Improved B-Picture Coding Scheme for Next Generation Video Compression

Seung-Jin BAEK†, Seung-Won JUNG†, Hahyun LEE††, Hui Yong KIM††, Nonmembers, and Sung-Jea KO††a, Member

SUMMARY In this paper, an improved B-picture coding algorithm based on the symmetric bi-directional motion estimation (ME) is proposed. In addition to the block match error between blocks in the forward and backward reference frames, the proposed method exploits the previously-reconstructed template regions in the current and reference frames for bi-directional ME. The side match error between the predicted target block and its template is also employed in order to alleviate block discontinuities. To efficiently perform ME, an initial motion vector (MV) is adaptively derived by exploiting temporal correlations. Experimental results show that the number of generated bits is reduced by up to 9.31% when the proposed algorithm is employed as a new macroblock (MB) coding mode for the H.264/AVC standard.

key words: B-picture coding, bi-directional motion estimation, H.264/AVC, decoder-side motion vector derivation

1. Introduction

Recently, the increasing popularity of broadcasting services and mobile devices which support high definition video contents is creating greater need for higher coding performance than existing coding standards can provide. Many methods have been introduced to improve the coding efficiency of the state-of-the-art standard H.264/AVC over the last several years. Finally, the joint collaborative team on video coding (JCT-VC) formed by the ISO/IEC MPEG and ITU-T/VCEG has recently begun standardization of a new generation video codec, called high efficiency video coding (HEVC), which aims to further reduce the bit rate by half that of H.264/AVC [1]–[3].

In video coding standards based on motion-compensated (MC) prediction, B-picture coding generally provides better overall compression performance than P-picture coding. B-picture is inter-coded using both the previous and following frames as references; the prediction signal of a block in a B-picture is obtained from a linear combination of the two MC signals obtained from bi-directional prediction [4], [5]. This prediction technique is referred to as MC interpolation (MCI).

In B-picture coding used in the previous video coding standards such as H.263 [6] and MPEG-4 [7], the motion vector (MV) and the prediction residual signal generated by the block-wise MCI technique are encoded and transmitted to the decoder. In order to improve the B-picture coding efficiency, H.264/AVC [8] adopts the B_SKIP prediction mode, where transmission of both motion and residual data is not required. In the B_SKIP mode, similarly to the P_SKIP mode, the reconstructed signal is replaced with a prediction signal using a single flag.

Bi-directional motion estimation (ME) has been widely used for MCI [9]–[11]. In [9] and [10], we proposed a symmetric bi-directional ME in order to find the MV of the to-be-interpolated block using the previous and following frames. In the area of video compression, especially B-picture coding of H.264/AVC, Murakami et al. [12], [13] and Mys et al. [14] proposed new macroblock (MB) coding modes in which the reconstructed signal is obtained from the MCI technique using bi-directional ME. Compared to the DIRECT mode that simply predicts the MV using the MVs of spatially or temporally neighboring blocks, these bi-directional ME methods can obtain relatively-accurate motion parameters. Moreover, it is not necessary to encode the motion and residual data since an identical MCI result can be produced in both the encoder and decoder by using the reconstructed previous and following frames. However, the bi-directional ME method often produces unreliable motion data, since only the previous and following frames are exploited in the ME process. As a result, the image quality of the interpolated signal can be degraded, resulting in inefficiency in terms of rate-distortion.

In block-based video coding, the left- and upper-blocks of the current coded block have already been reconstructed, since blocks in a frame are encoded in raster scan order. These previously-reconstructed pixels are used to construct a template in the template matching (TM) technique [15]–[17], where the prediction signal is obtained by minimizing the matching error between the template of the current block and that of a block in the search region. In the proposed method, the TM approach is adopted for bi-directional ME. In addition to the block match error, the proposed bi-directional ME scheme compares the reconstructed template region of the current block with that of the blocks in the neighbor reference frames. The side match error [10], [18], [19] between the predicted target block and its template is also employed in order to alleviate the block discontinuities. Moreover, the initial MV is adaptively derived by exploiting temporal correlations. The proposed algorithm...
finds the best MV within a small search range (SR) based on the initial MV. Simulation results demonstrate that there is significant improvement of B-picture coding performance when the proposed scheme is applied to H.264/AVC as a new coding mode.

