J Real-Time Image Proc
DOI 10.1007/s11554-013-0339-0

ORIGINAL RESEARCH PAPER

An adaptive motion compensation method using superimposed inter-frame signals

L.-M. Po · K.-H. Ng · X.-Y. Xu · K.-W. Cheung · L.-T. Feng

Received: 6 November 2012 / Accepted: 14 March 2013

Abstract In theory, multi-hypothesis motion compensation prediction (MHMCP) can enhance the prediction quality of motion compensation prediction. Traditional MHMCP methods use fixed weightings for the linear combination of the multiple signal sources which may not be optimum. Moreover, MHMCP requires the transmission of more than one motion vector, which will increase the side information to be transmitted. We discovered that using estimated distortion ratio, a weighting pair can be estimated adaptively for the linear combination of two signal blocks to form a prediction block with a lower distortion. The proposed MCP method does not require the transmission of additional side information yet has a better prediction accuracy than conventional motion compensation prediction. In addition, the proposed method has relatively low algorithmic decision overhead. It can be implemented in hardware easily to support the realization of real time high-quality video coding.

Keywords Video coding · Real time · Motion compensation prediction · Multi-hypothesis · Motion estimation

1 Introduction

By the Multi-Hypothesis MCP (MHMCP) theory, arbitrary number of prediction signals can be linearly combined to improve the performance of MCP [1, 2]. Bi-directional prediction for B-frames is one of the applications of MHMCP in which two prediction signals, one from the reference frame before and the other from the reference frame after the B-frame, are superimposed to form a prediction signal with better prediction quality. MHMCP requires the estimation of multiple motion vectors. The best prediction performance can be obtained when all the motion vectors are jointly estimated but this requires very high computation complexity. Suboptimal solutions can speed up the process [3]. In [4], it is reported that MHMCP can work together with variable block size MCP and multiple-reference frame MCP to enhance the efficiency of a rate-constrained coding scheme. Experimental results in [4] show that two jointly estimated prediction signals can achieve up to 30 % bit-rate reduction in coding. In [5], two-hypothesis MCP is used to boost the error resiliency in an error-prone environment. In [6], three-hypothesis MCP is used to reduce the error propagation as well as achieving rate-distortion gain.

In [4], the weighting coefficients applied to the multiple signal sources converge to $1/n$, where n is the number of signal sources. That means an averaging is applied to the multiple signal sources. The same averaging is also used for the linear combination of the multiple signals found by template matching in [7]. In both [6] and [8], fixed optimum weighting combinations for the multiple signal sources are found by empirical methods. However, we are not sure whether fixed weightings can obtain optimum prediction performance for MHMCP because the characteristics and conditions of the signal sources can differ a lot. The second problem of MHMCP is that it requires the transmission of more than one motion vector. As the residual data of MCP are nowadays getting smaller and smaller due to the advance in MCP technology, the increase in the number of motion vectors is very unfavorable. Moreover, most MHMCP methods proposed have very high computational complexity, which is unfavorable, especially for real time video coding applications.

L.-M. Po · K.-H. Ng (✉) · X.-Y. Xu · K.-W. Cheung · L.-T. Feng
Hong Kong, China
e-mail: kahomike@gmail.com

Published online: 05 April 2013

From the analysis of the linear combination of two inter-frame signal blocks, we found that the optimum weighting pair is correlated to the distortion ratio of the blocks. We also found that correlation exists for the estimated version of the distortion ratio too. Using a method similar to template matching, we can find the base signal block without the transmission of the respective motion vector. Combining these two methods, we propose an adaptive superimposed inter-frame search (ASIS) algorithm which has better prediction performance than conventional motion compensation prediction while eliminating the drawbacks of MHMCP.

2 Analysis of superimposition of inter-frame signals

In natural video sequences, local minima exist in the distortion error surface of block-matching motion estimation (BMME). This means that several prediction blocks can be found in the reference frame with similar distortion values. In other words, these prediction blocks with different MVs may all resemble the current block, with pixel differences here or there over the block. In conventional BMME, only one prediction block with the lowest distortion value is selected. Although this prediction block with the lowest distortion value can roughly represent the displacement of the current block in the previous frame, the other prediction blocks can also tell us how the current block should look like. Therefore, we can refine the prediction signal block with the lowest distortion value by superimpose it with another prediction signal block, with different weightings applying to these different signal sources.

