A Fast Normalized Cross Correlation-Based Block Matching Algorithm Using Multilevel Cauchy-Schwartz Inequality

Byung Cheol Song

This paper presents a fast block-matching algorithm based on the normalized cross-correlation, where the elimination order is determined based on the gradient magnitudes of subblocks in the current macroblock. Multilevel Cauchy-Schwartz inequality is derived to skip unnecessary block-matching calculations in the proposed algorithm. Also, additional complexity reduction is achieved re-using the normalized cross correlation values for the spatially neighboring macroblock because the search areas of adjacent macroblocks are overlapped. Simulation results show that the proposed algorithm can improve the speed-up ratio up to about 3 times in comparison with the existing algorithm.

Keywords: Fast full search, normalized cross-correlation, multilevel Cauchy-Schwartz.

Manuscript received June 1, 2010; revised Aug. 27, 2010; accepted Sept. 10, 2010.
This work was financially supported by Inha University Research Grant (INHA-39193).
Byung Cheol Song (phone: +82 32 860 7413, email: bcsong@inha.ac.kr) is with the School of Electronic Engineering, Inha University, Incheon, Rep. of Korea.
doi:10.4218/etrij.11.0110.0315

I. Introduction

Motion estimation has been employed by many video compression schemes to improve coding efficiency by removing the temporal redundancy that exists in video sequences. The block-matching algorithm (BMA) is the most popular approach applied to all video coding standards, such as MPEG and H.264/AVC, due to its structural simplicity. A full search algorithm (FSA) can be the best BMA for a given block distortion criterion, as it finds the block with minimum block-matching distortion among all candidates. However, its heavy computational cost is a crucial limiting factor in terms of software implementation as well as hardware implementation.

For several decades, many fast BMAs have been developed. These can be divided into two categories. The first category adopts pre-defined search patterns to locate candidate motion vectors (MVs) based on distortions of potential candidates [1]. The second category is entirely composed of optimal motion estimation methods, which can find the globally optimal MV within a search area [2]-[7]. Li and Salari proposed a well-known successive elimination algorithm (SEA) providing a decision boundary based on the sum norms of blocks to eliminate some checking points without the need for computationally intensive block matching [2]. Gao and others extended the SEA to a multilevel SEA (MSEA) that provides multiple levels of tighter boundaries using the sum norms of the macroblock (MB) and subblocks with reduced sizes [3]. MSEA reduces the necessary computation by detecting and rejecting unnecessary candidates from the lowest level to the highest level. Zhu and others proposed a fine granularity successive elimination (FGSE) scheme that extended the MSEA by adding greater detail levels [4]. Also, Liu and others presented an adaptive version of FGSE [5]. The FGSE is distinguished from the MSEA in that, if necessary, only a single subblock having the maximum complexity is chosen at each level and the subblock is partitioned into four smaller subblocks at the next level. Therefore, in the case of a 16×16 block, the total number of partition levels amounts to 86. Thus, the FGSE has more potential to prune out non-optimal candidates than MSEA before wholly performing block matching.

As a block distortion criterion, the sum of absolute differences (SAD) is commonly used in video compression. SAD is defined as

$$SAD(u,v)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left|I_{\mathrm{C}}(i,j)-I_{\mathrm{R}}(i+u,j+v)\right|, \qquad (1)$$

where (u, v) is a MV in the search area, and I_{C} and I_{R} denote the current and reference picture, respectively.

In addition to SAD and the sum of squared differences (SSD), the normalized cross correlation (NCC) is also a well-known similarity criterion. The NCC is defined as

$$NCC(u,v)=\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}I_{\mathrm{C}}(i,j)\cdot I_{\mathrm{R}}(i+u,j+v)}{\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}I_{\mathrm{C}}(i,j)^2}\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}I_{\mathrm{R}}(i+u,j+v)^2}}. \qquad (2)$$

The NCC is more robust than SAD and SSD under uniform illumination changes. Accordingly, it is widely used in object recognition and industrial inspection schemes.

