for the objects of interest at a reasonable sacrifice of the quality in the BG region.

A. Various Compression Options for Frame Memory Compression According to FG/BG Markers

Table II shows a PSNR comparison of the combination of the 1-D DWT and SPIHT according to various compression ratios and processing units. For these experiments, Table III summarizes the conditions used with eighteen HD video sequences (listed in the top row of Table III) [12]. Each number in the brackets represents the number of frames in each sequence and the test sequences that belong to the dark condition are marked in bold font. The encoding configurations of the H.264/AVC encoder are presented in the bottom row of Table III. In Table II, the PSNRs of processing units are nearly proportional to the compression ratios, whereas those of identical compression ratios are mostly proportional to the processing units. In other words, the PSNRs are large when the compression ratio is high and the processing unit is large. Considering their influences, the operation constraints of the FMC are selected according to the results of the PIBS.

TABLE II
AVERAGE PSNR VALUES OF 1-D DWT AND SPIHT ACCORDING TO THE COMPRESSION RATIOS AND PROCESSING UNITS

PSNR (dB)		CR 1/4	CR 3/8	CR 1/2	CR 3/4
PU	1×16	34.96	41.31	47.47	50.63
	1×32	36.22	43.22	49.48	52.46
	1×64	37.24	44.82	51.07	53.84

Table IV shows the operation options of the AFMC according to the information from the FG/BG markers. Each column represents various compression ratios according to whether the current MB is FG or BG. For FG MBs, it is important to maintain high video quality. Thus, a relatively low FMC compression ratio, 3/4, is used. In contrast, BG MBs do not contain the objects of interest, and it is more important to reduce the external BW. Therefore, a relatively high FMC compression ratio, 3/8, is selected. The selection of the compression ratios is based on experiments. Among various combinations of compression ratios, 3/4 and 3/8 FMC compression ratios for FG and BG regions, respectively, achieve the best performance in terms of the trade-off between the power saving and the video quality. An adaptively selected compression ratio for the FMC significantly reduces the external power consumption in a BG region, whereas it retains the video quality for the objects of interest in a FG region. On the other hand, each row represents various processing units according to whether consecutive MBs have identical values or not. If four consecutive MBs have identical values regardless of their FG/BG statuses, the four consecutive MBs are combined and compressed by 1×64 to improve the video quality. Even if the four MBs are not consecutively connected, it can be processed by 1×32 to realize an improvement in the video

TABLE III
EXPERIMENTAL CONDITIONS

Test Sequences	Access Door (250) Entrance (250) Overtaking Car (200)	Back Yard1 (200) Evening Fade (200) Parking Lot (500)	Back Yard2 (250) **Light Switch** (200) **Road** (200)	**Bund Nightscape** (200) **Low Light** (200) Slow Motion (200)	**Cafeteria** (200) No Activity1 (200) Sun Variation (100)	Car (200) No Activity2 (200) Walkway (200)
H.264/AVC Configuration	Baseline Profile		QP : 20,24,28,32,36		Intra Period : 60	

quality if two consecutive MBs have identical values. Otherwise, each MB is compressed by 1x16. The minimum processing unit of the FMC is 1×16 when consecutive MBs have different FG/BG values because the processing unit of the H.264/AVC encoder is 16×16. A larger processing unit such as 1×64 achieves better video quality, but it requires additional internal memory to combine the four consecutive 16×16 reconstructed MBs. The additional internal memory size is 1,536 (16×16×12×4/8) bytes, but it is an acceptable overhead considering the enhanced video quality. Especially, the use of a large processing unit mitigates video quality degradation in a BG region due to a low compression ratio because BG regions are more likely to be consecutive. It should be noted that the proposed AFMC does not require any additional operations for adaptive selection of compression ratios and processing units by utilizing the information from the FG/BG markers and significantly improves the video quality by considering both the compression ratios and processing units simultaneously.