Consider an example shown in Figs. 1 and 2. Figure 1 is the current frame and the current block to be encoded is highlighted in the red dotted box. Figure 2 is the previous frame, which is also the reference frame in which the current block will find the best match using BMME. The object on the upper-right corner of both frames is an illuminating object, which for example can be the sun. The triangle and the rectangle are moving towards each other, with the luminous of the rectangle becoming dimmer due to its increased distance from the illuminating object. The pixel luminous values are from 0 to 255, with 255 being the brightest and 0 being the darkest. Figure 3 shows the pixel luminous values of the reference frame.

If conventional block-matching motion estimation is used, the best match is the one with the lowest distortion value. In this example, the best match is the one indicated by the yellow dotted box in the reference frame in Fig. 3. Using sum of absolute differences (SAD) as the distortion measurement, the SAD of this best-matched block is 4,590. The pixel differences are shown in Fig. 4.

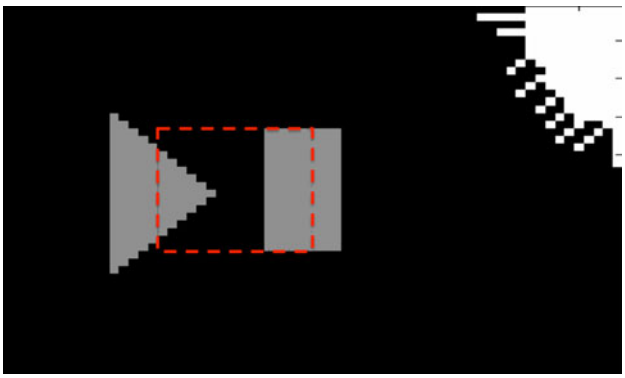

Fig. 1 Current frame and current block

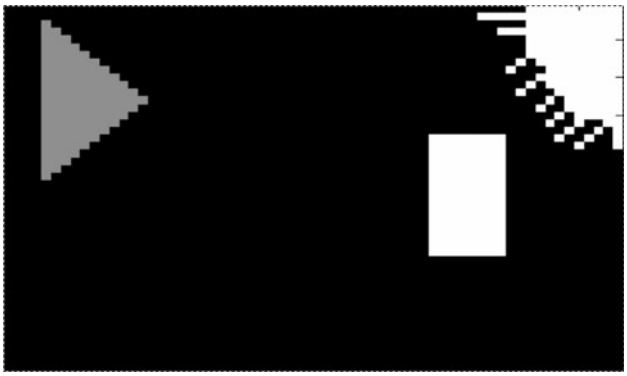

Fig. 2 Previous reference frame

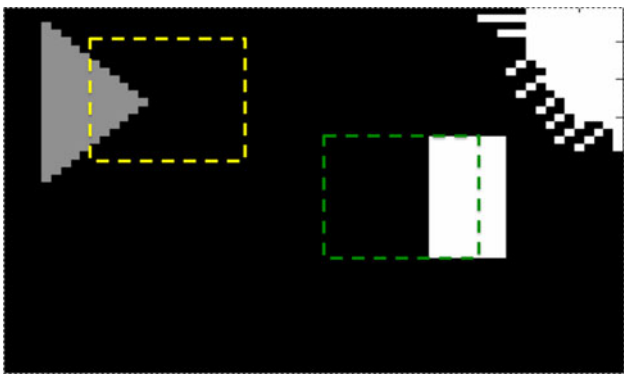

Fig. 3 Two block-matchings on the reference frame

Consider another matching indicated by the green dotted box in Fig. 3, the SAD is 10,200 as shown in Fig. 5. In conventional BMME, this matching is not used because it is not the lowest distortion matching. However, if we superimpose the best-matched block with this block linearly, with the pixel values of each of the block being halved, we can obtain a superimposed block with a SAD value of 2,295. This superimposed block, thus, has lower distortion value than the best-matched block found by conventional motion compensated prediction. Figure 6 shows this superimposed block and Fig. 7 shows the pixel

Fig. 4 Pixel differences between the matching indicated by the *yellow dotted box* in Fig. 3 and the current block