The remainder of this paper is organized as follows. In Sect. 2, we describe the bi-directional ME scheme that is generally used in MCI-based applications. Our proposed method is detailed in Sect. 3. Simulation results of the proposed method are presented in Sect. 4. Section 5 concludes this paper.

2. Symmetric Bi-directional Motion Estimation

Since the proposed coding technique is based on MCI using bi-directional ME, we briefly review bi-directional ME before introducing the proposed method.

In the block-wise MCI, estimation of MVs of blocks in a to-be-interpolated frame is not straightforward, since the original pixels of the frame do not exist. In our previous work [9], the forward and backward MVs of the block in the to-be-interpolated frame are estimated by comparing a block at a shifted position in the forward (previous) frame with its corresponding block at the linearly-opposite position in the backward (following) frame.

Let \( B \) be a target block in the to-be-interpolated frame \( f_f \) between the forward and backward frames, \( f_f \) and \( f_b \). Figure 1 illustrates the symmetric bi-directional ME scheme. Based on the assumption of a linear motion trajectory, the forward and backward vectors, \( v_f \) and \( v_b \), of block \( B \) are defined as follows:

\[
\begin{align*}
  v_f &= +v, \\
  v_b &= -v,
\end{align*}
\]

where \( v \) denotes a candidate MV of \( B \). For each \( v \), the mean absolute block match error \( E_B \) between the pixels in \( f_f \) and those in \( f_b \) is calculated by

\[
E_B(v, B) = \frac{1}{N_B} \sum_{c \in B} |f_f(c + v) - f_b(c - v)|,
\]

where \( c \) denotes the pixel coordinates of \( B \) in the frame and \( N_B \) indicates the number of pixels in the target block. Then, the MV \( v^* \) that provides the minimum sum of absolute difference (SAD) value is determined as the best MV by

\[
v^* = \arg \min_v E_B(v, B).
\]

Finally, we can obtain the target predicted block \( B \) in the interpolated frame from a linear combination of blocks in the forward and backward frames as follows:

\[
f_f(c) = \frac{1}{2} \left( f_f(c + v^*) + f_b(c - v^*) \right).
\]

In [13], the predicted MV in the DIRECT mode is replaced by the best MV resulting from bi-directional ME with cost \( E_B \), which improves coding efficiency in the non-hierarchical IBBP coding structure. In [14], forward ME is performed such that the best matching block in the forward frame is exhaustively searched with respect to the co-located block in the backward frame as an anchor block. Then the resultant MV of the forward ME serves as the initial MV, and this is used with \( E_B \) for symmetric bi-directional motion refinement.

3. Proposed New Coding Method

In order to not only improve the prediction accuracy of bi-directional ME but also minimize the increase of computational complexity, we propose an effective method using the block match criterion and initial MV derivation.

3.1 New Block Match Criterion for Bi-directional ME

In addition to the block match error between the blocks in the reference frames, the proposed block match criterion uses two additional cost functions for bi-directional ME. First, we adopt the TM error as shown in Fig. 2. For the current block \( B \), the adjacent inverse-L shaped region \( T_f \), which consists of the previously-reconstructed pixels, is used as a template. In the ME process, the template regions \( T_f \) and \( T_b \) that adjoin the candidate blocks \( B_f \) and \( B_b \), respectively, in the forward and backward frames, are used to formulate the mean absolute bi-directional template match error \( E_T \) as follows:

\[
E_T(v, T) = \frac{1}{N_T} \sum_{c \in T} \left| \frac{f_f(c + v) + f_b(c - v)}{2} - f_f(c) \right|,
\]

where \( T \) denotes the set of pixel coordinates in the template region and \( N_T \) indicates the number of pixels in the template region. The adoption of this cost means that the error between the reconstructed neighboring template in the current frame and the average of templates in the forward and backward frames is exploited in the ME process.

Second, the side match error is adopted in the proposed bi-directional ME scheme. The compensated block obtained by the proposed ME scheme needs to be smoothly connected to the upper and left reconstructed pixels. To this

![Fig. 1](image-url) Symmetric bi-directional ME.
end, the spatial smoothness across block boundaries is measured. Before calculating the side match error for each candidate MV $v$, the target block signal is first interpolated by a linear combination of blocks in the forward and backward frames with Eq. (5). Let $B_v$ be the interpolated block obtained by using each candidate MV $v$. The mean absolute side match error $E_S$ between the interpolated block $B_v$ and the template region $T$ is calculated by

$$E_S(v, T) = \frac{1}{N_S} \sum_{k=0}^{N_S-1} |f_i(t_k) - f_i(b_{k,v})|,$$  

(7)

where $t_k$ and $b_{k,v}$, respectively, indicate the positions of the $k$-th pixel at the boundary of the template region $T$ and the interpolated block $B_v$, as shown in Fig. 3, and $N_S$ denotes the number of pixels at the left and upper block boundaries in $B_v$.