TABLE IV
VARIOUS FRAME MEMORY COMPRESSION OPTIONS DEPENDING ON THE COMPRESSION RATIOS AND COMPRESSION UNITS

(CR, PU)		CR	
		FG	BG
PU	Four consecutive MBs	(3/4, 1×64)	(3/8, 1×64)
	Two consecutive MBs	(3/4, 1×32)	(3/8, 1×32)
	Otherwise	(3/4, 1×16)	(3/8, 1×16)

B. Memory Address Generation for Various Compression Ratios in the Frame Memory Compression

The AFMC significantly reduces the power consumption for external memory accesses while minimizing video quality degradation of the objects of interest. However, the compressed data size of each 16x16 reconstructed MB differs according to the compression ratio used. Thus, in contrast to a fixed compression ratio, an address table for random accessibility is required where the memory address for the corresponding MB is stored. However, a straightforward address table is too large to be stored in the internal memory. In order to store the memory address efficiently in the internal memory, this paper utilizes the FG/BG markers which contain the compression ratio of each MB. A key concept in this study is that the memory address for the current MB can be calculated by utilizing the number of previous FG and BG MBs. In other words, the memory address of a particular MB is calculated by adding the allocated memory size of the previous MBs from the starting address of the current frame. However, it takes a significant amount of time to calculate the address to access the latter MBs in the frame. To speed up the address calculation for arbitrary MBs, two auxiliary methods are devised. In Fig. 5, the memory address for the leftmost MB in each MB line is stored in the internal memory. Thus, the address calculation starts not from the address of the first MB in the frame but from the address of the leftmost MB in each line. However, this approach still requires a large amount of calculation because the maximum number of addition operations is proportional to the width of videos. In order to reduce the calculation time further, pre-calculated multiplication results are stored in another table utilizing the regularity of the allocated bit size. For HD sequences, one MB line consists of 80 MBs and thus, up to 79 additions need to be pre-calculated. Since the bit sizes for each FG MB and BG MB then become 288 (=256×1.5×(3/4)) bytes and 144 (=256×1.5×(3/8)) bytes, respectively, the corresponding multiplication results are stored in the pre-calculated multiplication table. As a result, it is sufficient to read two values from the pre-calculated multiplication table according to the number of previous FG and BG MBs and then to add them together.

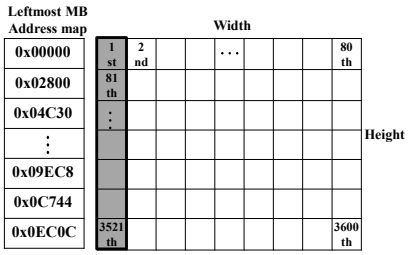


Fig. 5. The leftmost MB address map for an HD frame.

The two proposed methods facilitate real-time address calculation for the current MB in a single cycle. However, they require the additional internal memory of which the size for the HD frame is 294 bytes, a summation of 112.5 (=20×720/16/8) bytes for the leftmost MB address table and 181.5 bytes for the pre-calculated multiplication table. Even considering the FG/BG marker size of 450 bytes for the current frame, the proposed address calculation scheme reduces the embedded memory size remarkably by 91.73% compared to the straightforward table [22]. In addition, compared to the previous address generation scheme for the lossless FMC in [32], 63.71% of the embedded memory size is remarkably reduced by fixing the compression ratios for FG and BG regions and optimizing the embedded memory size.

IV. EXPERIMENTAL RESULTS

In this section, the power consumption and R-D performance of the proposed video surveillance system are estimated and the effectiveness of the proposed system is evaluated.

A. Power Estimation for the Proposed System

In this subsection, in order to estimate the power consumption of the proposed hardware platform shown in Fig. 1, the power consumption of the system is estimated with appropriate design/simulation tools using the data from a precise hardware model. The power consumption amounts for the long-term storage memory and the camera are not considered in this paper because these levels are nearly identical in every surveillance system. For estimating the internal logic circuits, the amounts of power consumed for the H.264/AVC encoder, the FMC encoder, the FMC decoder, and the additional SRAMs are obtained by a post-layout simulation using a netlist synthesized by the Synopsys Design Compiler with the 0.13 μm process technology. In order to estimate the power consumption for external memory accesses, the number of memory accesses and the necessary memory size are simulated by a register transistor logic (RTL) simulation environment. From the RTL simulation results, the power

consumption by LPDDR2 SDRAM (2Gb, ×32width, burst 8) is derived using a power calculator [33].