0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	127.5	127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	127.5	127.5	127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	127.5	127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
127.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 5 Pixel differences between the matching indicated by the *green dotted box* in Fig. 3 and the current block

0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.5	127.5	127.5	127.5	127.5

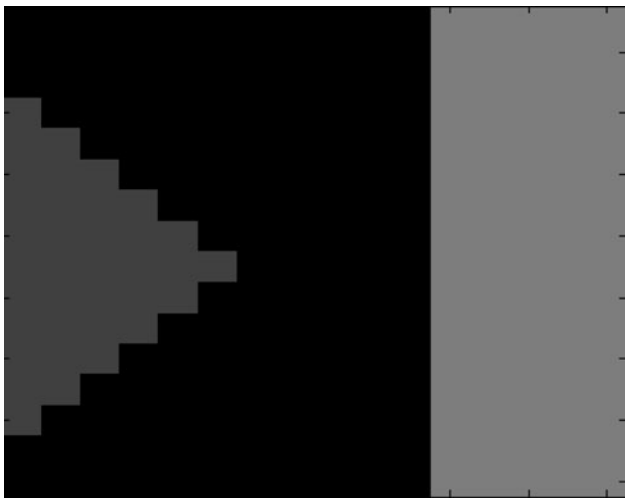

Fig. 6 The superimposed block

differences between this superimposed block and the current block.

We perform experiments to analyze the prediction quality of superimposition of inter-frame signals. First, we need to define the conventional block-matching motion estimation process. Consider the nth image frame $\mathbf{F}_n$ of size $W \times H$ of a video sequence with pixel value $F_n(s)$ for each pixel location $\boldsymbol{s} = (x, y)$, in which x and y are integers such that $0 \leq x < W$ and $0 \leq y < H$. The pixel value of the reconstructed reference frame is denoted as $\tilde{F}_n(s)$. The nth image frame is segmented into K non-overlapping blocks, $\{B_{i,n}\}_{i=1}^{K}$. In single reference frame conventional block-matching motion estimation, each block is predicted from the previous reconstructed frame $\tilde{F}_n$. The motion vector (MV), $v_{i,n} = (u, v)_{i,n}$, specifying a spatial displacement for motion compensation of ith block in $\mathbf{F}_n$, is determined by block-matching process as

$$v_{i,n} = \arg\min_{d} \mathrm{BDM}_{B_{i,n}}\{F_n(s), \tilde{F}_{n-1}(s-d)\}, \tag{1}$$

where BDM (block distortion measure) calculates the difference or distortion between two blocks and the 2D displacement vector $\mathbf{d}$ is limited to have finite vector component within a search area. A commonly used BDM is SAD (sum of absolute difference), which is defined as

$$\mathrm{SAD}_B\{F(s), G(s)\} = \sum_{s \in B} |F(s) - G(s)| \tag{2}$$

If all the candidate positions in the search area are searched, it is the exhaustive search. Exhaustive search guarantees the finding of the lowest distortion candidate block in the search area. To analyze the prediction quality of superimposition of inter-frame signals, we superimpose the lowest distortion candidate block found by conventional block-matching ME with another candidate block in the same reference frame. Therefore, in our new algorithm, each current block first performs a conventional exhaustive search using Eq. 1. The first motion vector $v_{i,n}$ is determined and the candidate block pointed by this motion vector is called the base signal block or base block. Then, a second exhaustive search is performed, in which each candidate block is superimposed with the base block and the distortions between the current block and the superimposed blocks are calculated. The second motion vector $r_{i,n} = (e,f)_{i,n}$, which points to the candidate signal block forming the lowest distortion superimposed block, is determined by

$$r_{i,n} = \arg\min_{d} \mathrm{BDM}_{B_{i,n}}\{F_n(s), (\tilde{F}_{n-1}(s-v)) \cdot w_{\text{base}} + (\tilde{F}_{n-1}(s-d) \cdot w_{\text{candidate}}\} \tag{3}$$

where w_{base} is the weighting applied to the lowest distortion block (base block) found in the first exhaustive search, $w_{\text{candidate}}$ is the weighting applied to the candidate signal blocks in the second exhaustive search, and $w_{\text{base}} + w_{\text{candidate}} = 1$.