Finally, we combine the three cost functions in Eqs. (3), (6), and (7) to find the best MV:

$$E_P(v, B, T) = w_B \cdot E_B(v, B) + w_T \cdot E_T(v, T) + w_S \cdot E_S(v, T),$$  

(8)

where $w_B$, $w_T$, and $w_S$, respectively, denote the weighting factors for the symmetric bi-directional block match, template match, and side match errors, and $w_B + w_T + w_S = 1$. In order to obtain the proposed weighting factors, we performed simulation using video sequences with 25 frames which are recommended in [20]. We empirically verified that the best performance was achieved when $w_B = 0.5$, $w_T = 0.35$, and $w_S = 0.15$. These weighting factors were used to evaluate our proposed method.

### Table 1

| Sequence       | SR (Q-pel) | $\Delta \text{Bitrate}$ | $\Delta T_E$ | $\Delta T_D$ |
|----------------|------------|--------------------------|--------------|--------------|
| Kimono         | ±2         | -3.55%                   | 2.32%        | 60.77%       |
| (1920×1080)    | ±16        | -10.23%                  | 41.92%       | 822.74%      |
| BQMall         | ±2         | -2.40%                   | 2.42%        | 36.43%       |
| (832×480)      | ±16        | -7.21%                   | 37.39%       | 508.93%      |

$\Delta T_E$: increase of mode decision time at the encoder  
$\Delta T_D$: increase of motion compensation time at the decoder

### 3.2 Adaptive Initial MV Derivation

The SR of bi-directional ME provides a tradeoff between ME performance and computational complexity. In other words, an increase of the SR in bi-directional ME leads to a more accurate MV, however, the computational cost of evaluating the matching error also increases. Especially, the complexity of the decoder drastically increases, since the decoder finds the best matching block instead of extracting the MV from a bitstream. In order to investigate how the SR affects the coding performance, we tested the proposed MCI method, which was integrated into the Joint Model (JM) 16.2 reference software [21] as a new coding mode, with the SR of ±2 and ±16 in the quarter-pel unit (see Sects. 3.3 and 4 for the experimental environment). Table 1 shows the performance of the proposed method in comparison to the original H.264/AVC. As shown in Table 1, the increase of SR not only improves coding performance but also increases decoding complexity. This is a crucial problem of the decoder-side MV derivation approach.

If the best MV is roughly predictable, the SR can be significantly reduced. We exploit the fact that the temporally co-located blocks in the forward and backward frames are strongly correlated with the block in the current frame. To this end, we adaptively assign an initial MV by exploiting the reference blocks of neighbor frames. The proposed bi-directional ME is then performed as a motion refinement with a small SR at the position indicated by the initial MV.
Figure 4 shows the flowchart of the proposed adaptive initial MV derivation. We categorize the conditions of temporal neighbors of the current block into the following four cases:

- Both co-located blocks are intra-coded.
- Only the forward co-located block is inter-coded.
- Only the backward co-located block is inter-coded.
- Both co-located blocks are inter-coded.

Let \( B_F \) and \( B_B \) be the co-located blocks in the forward and backward frames, respectively. We assume that there are no temporal correlations among \( B_F \), \( B_B \), and the current block if both the \( B_F \) and \( B_B \) blocks are intra-coded. In this case, a motion vector predictor (MVP) of H.264 standard, which is derived using spatially neighboring MVs, is set to the initial MV.

In the other cases, temporal relationships are considered in the initial MV derivation. If only \( B_F \) is inter-coded, the MV of \( B_F \), \( MV_F \), is set to the initial MV. In the opposite case, i.e., if only \( B_B \) is inter-coded, the MV of \( B_B \), \( MV_B \), is set to the initial MV.