Applying the NCC as the matching criterion to motion estimation leads to more uniform residuals. Hence, the NCC can improve subjective visual quality as well as coding efficiency in video compression [8]. Recently, visual quality measures focusing on the human visual system (HVS) have been devised in place of PSNR. Among these measures, structural similarity (SSIM) has become popular. Pan and others showed that the NCC helps increase the coding efficiency [8]. However, the NCC is a more complex criterion compared to SAD. Thus, a fast algorithm to speed up NCC-based block matching is required for computationally efficient video encoding.

This paper presents a fast BMA based on the NCC. In the proposed algorithm, the successive elimination order is determined based on the gradient magnitudes of subblocks in the current MB, as in the FGSE algorithm. The multilevel Cauchy-Schwartz inequality is derived to remove any unnecessary block-matching calculation in the proposed algorithm. Also, unnecessary block matching can be additionally avoided re-using the NCC values obtained from the motion estimation of the spatially adjacent MB. The experimental results show that the proposed algorithm outperforms the existing methods in terms of the search speed.

The remainder of this paper is organized as follows. Section II introduces the previous works, and section III presents the proposed algorithm in detail. We present experimental results in section IV. Finally, our conclusions are given in section V.

II. Previous Works and SSIM

This section introduces two earlier works related to the proposed algorithm: MSEA [3] and FGSE [4]. SSIM is also described in this section.

1. Review of MSEA

In MSEA, using what is known as a block sum pyramid, multiple boundary levels tighter than the SEA are obtained by dividing each $N\times N$ block into subblocks of $N/2\times N/2$ and further dividing these into $N/4\times N/4$ until 1×1 is reached.

In total, $L=\log_2 N$ boundary levels are established. All subblocks at one level are of the same size. At the l-th level, where $0\leq l\leq L$, the number of subblocks is $2^{2l}=4^l$, and the size of each subblock is $N_l\times N_l$, where $N_l=N/2^l$. For simplicity, let $C_{i,j}$ and $R_{i,j}$ denote $I_{\mathrm{C}}(i, j)$ and $I_{\mathrm{R}}(i+u, j+v)$ in (2), respectively. Then, based on the triangle inequality, the following can be obtained:

$$SAD\equiv\sum_{i,j}^{N-1}\left|C_{i,j}-R_{i,j}\right|\geq\cdots\geq\sum_{i,j}^{\frac{N}{2^{L-l}}-1}\left|C_{i,j}^l-R_{i,j}^l\right|\geq\cdots\geq\left|C^0-R^0\right|. \qquad (3)$$

Here, $X_{i,j}^l=\sum_{m=2i}^{2i+1}\sum_{n=2j}^{2j+1}X_{m,n}^{l+1}$, $X^0=\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}X_{m,n}$, and X represents C or R. It is clear from (3) that the boundary value increases along with boundary level l and is bounded by the matching error. In MSEA, one candidate is evaluated sequentially from the lowest level 0 to the highest level L. If a candidate survives until level L–1, its matching error, that is, SAD, will be finally calculated at level L. It is important to note that there are only a small number of candidates remaining for final SAD calculations.