In Table V, the amounts of power consumed for the conventional system (denoted as "Conventional"), the proposed system with the PIBS (denoted as "PIBS"), the proposed system with the PIBS and AFMC (denoted as "PIBS+AFMC"), and the system with the PIBS and the previous FMC in [31] for a comparison with "PIBS+AFMC" (denoted by "PIBS+FMCCOMP") are estimated with the experimental conditions in Table III and the results are summarized. In the second column, the amount of power consumptions in the internal logic circuits (PC_{INT}) is presented. When compared to "Conventional", both "PIBS" and "PIBS and AFMC" reduce the PC_{INT} by 83.87mW and 73.34mW, respectively, because the computational complexity levels in those systems are significantly reduced by the PIBS. However, "PIBS+AFMC" consumes slightly more power than "PIBS" due to additional modules for the AFMC. The PC_{INT} of "PIBS+FMCCOMP" is somewhat smaller than "PIBS+AFMC" because "PIBS+FMCCOMP" does not use additional SRAMs for the proposed AFMC and memory address scheme. In the third column, the amount of power consumptions for external memory accesses (PC_{EXT}) is presented. When compared to the "Conventional", both "PIBS" and "PIBS+AFMC" reduce the PC_{EXT} by 74.53mW and 117.11mW, respectively. "PIBS" achieves external power savings by reducing the external memory BW for reference frames as the SR of IME is reduced by the PIBS. On the other hand, "PIBS+AFMC" achieves greater power savings than "PIBS" because the external data for reference frames which are already reduced by the PIBS are additionally compressed by the AFMC in "PIBS+AFMC." The PC_{EXT} of "PIBS+FMCCOMP" is nearly identical to that of "PIBS+AFMC." The PS_{TOT}, the power savings of the proposed systems compared to the conventional system in the fifth column, shows that both "PIBS" and "PIBS+AFMC" reduce the PC_{TOT} in the fourth column by 51.2% and 61.6%, respectively. The PC_{TOT} of "PIBS+FMCCOMP" is nearly identical to that of "PIBS+AFMC."

TABLE V
POWER CONSUMPTION OF VARIOUS VIDEO SURVEILLANCE SYSTEMS

Video Surveillance System	PC_{INT} (mW)	PC_{EXT} (mW)	PC_{TOT} (mW)	PS_{TOT} (%)
Conventional	133.54	175.77	309.31	-
PIBS	49.69	101.24	150.93	51.2
PIBS + AFMC	60.2	58.66	118.86	61.6
PIBS + FMCCOMP	58.72	58.69	117.41	62

B. Performance Estimation for the Proposed System

In this subsection, in order to demonstrate the video quality of each system for various output bitrates, Figs. 6(a) and (b) present the R-D curves for entire frames and the objects of interest in FG regions, respectively. The horizontal and vertical axes represent the bitrates and PSNRs, respectively. The results are tested with the eighteen video sequences in Table III and the average R-D curves are presented. In Fig. 6(a), the difference between the dark gray curve labeled "PIBS" and the light gray curve labeled "Conventional" is very small regardless of the bitrate. The maximum difference in the PSNR between them is approximately 0.2dB at 2,500kbps, a negligible difference. Furthermore, in Fig. 6(b) for the objects of interest in FG regions, the "PIBS" curve shows the performance nearly identical to that of the "Conventional" curve. These results show that the proposed PIBS can achieve power savings with negligible quality degradation. The black curve labeled "PIBS+ AFMC" and the dashed curve labeled "PIBS+FMCCOMP" cause considerable video quality degradation compared to the "Conventional" and "PIBS" curves due to the utilization of the FMC. However, the video quality degradation resulting from the FMC is considerably mitigated in "PIBS+ AFMC" by utilizing the proposed AFMC scheme, which achieves a significant BDPSNR improvement [34] of an average of 2.22dB for entire frames and 3.9dB for the objects of interest in FG regions compared to "PIBS+FMCCOMP". Furthermore, the PSNR difference between the "Conventional" and the "PIBS+AFMC" curves for the objects of interest in FG regions is much smaller than that for entire frames.