If both \( B_F \) and \( B_B \) are inter-coded, each corresponding MV can be a candidate of the initial MV. Let \( D_F \) (\( D_B \)) be the distance between the current frame and the reference frame of the forward (backward) co-located block. In general, temporal relationships tend to be considerably reduced when the temporal distance between frames is increased [22]. Hence, \( D_F \) and \( D_B \) are compared and the initial MV is determined accordingly. Specifically, if \( D_B \leq D_F \), \( MV_B \) is set to the initial MV of the current block. Otherwise, \( MV_F \) is set to the initial MV of the current block.

The resultant MV obtained from the above decision procedure is scaled in order to normalize its temporal distance to the distance between the nearest forward frame and the current frame. Finally, the proposed algorithm finds the best bi-directional MV based on the cost, which is calculated by

\[
E_P(v + v_i, B, T) = w_B \cdot E_B(v + v_i, B) + w_T \cdot E_T(v + v_i, T) + w_S \cdot E_S(v + v_i, T),
\]

and the best MV \( v_P^* \) is determined by

\[
v_P^* = \arg \min_v E_P(v + v_i, B, T),
\]

where \( v \) and \( v_i \) denote the candidate MV within the SR and the proposed initial MV, respectively.

Let \( v_0^* \) be the best MV resulting from the proposed bi-directional ME scheme when the SR = ±16 without using the initial MV. To evaluate the accuracy of the proposed initial MV, we measured the distances between \( v_0^* \) and \( v_i \) along the x and y-axes, \( d_x \) and \( d_y \). Table 2 lists the probability distribution of \(|d_x| < 1 \) and \(|d_y| < 1 \) (in the quarter-pel unit) when the proposed coding mode is selected as the best mode of the current MB. As shown in Table 2, the probability that the positions indicated by both MVs are very close is considerably high. This result demonstrates that the proposed algorithm can find an accurate MV within a small SR. In this paper, the SR of the proposed algorithm is set to ±1 in the quarter-pel unit.

### Table 2: Probability distribution of \(|d_x| < 1 \) and \(|d_y| < 1 \).

| Sequence (size)      | Distribution (%) |
|----------------------|------------------|
| City (352×288)       | 91.70            |
| BlowingBubbles (416×240) | 95.55         |
| PartyScene (832×480) | 89.05            |
| Cactus (1920×1080)  | 89.53            |

The data values for each sequence are the averaged values when QP=22, 27, 32, and 37 are used.

3.3 Implementation

We utilize the proposed method as a new MB coding mode, which requires a slight modification of the syntax structure of the H.264/AVC standard. In the syntax of the current standard, the flag \( mb\_skip\_flag \) signals whether the best mode of the MB is DIRECT. When \( mb\_skip\_flag = 1 \), we add the \( MCI\_skip\_flag \) indicating whether the proposed coding mode has been selected or not.

- If \( MCI\_skip\_flag = 1 \), the MB is coded in the proposed coding mode without any residual signal (\( MCI\_SKIP \) mode).
- If \( MCI\_skip\_flag = 0 \), the MB is coded in the DIRECT mode without any residual signal (B\_SKIP mode).

As a result, our new coding mode only produces one additional bit when no residual signal needs to be transmitted.
4. Performance Analysis

In order to evaluate the coding performance, the proposed algorithm is compared with Murakami’s [13] and Mys’s [14] methods. All methods were implemented in the JM 16.2 reference software [21]. Several CIF (352 × 288), WQVGA (416 × 240), WVGA (832 × 480), and 1080p (1920 × 1080) video sequences were tested using various quantization parameters (QPs). In all experiments, ME is conducted in the quarter-pel precision. Experiments were based on the recommended simulation conditions [20]. The detailed simulation conditions are described in Table 3.

4.1 Compression Performance

Table 4 lists the respective performances of the conventional and proposed methods, which are compared to the H.264/AVC standard. In all experiments, the MCI block size is fixed to 16 (equal to the MB size). The SR of ±1 and ±4 in the quarter-pel unit are used for both Murakami’s and Mys’s methods. For the proposed method, the template size is set to 4. In order to demonstrate the performance of each method, we calculate the Bjøntegaard delta bit rate ($\Delta$Bitrate) [23] which provides the average difference in bit rate. As shown in Table 4, Murakami’s method improves the coding efficiency by 1.84% and 3.49% on average for the SR of ±1 and ±4, respectively, compared to H.264/AVC. Mys’s method improves the coding efficiency by 2.05% and 3.78% on average for the SR of ±1 and ±4, respectively, compared to H.264/AVC. Compared with the conventional methods,