2. Review of FGSE

The FGSE algorithm establishes a total of $L=(N^2-1)/3$

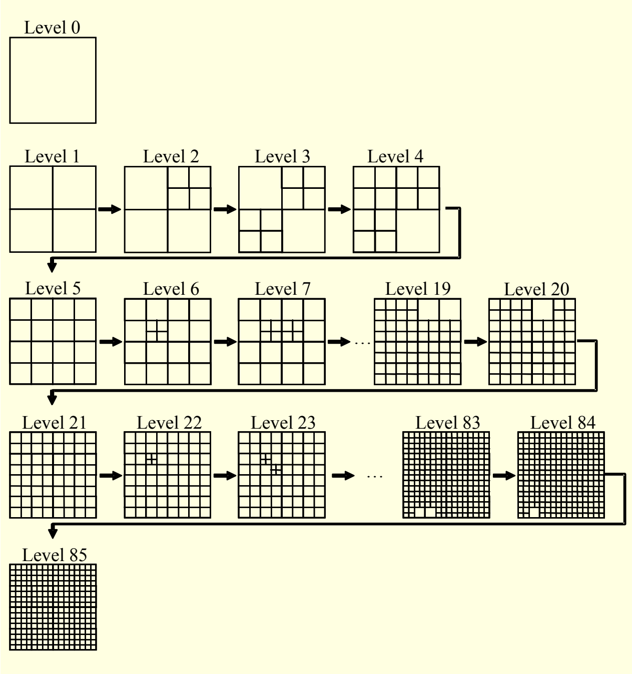

Fig. 1. Partition process in FGSE algorithm.

boundary levels by partitioning only one block of size $N \times N$ into four subblocks of size $N/2 \times N/2$ at each level according to a predefined rule. Here, subblocks of a smaller size are not partitioned further until all other larger subblocks are partitioned. Consequently, at the same level, subblocks of two different sizes may coexist. At the l-th level, where $0 \le l \le L$, the number of subblocks is $3l+1$ (Fig. 1). If SAD at level l is lower than the current minimum, a subblock in the MB is partitioned into four subblocks at the next level $l+1$. The selection of the subblock to be partitioned is based on the complexity of the image. In other words, image complexity measures such as gradient magnitudes are utilized to determine the partition order, where subblocks of a higher complexity will be partitioned first so that larger boundary values can be obtained as early as possible.

3. Structural Similarity (SSIM)

Recently, Wang and others proposed a measure based on image structural distortion termed SSIM [9]. The SSIM index is more consistent with human perception and is designed to measure structural information degradation, including the three comparison points of luminance, contrast, and structure. Its definition is as follows:

$$SSIM(x,y) = l(x,y)\cdot c(x,y)\cdot s(x,y). \tag{4}$$

Here, $l(x, y)$, $c(x, y)$, and $s(x, y)$ are the luminance, contrast, and structure measures, respectively. They are defined as

$$\begin{aligned} l(x,y) &= \frac{2\mu_x\mu_y + K_1}{\mu_x^2+\mu_y^2+K_1}, \\ c(x,y) &= \frac{2\sigma_x\sigma_y + K_2}{\sigma_x^2+\sigma_y^2+K_2}, \\ s(x,y) &= \frac{\sigma_{xy} + K_3}{\sigma_x\sigma_y+K_3}, \end{aligned} \tag{5}$$

where x and y are two vectors obtained from the image in the corresponding local windows, μ_x and μ_y are the sample means of x and y, respectively, and σ_x^2 and σ_y^2 are the variances of x and y, respectively. In addition, σ_{xy} denotes the covariance of x and y. K_1, K_2, and K_3 are small constants to avoid a condition in which the denominator is zero. They are recommended in an earlier study [9] as

$$K_1 = (K_1 D)^2,\ K_2 = (K_2 D)^2,\ K_3 = K_2/2, \tag{6}$$

where K_1, K_2 << 1, and D is the dynamic range of the pixel values. In addition, the higher the value of SSIM (x, y) is, the more similar the images x and y.

III. Proposed Algorithm

This section introduces a fast NCC-based BMA for video encoding. The NCC can be often a better criterion than the SAD in terms of SSIM. In order to prove this, a full search based on SAD was compared to that based on NCC, where the search range and matching block size were fixed to ±15 pixels and 16×16 pixels in terms of integer-pel accuracy. Table 1 shows the experimental results for eight CIF sequences. Here, the PSNR and SSIM values are computed from the motion-compensated version of the second frame of each sequence in order to evaluate the performance of the motion estimation only. The results show that the NCC provides better SSIM performance than the SAD, while the former is comparable to or slightly weaker than the latter in terms of PSNR. This means that the motion compensated images using NCC-based motion estimation are visually better than that those using SAD-based motion estimation even though the former does not outperform the latter in terms of PSNR.