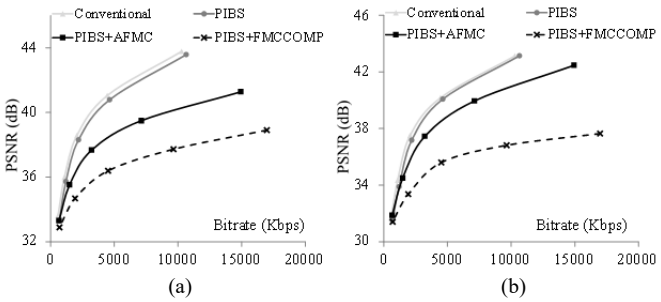


Fig. 6. Comparison of the R-D performance of the proposed system: (a) Entire frame. (b) Objects of interest in the FG region.

In order to verify the accuracy of the PIBS, Table VI shows the ratios of false-negative and false-positive errors. The second column presents the false-negative ratios of each case with the "Basic" algorithm, where fundamental conditions are utilized without the two additional processes in Fig. 3 and the third column presents those with the "Basic+Additional" algorithm in which all conditions including the two additional processes in Fig. 3 are performed. The results show that the two additional processes decrease the false-negative ratio by 3.1%. As a result, the proposed PIBS achieves outstanding accuracy with an average of 98.896%. In the fourth and fifth columns, the false-positive ratios of each case in the "Basic" and "Basic+Additional" algorithms are presented, respectively. The results show that the two additional processes in Fig. 3 increase the false-positive ratios by 0.092%, but the amount of the increase is very small and the average false-positive ratio of 0.441% is negligible. Furthermore, the false-positive errors are not crucial because false inclusions of the BG regions into FG regions cause a slight increase in the power consumption without an important data loss in the objects of interest. Fig. 7 shows the results of the PIBS in the two bright test sequences of "Access Door" and "Entrance" and two dark test sequences of "Light Switch" and "Low Light". In each result, the pink blocks in each figure denote the MBs which are determined as belonging to FG regions. These results show that the PIBS segments out FG objects with great accuracy regardless of brightness. Especially, false-positive errors in Figs. 7 (a) and (b) are removed at a significant rate by two additional processes in Fig. 3 as compared to the results shown in Figs. 4(a) and (b).

C. Comparison with Previous Video Surveillance Systems

In Table VI, for a further evaluation, the state-of-the-art techniques on video surveillance systems in [9] and [11] are compared with the proposed systems. The scheme proposed in [9] detects content differences in consecutive input frames by analyzing the color and moving correlation and reduces the power consumption in the video encoder. If a difference is detected, the MB is passed to the regular mode decision module; otherwise, the MB is determined as the skip-mode without mode decision. The scheme proposed in [11] utilizes multi-video codecs at a frame level such that the light-weight compression (LWC) which offers very low power consumption compresses video data for temporal storage before event detection and the H.264/AVC encoder is used only for permanent storage when events take place. As a result, a significant power consumption is saved if there is no change in the frame. The proposed systems and these state-of-the-art techniques are compared under the same conditions and the comparison results are presented separately according to the bright and dark conditions. The bright and dark results in Table VII are obtained by simulating with three bright test sequences ("Access Door", "Back Yark1", and "Entrance") and three dark test sequences ("Bund Nightscape", "Light Switch", and "Low Light") in Table III, respectively, which relatively have many FG regions. In the table, the BDPSNR and power savings (PS) compared to the conventional system are presented. In addition, the absolute value of the ratio of power savings for BDPSNR (|PS/BDPSNR|), which represents the effectiveness of each system, is also given. The values in bold font in this table represent the best results in each list.