### Table 3 Simulation configuration.

| Option                  | Value                        |
|-------------------------|------------------------------|
| Sequence resolution     | CIF, WQVGA, WVGA, 1080p      |
| Quantization parameters | QPP=22,27,32,37              |
| Profile                 | High profile                 |
| Coding structure        | Hierarchical B structure     |
| ME precision            | Quarter-pel unit             |
| Entropy coding          | CABAC                        |
| Rate-distortion (RD) optimization | Enabled                   |
| RD picture decision     | Enabled                       |
| RD optimization quantization | Disabled               |

### Table 4 Performance comparison of the conventional and proposed methods.

| Sequence | $\Delta$Bitrate (%) | $\Delta T_B$ (%) |
|----------|----------------------|-----------------|
|          | Murakami’s          | Mys’s           | Proposed | Murakami’s | Mys’s | Proposed |
|          | SR ±1 | SR ±4 | SR ±1 | SR ±4 | SR ±1 | SR ±4 | SR ±1 | SR ±4 | SR ±1 | SR ±4 | SR ±1 | SR ±4 | SR ±1 | SR ±4 | Proposed |
| CIF (352×288) | Soccer | -0.47 | -1.39 | -0.36 | -1.48 | 2.52 | 9.48 | 21.83 | 153.85 | 395.22 | 24.32 |
|            | City    | -3.49 | -4.39 | -3.74 | -4.48 | -4.79 | 10.20 | 19.27 | 173.93 | 446.14 | 16.13 |
|            | Foreman | -2.72 | -3.73 | -2.85 | -3.88 | -4.49 | 20.58 | 38.07 | 181.89 | 460.09 | 33.66 |
| Average on CIF sequences | -2.23 | -3.17 | -2.32 | -3.28 | -3.93 | 13.42 | 26.39 | 169.89 | 433.82 | 24.70 |
| WQVGA (416×240) | BasketballPass | -1.94 | -2.89 | -2.08 | -2.97 | -3.60 | 28.39 | 56.37 | 174.46 | 430.48 | 42.63 |
|            | BQSquare | -3.77 | -3.97 | -4.13 | -4.12 | -4.37 | 23.66 | 39.85 | 154.29 | 376.25 | 28.57 |
|            | RaceHorses | -1.49 | -3.42 | -1.67 | -3.71 | -4.02 | 14.17 | 35.06 | 126.80 | 319.52 | 37.95 |
| Average on WQVGA sequences | -2.40 | -3.43 | -2.63 | -3.60 | -4.00 | 22.08 | 43.76 | 151.85 | 375.42 | 36.39 |
| WVGA (832×480) | BasketballDrill | -0.16 | -1.21 | -0.07 | -1.68 | -2.68 | 14.16 | 25.83 | 181.20 | 491.18 | 27.71 |
|            | BQMall    | -0.77 | -4.33 | -1.27 | -4.81 | -6.56 | 20.71 | 66.89 | 182.46 | 499.21 | 58.35 |
|            | PartyScene | -2.62 | -3.01 | -2.73 | -2.98 | -3.13 | 14.66 | 28.22 | 126.52 | 334.19 | 22.79 |
|            | RaceHorses | -0.59 | -1.33 | -0.60 | -1.75 | -2.71 | 06.67 | 19.24 | 111.43 | 304.70 | 33.98 |
| Average on WVGA sequences | -1.03 | -2.47 | -1.17 | -2.81 | -3.77 | 14.05 | 35.04 | 150.40 | 407.32 | 35.71 |
| 1080p (1920×1080) | Kimono    | -1.78 | -5.64 | -2.31 | -6.30 | -9.13 | 13.87 | 56.65 | 149.92 | 434.90 | 83.22 |
|            | BasketballDrive | +1.59 | +0.48 | +1.63 | +0.05 | -6.12 | 3.34 | 13.51 | 136.75 | 403.50 | 75.98 |
|            | Cactus    | -3.28 | -7.21 | -3.81 | -7.44 | -9.31 | 12.78 | 149.44 | 175.76 | 511.17 | 107.17 |
|            | ParkScene | -4.33 | -6.85 | -4.71 | -7.33 | -9.00 | 10.78 | 45.57 | 143.49 | 417.31 | 48.06 |
| Average on 1080p sequences | -1.95 | -4.80 | -2.30 | -5.26 | -8.39 | 15.19 | 66.29 | 151.48 | 441.72 | 78.61 |
| Average on overall sequences | -1.84 | -3.49 | -2.05 | -3.78 | -5.17 | 15.96 | 43.99 | 155.20 | 415.99 | 45.75 |
the proposed method achieves a significant improvement of the B-picture coding efficiency, especially in high resolution sequences. For the BasketballDrive 1080p sequence, only the proposed method exhibits the improvement of coding efficiency. The proposed method reduces the bit rate by 2.52% to 9.31% and 5.17% on average. Figure 5 shows the RD curves for four simulation video sequences. The proposed method outperforms the other methods, especially at the low bit rate. As the video resolution becomes larger, the superior performance of the proposed method is revealed more clearly.