The proposed BMA is a type of NCC version of FGSE that uses multilevel Cauchy-Schwartz inequality [9]. As in an earlier study [10], it is possible to apply the multilevel Cauchy-Schwartz inequality based on a L_2-norm pyramid to the numerator of (2) as follows:

$$\sum_{i,j}^{N-1} C_{i,j}\cdot R_{i,j} \le \cdots \le \sum_{i,j}^{\frac{N}{2^{L-l}}-1} C_{i,j}^l \cdot R_{i,j}^l \le \cdots \le C^0\cdot R^0. \tag{7}$$

Table 1. Comparison of PSNR and SSIM.

	PSNR (dB)		SSIM	
	SAD	NCC	SAD	NCC
Foreman	34.71	33.89	0.9153	0.9335
News	33.84	33.78	0.9537	0.9587
Coastguard	31.89	31.90	0.9126	0.9145
Containership	37.27	37.26	0.9798	0.9807
Hall Monitor	38.49	38.10	0.9294	0.9538
Mobile	24.36	24.39	0.8629	0.8914
Stefan	26.85	26.82	0.9069	0.8973
Tempete	27.59	27.54	0.9104	0.9161

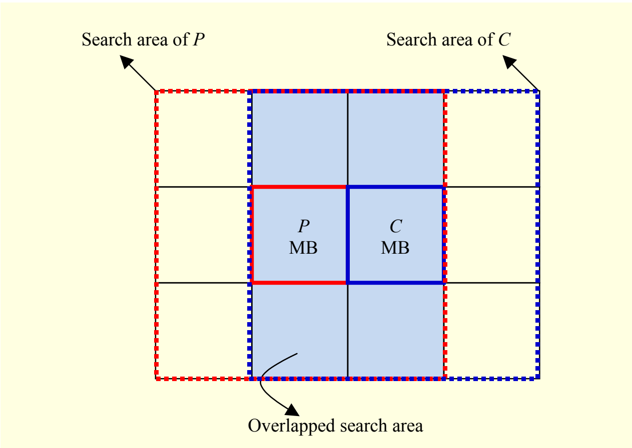

Fig. 2. Overlapping of search areas of adjacent MBs.

Here, $X_{i,j}=X^{L}_{i,j},\ X^{l}_{i,j}=\sqrt{\sum_{m=2i}^{2i+1}\sum_{n=2j}^{2j+1}\left(X^{l+1}_{m,n}\right)^2}$ and $L=\log_2 N$.

Based on (7), it is possible to derive the following inequality:

$$NCC^{0}\geq NCC^{1}\geq\cdots\geq NCC^{l}\geq\cdots\geq NCC^{L}. \tag{8}$$

In (8), NCC^l is defined as

$$NCC^{l}=\frac{\sum_{i,j}^{\frac{N}{2^{L-l}}-1} C^{l}_{i,j}\cdot R^{l}_{i,j}}{\sqrt{\sum_{i,j}^{N-1} C_{i,j}{}^{2}}\sqrt{\sum_{i,j}^{N-1} R_{i,j}{}^{2}}}. \tag{9}$$

As in the aforementioned studies [3], [10], a multilevel successive elimination algorithm is developed to determine the best MB with the maximal NCC value according to (8).

Furthermore, the total number of partition levels can be extended to 86, that is, $0\leq l\leq 85$, by partitioning each of four subblocks at a certain level one by one into another four subblocks at the next level in the descending order of the complexities of larger-sized subblocks, as in a more recent study [4]. It is easily proved that (7) and (8) still hold for sequential partition levels, as shown in Fig. 1. Here, the partition order is determined in the manner used in the above-mentioned study [4] by employing the gradient magnitude as an image complexity measure.