TABLE VI
THE RATIOS OF FALSE-NEGATIVE AND FALSE-POSITIVE ERRORS

Case	False-negative (%)		False-positive (%)	
	Basic	Basic+Additional	Basic	Basic+Additional
Average	4.204	1.104	0.349	0.441

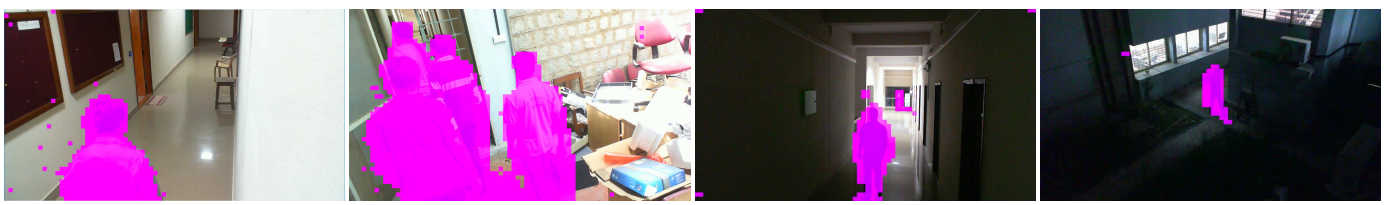
TABLE VII
COMPARISONS OF THE R-D PERFORMANCE IN FOREGROUND REGIONS AND THE POWER SAVINGS WITH THE STATE-OF-THE-ART TECHNIQUES

		PIBS	PIBS+AFMC	[9]	[11]
Bright	BDPSNR (dB)	**-0.422**	-1.164	-0.468	-0.695
	PS (%)	49.09	**59.19**	45.95	42.3
	|PS/BDPSNR|	**116.33**	50.85	98.18	60.86
Dark	BDPSNR (dB)	**-0.682**	-1.572	-0.847	-1.136
	PS (%)	51.64	**62.79**	47.51	36.51
	|PS/BDPSNR|	**75.72**	39.94	56.09	32.14

Simulation results show that the proposed PIBS system achieves greater power savings as compared to the two state-of-the-art techniques while also achieving better video quality than both schemes regardless of brightness. Thus, the proposed PIBS scheme offers the highest |PS/BDPSNR|. Furthermore, the amount of the difference of |PS/BDPSNR| between the proposed PIBS and two state-of-the-art techniques is further increased in the dark condition thanks to classification according to brightness. In comparison with [9], the proposed PIBS scheme achieves better performance because it significantly enhances the accuracy of segmentation by using various computation results such as MVs, PMVs, and SADs, effectively reduces the computation complexity according to the accurate segmentation results by considering the coding options for both current and next frames, and avoids the additional computation for BG subtraction. In comparison with [11], the proposed PIBS achieves better power saving because it performs BG subtraction at the MB level instead of the frame level. The scheme in [11] which operates at the frame level cannot achieve the power saving even if a slight change of the frame occurs. Furthermore, the proposed PIBS achieves better video quality by avoiding the unnecessary quality degradation by using the LWC in [11]. On the other hand, the proposed PIBS+AFMC system causes relatively significant quality degradation compared to other systems by utilizing the AFMC. However, it can achieve the highest power savings compared to those by other systems. For this reason, the proposed PIBS+AFMC system is suitable when a surveillance system requires very low-power operation.

V. CONCLUSION

Security concerns have led to the wider use of video surveillance systems and various commercial devices for video surveillance are being released. There are three main contributions in the proposed video surveillance system. First, the proposed video surveillance system performs very accurate BG subtraction without an additional module by utilizing various types of motion information. Second, the proposed system significantly reduces the encoding complexity based on the results from BG subtraction by utilizing the hardware pipelined structure of the video encoder. Experimental results show that maximum power savings of up to 51.2% can be achieved with minimum degradation of the image quality when compared to the results from a conventional system. Third, in order to achieve further power savings, an external BW is considered. An adaptive FMC significantly reduces unnecessary power consumption while retaining the video quality in the objects of interest. Accordingly, maximum power savings of as much as 61.6% is achieved. It is possible to apply the proposed PIBS scheme to the HEVC standard as well as the H.264/AVC standard because the fundamental operations in the HEVC standard and the H.264/AVC standard are very similar and thus, the proposed system is expected to contribute greatly to a wide use of battery-powered video surveillance system.



(a) (b) (c) (d)

Fig. 7. The results of the proposed background subtraction in two test sequences: (a) "Access Door". (b) "Entrance". (c) "Light Switch". (d) "Low Light".