The weightings w_{base} and $w_{\text{candidate}}$ can be regarded as the signal strengths given to the base signal block and the candidate block, respectively. For example, if $w_{\text{base}} = 0.7$ and $w_{\text{candidate}} = 0.3$, we can interpret the superimposed signal block as a linear combination of the lowest distortion block with signal strength of 70 % and a candidate signal block with signal strength of 30 %. Since we do not know yet which weighting pair can achieve the best prediction quality, we perform experiment using the following Algorithm 1:

Step 1 Perform block matching between the current block and the candidate blocks in the search window based on Eq. 1. The candidate block with the lowest distortion and its MV are found. They are the base block and the base MV, respectively

Step 2 Based on Eq. 3, perform block matching between A) the current block and B) the weighted combinations of candidate blocks in the search window and the base block found in Step 1. The superimposed block with the lowest distortion together with the weighting used and the MV are found

Experiment is performed using 81 sets of weightings w_{base} and $w_{\text{candidate}}$:

w_{base}	$w_{\text{candidate}}$
0.0000	1.0000
0.0125	0.9875
0.0250	0.9750
0.0375	0.9625

Table a continued

w_{base}	$w_{candidate}$
0.0500	0.9500
0.0625	0.9375
0.0750	0.9250
0.0875	0.9125
0.1000	0.9000
0.1125	0.8875
.	.
.	.
.	.
0.9000	0.1000
0.9125	0.0875
0.9250	0.0750
0.9375	0.0625
0.9500	0.0500
0.9625	0.0375
0.9750	0.0250
0.9875	0.0125
1.0000	0.0000

The difference between each weighting pair is 0.0125. At each candidate position, 81 block-matchings are performed between the current block and the candidate block, using the 81 weighting pairs. That means at each search position, different signal strengths are applied to the candidate block signal and the base block signal. We want to find out the optimum signal strength balance between the two sources.

Table 1 shows the average SAD per block and the average PSNR per frame achieved using the above algorithm compared with that achieved using conventional motion prediction. The sequences are in size CIF (352×288) and 4CIF (704×576). 100 and 50 frames are used for the CIF and 4CIF sequences, respectively. The block size is 16×16 pixels. The search window size is ± 16 pixels. Fractional-pixel motion estimation is implemented with quarter-pixel (1/4-pixel) accuracy. Different from a real video codec that uses reconstructed reference frames, in these experiments, original frames are used as reference frames. Bit-rate is not available in these simulation experiments.

We can observe that two inter-frame prediction signals can be linearly combined to form a signal with lower distortion and thus increase the overall prediction quality. However, we have to solve two problems in real coding. The first is that the decoder side needs to know which is the best weighting pair being used for the superimposition. Second, the decoder needs to know where to find the two inter-frame prediction signals. That means two MVs are

Fig. 7 Pixel differences between the superimposed block (Fig. 6) and the current block

0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	63.8	63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	63.8	63.8	63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	63.8	63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1 Average SAD per block and average PSNR per frame of Algorithm 1

	Akiyo (CIF)	Coastguard (CIF)	Foreman (CIF)	Mobile (CIF)	Mother (OF)
Algorithm 1					
Average SAD per block	113.72	1,029.91	533.00	1,186.51	291.46
Exhaustive search					
Average SAD per block	132.43	1,140.81	602.03	1,329.97	334.22
Algorithm 1					
Average PSNR (dB) per frame	46.16	31.58	35.71	28.11	42.69
Exhaustive search					
Average PSNR (dB) per frame	44.93	30.88	34.79	27.52	41.55
PSNR increment	1.23	0.70	0.92	0.58	1.14
	Sean (CIF)	Stefan (CIF)	City (4CIF)	Crew (4CIF)	Harbor (4CIF)
Algorithm 1					
Average SAD per block	332.66	1,082.08	844.20	865.69	1,005.51
Exhaustive search					
Average SAD per block	370.11	1,197.33	938.85	1,017.22	1,124.01
Algorithm 1					
Average PSNR (dB) per frame	42.32	28.21	33.66	35.57	32.49
Exhaustive search					
Average PSNR (dB) per frame	41.13	27.44	32.69	34.50	31.54
PSNR increment	1.19	0.77	0.97	1.07	0.95