Note that the search areas of the current MB and its left neighboring MB are mostly overlapped (Fig. 2). For the search range of ±16 pixels in Fig. 2, two adjacent MBs share about 2/3 of their entire search areas. We propose a method to prevent unnecessary computations of NCC using the NCC values obtained from motion estimation of the spatially neighboring MB. Assume that P_{ij} denotes a pixel at (i, j) in the spatially adjacent 16×16 MB, that is, $I_C(i, j\text{–}16)$ (Fig. 2). For instance, we consider an example when l is equal to 1. For the same candidate in the overlapped search area, the NCC^1s of P and C are as

$$\frac{\sum_{i,j=0}^{1} C^{1}_{i,j}\cdot R^{1}_{i,j}}{\|C\|_2\cdot\|R\|_2}\ \text{and}\ \frac{\sum_{i,j=0}^{1} P^{1}_{i,j}\cdot R^{1}_{i,j}}{\|P\|_2\cdot\|R\|_2}, \tag{10}$$

where $\|C\|_2$ stands for the L_2-norm of the current MB C. Since two adjacent MBs generally have significant spatial correlation, the possibility that P_{ij} and C_{ij} are equivalent is high. For example, the two NCC^1s of P and C in (10) can be the same at level 1. Then, we can replace the NCC^1 of C with the NCC^1 of P without computation. Note that as l becomes smaller, the probability that P^l_{ij} and C^l_{ij} are equivalent becomes higher. Thus, the NCC values of the block candidates in the overlapped search area are stored during motion estimation of the spatially neighboring MB, and if P^l_{ij} and C^l_{ij} are equivalent at level l of motion estimation of the current MB, the NCC^l of C is replaced with the stored NCC^l of P.

On the other hand, it is very important to choose the initial search point and search pattern properly in the search area. In this paper, the initial search point is set to the median MV of MVs of three spatially adjacent MBs (Fig. 3(a)). In order to maximize the elimination effect, a spiral search pattern with the initial MV is presented as the starting point, as in the example shown in Fig. 3(b).

As a result, the proposed algorithm is summarized as follows:

1) Offline pre-processing: Build L_2-norm pyramids for the reference frame.

2) Online processing: For each MB in the current frame, the following procedure is applied.

(a) Compute the L_2-norm pyramid of the current MB.

(b) Compute the NCC corresponding to the initial MV and set the current maximum cost (C_{max}) to the computed NCC.