Table 5 shows the selection ratio of each skip mode for the conventional and proposed methods.

![Fig. 5](image)

Table 5  Selection ratios of each skip mode in the conventional and proposed methods.

| Sequence       | $R_{SR}$ (%) | $R_{MSP}$ (%) |
|----------------|--------------|---------------|
|                | Original     | Murakami's    | Mys's         | Proposed    | Murakami's    | Mys's         | Proposed    |
|                | SR ±1        | SR ±4        | SR ±1        | SR ±4        | SR ±1        | SR ±4        | SR ±1       | SR ±4       |
| CIF Soccer     | 42.63        | 43.42        | 44.82        | 43.55        | 44.77        | 45.66        | 4.49        | 8.64        | 4.40        | 8.35        | 11.91       |
| WQVGA RaceHorse| 25.42        | 28.04        | 30.53        | 28.19        | 30.85        | 31.47        | 16.30       | 27.69       | 17.10       | 29.54       | 34.55       |
| WVGA BQMall    | 59.21        | 60.48        | 63.86        | 60.50        | 64.40        | 66.17        | 12.01       | 19.14       | 12.25       | 18.27       | 24.46       |
| 1080p ParkScene| 57.76        | 60.28        | 62.31        | 60.75        | 62.43        | 64.05        | 11.62       | 18.18       | 14.12       | 19.03       | 26.31       |
Sect. 3.3. \( R_{DT} \) and \( R_{MD} \) in Table 5 are computed by

\[
R_{DT}(\%) = \frac{N_D}{N_T} \times 100, \tag{11}
\]

\[
R_{MD}(\%) = \frac{N_M}{N_D} \times 100, \tag{12}
\]

where \( N_D \) and \( N_M \), respectively, denote the numbers of MBs whose modes are DIRECT (\( MCI_{\text{skip \_flag}} = 0 \) or 1) and MCI_SKIP, and \( N_T \) indicates the total number of MBs for each simulation sequence. A higher \( R_{DT} \) value represents the stronger influence of the DIRECT mode. In addition, \( R_{MD} \) specifies the significance of the conventional and proposed methods. The results in Table 5 demonstrate that the proposed method is more frequently selected as the best mode in comparison to the other methods, which leads to the improvement of the overall performance.

### 4.2 Analysis of the Proposed Block Match Criterion

The proposed block match criterion consists of three measurements; symmetric bi-directional block match, template match, and side match errors. To verify the contributions of these match errors, we carried out the simulations using three different combinations of the weighting factors of the proposed block match criterion, as shown in Table 6. In these simulations, the adaptive initial MV derivation is not conducted in order to investigate the effect of the block match criterion only, and the SR of ME is set to \( \pm 16 \) in the quarter-pel unit. Compared with \((w_B, w_T, w_S) = (1, 0, 0)\) (i.e., only the symmetric bi-directional block match is applied), the cases of \((0.5, 0.5, 0)\) show that the use of the template match improves the coding performance. When the side match error is included into the proposed block match criterion (the last column in Table 6), its coding performance is superior to the other cases.

Figure 6 illustrates the subjective performance comparison that helps to understand the contributions of the additional match errors. Figures 6 (a) and (e) show the enlarged regions of the original frames of the RaceHorses and BQMall sequences, respectively, and Figs. 6 (b)–(d) and (f)–(h) show the same regions of the MCI results obtained from employing the reconstructed neighboring frames with \( QP = 27 \). Each MCI is performed using the proposed block match criterion with different weighting factors, as shown in Table 6.