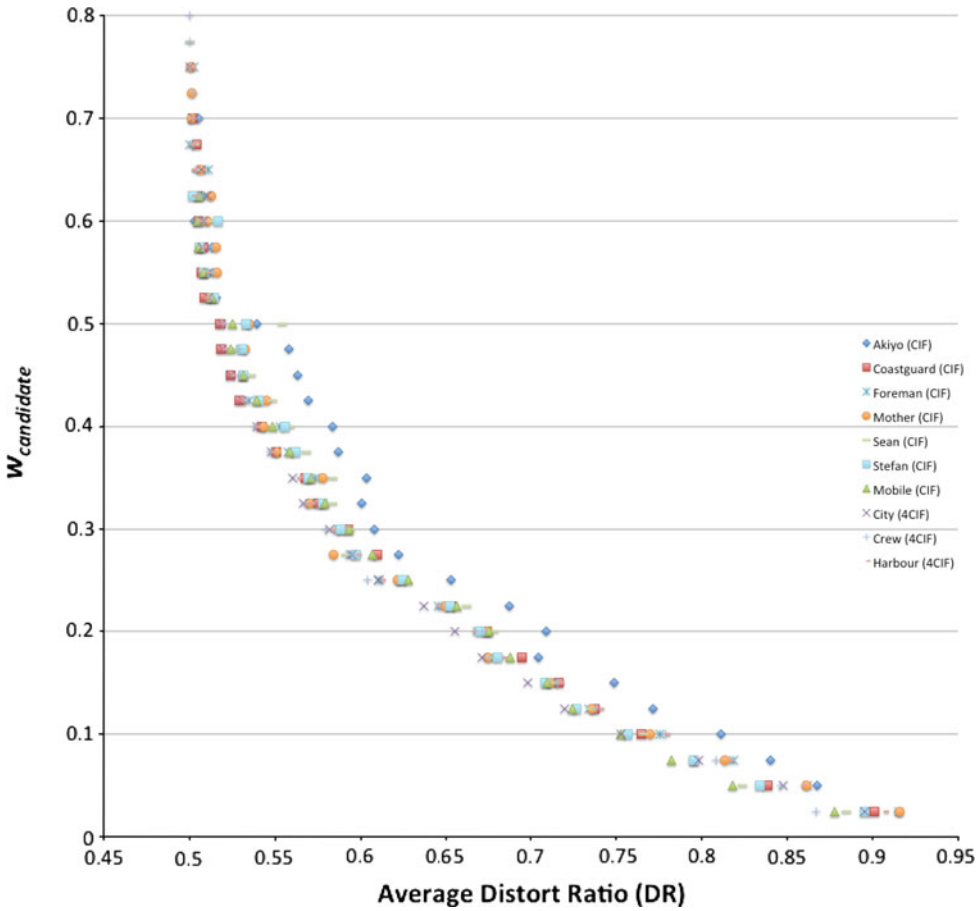

Fig. 8 Best $w_{candidate}$ selected versus average distortion ratio (DR) using Algorithm 1