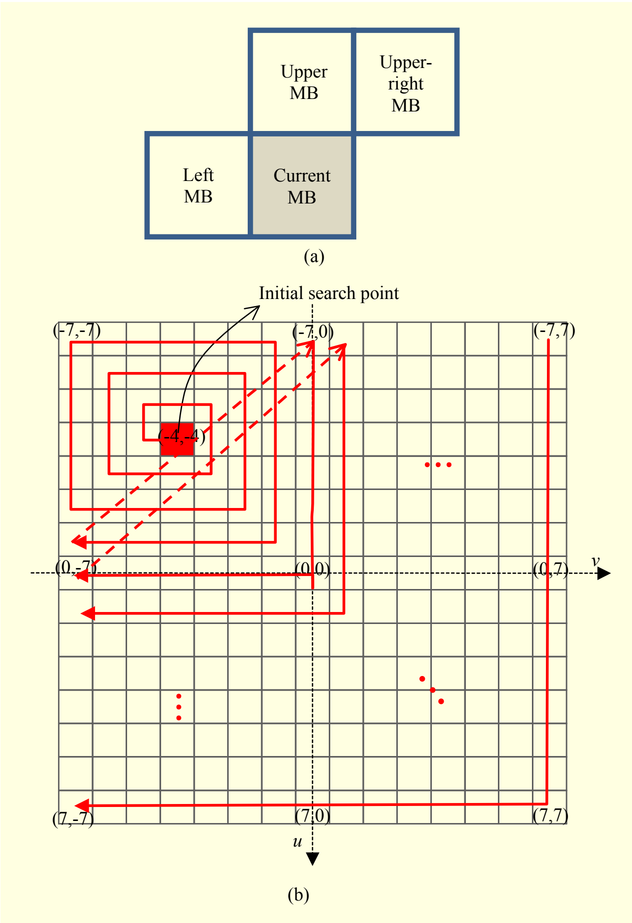

Fig. 3. (a) Neighboring MBs that are used to determine initial MV and (b) initial MV-centric search pattern.

(c) Set l to 0. Compute NCC^l. If C^0 and P^0 are the same at this level and the NCC for the reference block is available, employ the stored NCC instead of the NCC^l computation.

(d) If $NCC^l \geq C_{max}$, replace l with l+1 and find a subblock with the largest complexity according to the given partition rule. Otherwise, reset l to 0 and go to step (c) with the next MV candidate. If no additional MV candidate exists, go to step (h).

(e) Partition the subblock having the largest complexity and compute its corresponding NCC. Compute NCC^l by updating the NCC of the partitioned subblock only. If C^l and P^l are same at this level and the NCC for the reference block is available, employ the stored NCC instead of NCC^l computation. If l is equal to 85, go to step (g).

(f) If $NCC^l \geq C_{max}$, replace l with l+1, find a subblock with the largest complexity according to the given partition rule, and go to step (e). Otherwise, reset l to 0 and go to step (c) with the next MV candidate. If no more MV candidates exist, go to step (h).

(g) If $NCC^{85} \geq C_{max}$, update C_{max} to the NCC^{85}. Reset l to 0 and go to step (c) with the next MV candidate. If no additional MV candidates exist, go to step (h).

(h) Select the MV corresponding to the final C_{max} as the best match to the current MB.

The computed NCC values are stored for the next MB.

IV. Experimental Results

In order to evaluate the performance of the proposed algorithm, eight CIF (352×288) video sequences were used, that is, Foreman, Containership, Coastguard, Mobile and Calendar (Mobile), News, Stefan, Tempete, and Hall Monitor. The luminance components of the first 100 frames of these video sequences are adopted in the simulation. The MB size and the search range are fixed to 16×16 and ±15 pixels in both the horizontal and vertical directions, respectively. Simulation was performed on a dual core CPU at 2.66 GHz.

The proposed algorithm was compared with FSA and MSEA [3]. All the algorithms are based on the NCC distortion criterion. To do this, the MSEA was revised based on NCC rather than SAD. Here, since all the algorithms provide the same SSIM and the same PSNR, we do not provide those results in this section. Table 2 summarizes the average number

Table 2. Average number of operations per MB for Mobile sequence.

	FSA	MSEA	Proposed
ADD/SUB	443,360	21,750	11,483
MLP	445,098	21,250	9,344
DIV	869	2,191	838
COMP	869	2,190	4,038
SQRT	869	1,067	1,067
Total	891,065	48,448	26,770

Table 3. Computational complexity comparison of various sequences.

	FSA	MSEA		Proposed	
	ANOP	ANOP	Speed-up ratio	ANOP	Speed-up ratio
Foreman	891,065	52,043	17.1	23,964	37.2
News	891,065	25,161	35.4	13,405	66.5
Coastguard	891,065	47,313	18.8	27,485	32.4
Containership	891,065	47,063	18.9	28,993	30.7
Hall Monitor	891,065	99,091	9.0	33,560	26.6
Mobile	891,065	48,448	18.4	26,770	23.3
Stefan	891,065	43,603	20.4	25,543	34.9
Tempete	891,065	28,232	31.6	14,582	61.1

of operations per MB (ANOP) for the 'Mobile' video sequence. Here, all operations including addition (ADD), subtraction (SUB), multiplication (MLP), division (DIV), comparison (COMP), and square root (SQRT) are considered. Table 2 shows that the proposed algorithm considerably reduces the complexity of the MSEA. The proposed algorithm substantially reduces the number of ADD and MLP operations, although there are slight increases of the number of COMP operations due to additional comparisons based on (10). Table 3 compares the ANOP and the speed-up ratio over the FSA for various video sequences.