#### Table 6 Comparison of the coding performance (ΔBitrate) for different block match criterions.

| Sequence    | Weighting factor \((w_B, w_T, w_S)\) | \((1,0,0)\) | \((0.5,0.5,0)\) | \((0.5,0.35,0.15)\) |
|-------------|-------------------------------------|-------------|-----------------|-------------------|
| CIF         | Soccer                              | -2.10       | -2.52           | **-2.82**         |
| WQVGA       | RaceHorses                          | -3.93       | -4.42           | **-4.55**         |
| WVGA        | BQMall                              | -5.65       | -7.13           | **-7.21**         |
| 1080p       | ParkScene                           | -7.38       | -8.85           | **-8.92**         |

![Fig. 6](https://example.com/fig6.png)

**Fig. 6** Enlarged regions of the original (first column) and block-based MCI frames (second to fourth columns) with different block match criterions: (a) The 150-th frame of the RaceHorses (WQVGA) sequence, (b) \((w_B, w_T, w_S) = (1, 0, 0)\), (c) \((w_B, w_T, w_S) = (0.5, 0.5, 0)\), (d) \((w_B, w_T, w_S) = (0.5, 0.35, 0.15)\), (e) the 173-th frame of the BQMall (WVGA) sequence, (f) \((w_B, w_T, w_S) = (1, 0, 0)\), (g) \((w_B, w_T, w_S) = (0.5, 0.5, 0)\), (h) \((w_B, w_T, w_S) = (0.5, 0.35, 0.15)\). Red boxed regions indicate the inaccurately-compensated blocks.
In Fig. 6 (b), there exist multiple inaccurately-compensated blocks (red boxed regions). As the template match and side match errors are exploited additionally, it is seen that the visual quality of the MCI results becomes more improved. This improvement yields that more MCI results are selected as the best modes, as a consequence, the improvement of coding performance can be achieved.

### 4.3 Computational Complexity

Table 4 also shows the decoding computational complexity, which is measured by the following equation,

\[ \Delta T_D(\%) = \frac{(T_m - T_a)}{T_a} \times 100, \]

where \( T_m \) and \( T_a \) are, respectively, the elapsed motion compensation time of each method and the original H.264/AVC at the decoder, and \( \Delta T_D \) is the percentage increase of elapsed time for each method. As shown in Table 4, an increase of decoding complexity is unavoidable for all methods. Mys’s method drastically increases the decoding complexity, since full-searching forward ME is required to obtain the initial MV. It is shown that the complexity of the proposed method is similar to that of Murakami’s method with the SR of ±4.

In order to make a fair comparison of the coding performance in terms of decoding complexity, Fig. 7 represents the relationships between \( \Delta \text{Bitrate} \) and \( \Delta T_D \) of each method for the average of overall sequences with various SRs. In Fig. 7, P_ZIM indicates the proposed method without the adaptive initial MV derivation (i.e., the initial MV is set to a zero vector) and P_AIM denotes the proposed method with the adaptive initial MV derivation. It is shown that P_ZIM with the SR of ±16 outperforms the other cases at the expense of decoding complexity. Compared to P_ZIM with the SR of ±16, P_AIM significantly reduces the decoding complexity with a slight reduction of coding performance. Note that P_AIM not only achieves a considerable improvement of coding performance but also has a lower decoding complexity in comparison to Murakami’s method with the SR of ±16. From Table 4 and Fig. 7, it is demonstrated that the proposed method shows the best coding efficiency among conventional methods within a similar complexity range.

### 5. Conclusion

We introduced an improved B-picture coding scheme in this paper. In order to minimize the increase of computational complexity, the initial MV is adaptively derived and the best matching block is searched by performing bi-directional ME within the small SR. In the bi-directional ME, the previously-reconstructed template region is additionally used to find the best prediction signal. The side match error is also utilized to alleviate the block discontinuities. Transmission of both residual signals and motion data is not needed, since both the encoder and decoder have the same reconstructed references. Experimental results indicate that a significant improvement of B-picture coding efficiency can be achieved by using our proposed method as a new coding mode in H.264/AVC.

### Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A2A4A01008384).