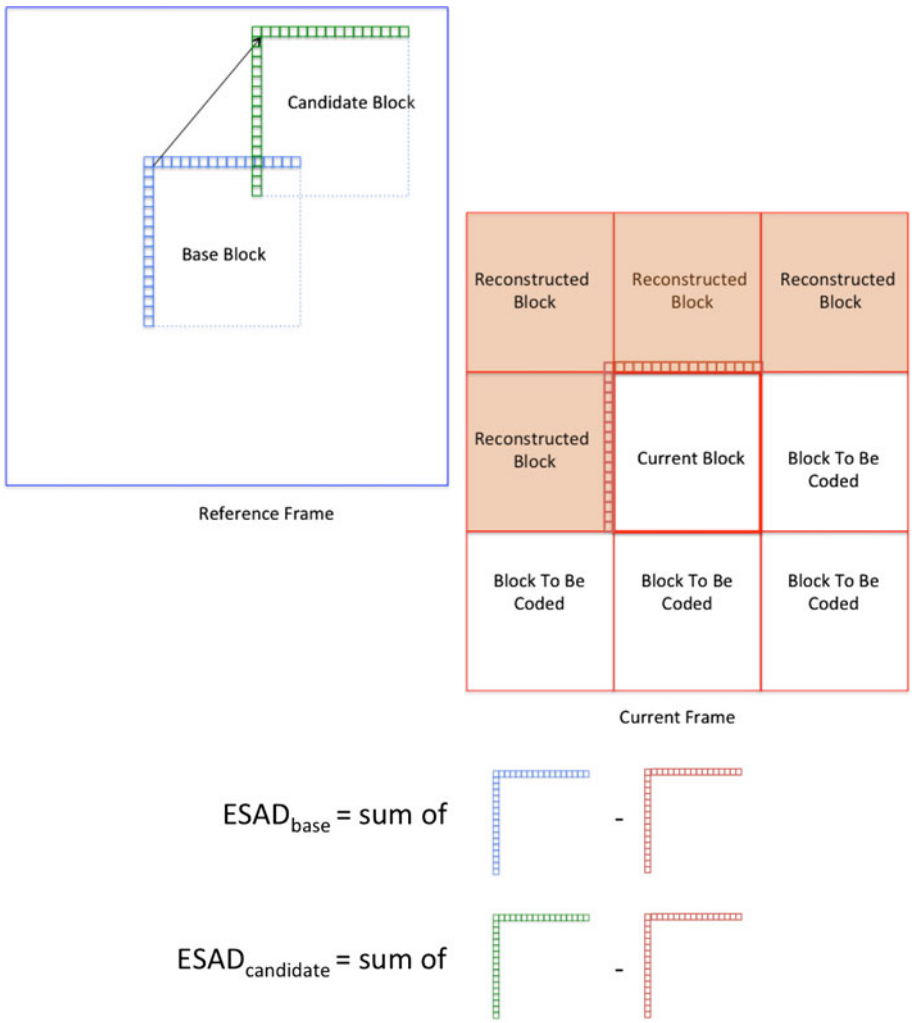

Fig. 9 Calculation of estimated sum of differences (ESAD)

required to be transmitted. Moreover, we need to tell the decoder the value of the optimal weighting pair being selected. This overhead information requires extra bits to be transmitted, thus affecting the overall bit-rate performance. Although there might be a chance that the better prediction gain can compensate the overhead bit requirement, we take another approach to address these two problems. We want to solve them without extra overhead information being sent. Then, we can obtain pure prediction gain.

First, we address the problem of instructing the decoder side the best weighting pair being used. We found that the best weighting pair selected has high correlation with both the distortion ratio (DR) and the estimated distortion ratio (EDR) of the base block and the candidate blocks. This will be discussed in Sect. 3. Second, we address the problem of using more than one MV. By finding the block with the lowest ESAD, we can obtain an approximation of the base block without exhaustive search and so eliminate the use of more than one MV. This will be discussed in Sect. 4.

3 Distortion ratio

We hypothesize that the optimum weighting pair is related to the distortions of the base blocks and the candidate blocks. To find the relationship of the distortions of the base blocks, we define:

$$\mathrm{DR} = \frac{\mathrm{SAD}_{\mathrm{candidate}}}{\mathrm{SAD}_{\mathrm{base}} + \mathrm{SAD}_{\mathrm{candidate}}} \tag{4}$$

where DR is the distortion ratio, $\mathrm{SAD}_{\mathrm{candidate}}$ is the sum of absolute differences (SAD) between a candidate block and the current block, $\mathrm{SAD}_{\mathrm{base}}$ is the sum of absolute differences (SAD) between a base block and the current block. Algorithm 1 stated in the last section is repeated. For each superimposed block with the lowest distortion, the distortion ratio is calculated and the optimum weighting pair w_{base} and $w_{\mathrm{candidate}}$ selection is recorded. The average value of the distortion ratios of each weighting pair is shown in Fig. 8 ($w_{\mathrm{candidate}}$ is to represent the weighting pairs as $w_{\mathrm{base}} = 1 - w_{\mathrm{candidate}}$).