Here, the computational complexity of the proposed algorithm is significantly lower than the existing algorithms, regardless of the sequence type. For instance, the proposed algorithm can improve the speed-up ratio up to about 3 times (Hall Monitor) in comparison with the MSEA.

V. Conclusion

This paper proposes a fast BMA based on NCC, where multilevel Cauchy-Schwartz inequality is employed to skip unnecessary block-matching calculation and the elimination order is determined based on the image complexities of subblocks in the current MB. Also, additional complexity reduction is possible re-using the NCC values for the spatially neighboring MB. The proposed NCC-based algorithm considerably reduces the computational complexity of a video encoder.

References

[1] S. Zhu and K.K. Ma, "A New Diamond Search Algorithm for Fast Block Matching Motion Estimation," *IEEE Trans. Image Process.*, vol. 9, no. 2, 2000, pp. 287-290.

[2] W. Li and E. Salari, "Successive Elimination Algorithm for Motion Estimation," *IEEE Trans. Image Process.*, vol. 4, no. 1, 1995, pp. 105-107.

[3] X.Q. Gao, C.J. Duanmu, and C.R. Zou, "A Multilevel Successive Elimination Algorithm for Block Matching Motion Estimation," *IEEE Trans. Image Process.*, vol. 9, no. 3, 2000, pp. 501-504.

[4] C. Zhu, W.S. Qi, and W. Ser, "Predictive Fine Granularity Successive Elimination for Fast Optimal Block-Matching Motion Estimation," *IEEE Trans. Image Process.*, vol. 14, no. 2, 2005, pp. 213-221.

[5] S.W. Liu, S.D Wei, and S.H. Lai, "Fast Optimal Motion Estimation Based on Gradient-Based Adaptive Multilevel Successive Elimination," *IEEE Trans. Circ. Syst. for Video Technol.*, vol. 18, no. 2, 2008, pp. 263-267.

[6] J.N. Kim and T.S. Choi, "Adaptive Matching Scan Algorithm Based on Gradient Magnitude for Fast Full Search in Motion Estimation," *IEEE Trans. Consum. Electron.*, vol. 45, no. 3, 1999, pp. 762-772.

[7] C.Y. Choi and J.C. Jeong, "New Sorting-Based Partial Distortion Elimination Algorithm for Fast Optimal Motion Estimation," *IEEE Trans. Consum. Electron.*, vol. 55, no. 4, Nov. 2009, pp. 2335-2340.

[8] W.H. Pan et al., "A Hybrid Motion Estimation Approach Based on Normalized Cross Correlation for Video Compression," *IEEE Proc. ICASSP*, 2008, pp. 1037-1040.

[9] Z. Wang et al., "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, 2004, pp. 600-612.

[10] B.C. Song and J.B. Ra, "A Fast Search Algorithm for Vector Quantization Using L_2-norm Pyramid of Codewords," *IEEE Trans. Image Process.*, vol. 11, no. 1, 2002, pp. 10-15.

Byung Cheol Song received the BS, MS, and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 1994, 1996, and 2001, respectively. From 2001 to 2008, he was a senior engineer at Digital Media R&D Center, Samsung Electronics Co., Ltd., Suwon, Rep. of Korea. In March 2008, he joined the School of Electronic Engineering, Inha University, Incheon, Korea, and currently is an assistant professor. His research interests are in the general areas of video coding, video processing, super-resolution, stereo vision, multimedia system design, image coding, content-based multimedia retrieval, and data mining.