REFERENCES

[1] C. R. del-Blanco, F. Jaureguizar, and N. Garcia, "An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications," *IEEE Trans. Consumer Electro.,* vol. 58, no. 3, pp. 857-862, Aug. 2012.

[2] P. Fossia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1545-1556, Sep. 2015.

[3] M. J. H. Loomans, C. J. Koeleman, and P. H. N. De With, "Low-complexity wavelet-based scalable image & video coding for home-use surveillance," *IEEE Trans. Consumer Electro.,* vol. 57, no. 2, pp. 507-515, May 2011.

[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[5] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol*., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[6] B. Dey, and M. K. Kundu, "Robust Background Subtraction for Network Surveillance in H.264 Streaming Video," *IEEE Trans. Circuits Syst. Video Technol*., vol. 23, no. 10, pp. 1695–1703, Oct. 2013.

[7] X. Zhang, Y. Tian, T. Huang, S. Dong, and W. Gao, "Optimizing the Hierarchical Prediction and Coding in HEVC for Surveillance and Conference Videos With Background Modeling," *IEEE Trans. Image Process.,* vol. 23, no. 10, pp. 4511-4526, Oct. 2014.

[8] F. Chen, H. Li, L. Li, D. Liu, and F. Wu, "Block-Composed Background Reference for High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, will be published.

[9] X. Jin and S. Goto, "Encoder adaptable difference detection for low power video compression in surveillance system," *Signal. Process.: Image Commun.*, vol. 26, no. 3, pp. 130–142, 2011.

[10] Y. Lee, J. Kim, and C. -M. Kyung, "Energy-Aware Video Encoding for Image Quality Improvement in Battery-Operated Surveillance Camera," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 2, pp. 310–318, Feb. 2012.

[11] H. Kim, C. E. Rhee, and H. J. Lee, "A Low-Power Video Recording System with Multiple Operation Modes for H.264 and Light-Weight Compression," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 603-613, Apr. 2016.

[12] P. Gorur and B. Amrutur, "Skip decision and reference frame selection for low-complexity H.264/AVC surveillance video coding," *IEEE Trans. Circuits Syst. Video Technol*., vol. 24, no. 7, pp. 1156–1169, Jul. 2014.

[13] S. Guo, Y. Wang, Y. Tian, P. Xing, and W. Gao, "Quality-progressive coding for high bit-rate background frames on surveillance videos," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2015, pp. 2764-2767.

[14] D. Liu, D. Zhao, X. Ji, and W. Gao, "Dual frame motion compensation with optimal long-term reference frame selection and bit allocation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 3, pp. 325–339, Mar. 2010.

[15] M. Paul, W. Lin, C.-T. Lau, and B.-S. Lee, "Explore and model better I-frames for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1242–1254, Sep. 2011.

[16] X. Li, Z. Wang, and X. Lu, "Surveillance Video Synopsis via Scaling Down Objects," *IEEE Trans. Image Process.,* vol. 25, no. 2, pp. 740-755, Feb. 2016.

[17] M. A. M. Razif, M. Mokji, and M. M. A. Zabidi, "Adaptive Search Window for Low Complexity Surveillance Video Coding," in *Proc. IEEE Int. Symp. Technol. Manage. Emer. Technol.*, Aug. 2015, pp. 360-363.

[18] M. Qin, Y. Lu, H, Di, and W. Huang, "A Background Basis Selection-Based Foreground Detection Method," *IEEE Trans. Multimedia*, vol. 18, no. 7, pp. 1283-1296, Jul. 2016.

[19] M. A. M. Razif, M. Mokji, and M. M. A. Zabidi, "Low Complexity Maritime Surveillance Video using Background Subtraction on H.264," in *Proc. IEEE Int. Symp. Technol. Manage. Emer. Technol.*, Aug. 2015, pp. 364-368.

[20] X. Zhang et al., "Fast and efficient transcoding based on low-complexity background modeling and adaptive block classification," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1769-1785, Dec. 2013.