### References

1. ITU-T Q.6/SG16 VCEG and ISO/IEC JTC1/SC29/WG11 MPEG, “Terms of reference of the joint collaborative team on video coding standard development,” ISO/IEC JTC1/SC29/WG11/ N11112, Kyoto, Japan, Jan. 2010.
2. ITU-T Q.6/SG16 VCEG and ISO/IEC JTC1/SC29/WG11 MPEG, “Joint call for proposals on video compression technology,” ISO/IEC JTC1/SC29/WG11/N11113, Kyoto, Japan, Jan. 2010.
3. G.J. Sullivan and J.-R. Ohm, “Recent developments in standardization of high efficiency video coding (HEVC),” Proc. SPIE, SPIE Applications of Digital Image Processing XXXIII, vol.7798, no.7798–30, Aug. 2010.
4. M. Flierl and B. Girod, “Generalized B pictures and the draft H.264/AVC video-compression standard,” IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp.587–597, July 2003.
5. A. Puri, R. Aravind, B.G. Haskell, and R. Leonardi, “Video coding with motion-compensated interpolation for CD-ROM applications,” Signal Process. Image Commun., vol.2, no.2, pp.127–144, Aug. 1990.
6. ITU-T Recommendation H.263 Version 2, Video coding for low bit rate communication, ITU-T, 1998.
7. ISO/IEC Standard 14496-2 (MPEG-4 Visual Version 1), Coding of Audio-Visual Objects–Part 2: Visual, ISO/IEC MPEG, 1999.
8. T. Wiegand, G.J. Sullivan, and A. Luthra, “Draft ITU-T recommendation H.264 and final draft international standard 14496-10 advanced video coding,” Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Doc. JVT-G050r1, Geneva, Switzerland, May 2003.
9. B.-T. Choi, S.-H. Lee, and S.-J. Ko, “New frame rate up-conversion using bi-directional motion estimation,” IEEE Trans. Consum. Electron., vol.46, no.3, pp.603–609, Aug. 2000.
10. B.-D. Choi, J.-W. Han, C.-S. Kim, and S.-J. Ko, “Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation,” IEEE Trans.
Seung-Jin Back received the B.S. degree in Electrical Engineering from Korea University, Seoul, in Feb. 2007. He is currently working toward the Ph.D. degree in School of Electrical Engineering at Korea University. His research interests are in the areas of video compression codec, multimedia communication, and video signal processing.

Seung-Won Jung received the B.S. and Ph.D. degrees in electrical engineering in 2005 and 2011, respectively, from Korea University, Seoul, Korea. He is now a research professor in the Research Institute of Information and Communication Technology, Korea University, Seoul, Korea. His research interests include image enhancement, image restoration, video compression, and computer vision.

Habyun Lee received the BS in electronics engineering from Korea Aerospace University, Korea, in 2002, and the MS in mobile communication and digital broadcasting engineering from the University of Science and Technology (UST), Daejeon, Korea, in 2007. From 2003 to 2005, he was a member of engineering staff in LG Electronics. Since 2008, he has been a member of engineering staff in Broadcasting and Telecommunications Media Research Department of ETRI, Daejeon, Korea. His research interests include video coding, image processing, and video transmission.

Hui Yong Kim received his BS, MS, and Ph.D degrees from Korea Advanced Institute of Science and Technology (KAIST) in 1994, 1998, and 2004, respectively. From 2003 to 2005, he was the Multimedia Team leader of AddPac Technology Co. Ltd. In November 2005, he joined the Broadcasting Media Research Group of Electronics and Telecommunications Research Institute (ETRI) as a senior member of research staff. His research interests include video signal processing, multimedia systems, and interactive media.

Sung-Jea Ko received the Ph.D. degree in 1988 and the M.S. degree in 1986, both in Electrical and Computer Engineering, from State University of New York at Buffalo, and the B.S. degree in Electronic Engineering at Korea University in 1980. In 1992, he joined the Department of Electronic Engineering at Korea University where he is currently a Professor. From 1988 to 1992, he was an Assistant Professor of the Department of Electrical and Computer Engineering at the University of Michigan-Dearborn. He has published over 160 international journal articles. He also holds over 60 patents registered on video signal processing, computer vision, and multimedia communications. He is the 1999 Recipient of the LG Research Award given to the Outstanding Information Technology and Communication Researcher. He received the Hae-Dong best paper award from the Institute of Electronics Engineers of Korea (IEEEK) (1997) and the best paper award from the IEEE Asia Pacific Conference on Circuits and Systems (1996), and the research excellence award from Korea University (2004). He has served as a TPC member for the IEEE Conference on Consumer Electronics (ICCE) since 1997 and received a 10-year service award from the TPC of ICCE in 2008. He is a member of the editorial board of the IEEE Transactions on Consumer Electronics. He is the President-Elect of the IEEK and an IEEE Fellow.