[21] H. Kim, C. E. Rhee, and H.-J. Lee, "An Effective Combination of Power Scaling for H.264/AVC Compression," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 11, pp. 2685–2689, Nov. 2015.

[22] S. Lee, M. Chung, S. Park, and C. Kyung, "Lossless frame memory recompression for video codec preserving random accessibility of coding unit," *IEEE Trans. Consumer Electro*., vol. 55, no. 4, Nov. 2009.

[23] Y.-W. Huang et al., "A 1.3TOPS H.264/AVC single-chip encoder for HDTV applications," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2005, pp. 128-129.

[24] C. E. Rhee, J.-S. Jung, and H.-J. Lee, "A Real-time H.264/AVC Encoder with Complexity-Aware Time Allocation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 7, pp. 1848-1862, Dec. 2010.

[25] S. F. Tsai, C. H. Tsai, and L. G. Chen, "*Encoder Hardware Architecture for HEVC*," in High Efficiency Video Coding (HEVC), ser. Integrated Circuits and Systems, V. Sze, M. Budagavi, and G. J. Sullivan, Eds. Springer International Publishing, 2014, pp. 343-375.

[26] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243-250, Jun. 1996.

[27] P.-Y. Chen, "VLSI implementation for one dimensional multilevel lifting-based wavelet transform," *IEEE Trans. Comput*., vol. 53, no. 4, pp. 386–398, Apr. 2004.

[28] J. A. Sans, F. J. Manjón, A. L. J. Pereira, J. A. Gómez-Tejedor, and J. A. Monsoriu, "Oscillations studied with the smartphone ambient light sensor," *European Journal of Physics*, vol. 34, pp. 1349-1354, 2013.

[29] J. -S. Jung, D.-U. Moon, and H. -J. Lee, "Computation reduction of H.264/AVC motion estimation by search range adjustment and partial cost evaluation," in *Proc. of Int. Conf. Electronics, Information and Communication*, Jun. 2008, pp. 229-233.

[30] L. Su, Y. Lu, F. Wu, S. Li, and W. Gao, "Real-time video coding under power constraint based on H.264 codec," in *Proc. SPIE 6508, Visual Commun. Image Process. (VCIP)*, Jan. 2007, vol. 6508, pp. 1-12.

[31] S. Kim, D. Lee, J.-S. Kim, and H.-J. Lee, "A high-throughput hardware design of a one-dimensional SPIHT algorithm," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 392-404, Mar. 2016.

[32] H. Kim, C. E. Rhee, and H.-J. Lee, "Address generation for lossless frame memory compression in an H.264/AVC encoder," in *Proc. IEEE Int. Conf. Consumer Electon.*, Jan. 2013, pp. 570-571.

[33] *Micron LPDDR2 power calculator* [online]. Available: http://www.micron.com/~/media/documents/products/power-calculator/tn4201_lpddr2_system_power_calculator.xlsx

[34] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," in *VCEG-M33 of ITU-T Q6/16*, Austin TX, USA, April, 2001.

Hyun Kim received the B.S., M.S. and Ph.D. degrees in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea, in 2009, 2011 and 2015, respectively.

From 2015, he worked in the BK21 Creative Research Engineer Development for IT, Seoul National University, Seoul, Korea, where he is currently working as an Assistant Professor. His research interests are low-power algorithm and SoC design for video coding such as HEVC and H.264/AVC and configurable video coding for real time systems.

Hyuk-Jae Lee received the B.S. and M.S. degrees in Electronics Engineering from Seoul National University, Korea, in 1987 and 1989, respectively, and the Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, in 1996.

From 1998 to 2001, he worked at the Server and Workstation Chipset Division of Intel Corporation in Hillsboro, Oregon as a Senior Component Design Engineer. From 1996 to 1998, he was on the faculty of the Department of Computer Science of Louisiana Tech University at Ruston, Louisiana. In 2001, he joined the School of Electrical Engineering and Computer Science at Seoul National University, Korea, where he is currently working as a Professor. He is a founder of Mamurian Design, Inc., a fabless SoC design house for multimedia applications. His research interests are in the areas of computer architecture and SoC design for multimedia applications.