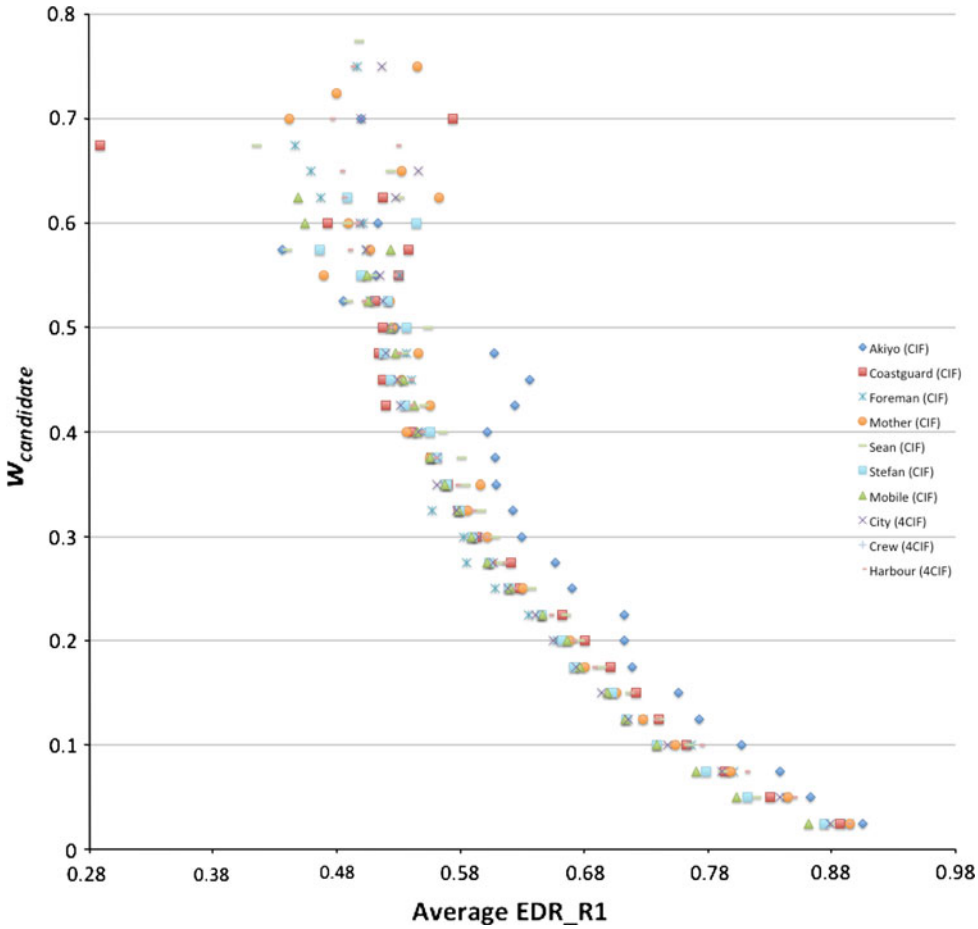

Fig. 10 Best $w_{candidate}$ selected versus average estimated distortion ratio (EDR) using Algorithm 1

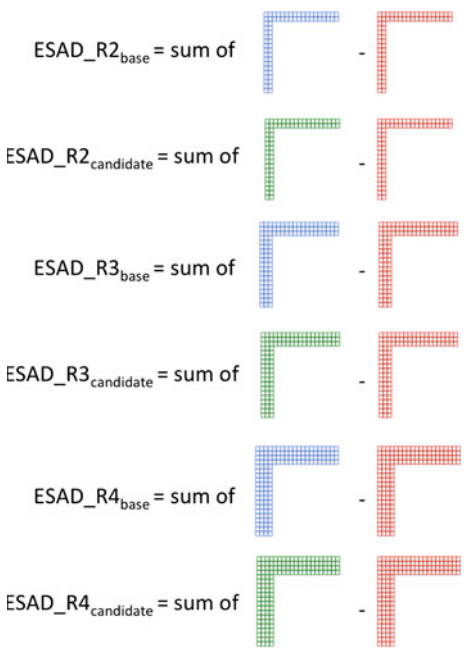

Fig. 11 ESAD with different R numbers

We can see that there is an inverse relationship between the best $w_{candidate}$ selected and the average distortion ratio. The higher the $w_{candidate}$ selected, the lower the average distortion ratio, and vice versa. If, for example, the distortion of a base block is high and that of a candidate block is low, it means the semblance of the base block with the current block is lower than the semblance of the candidate block with the current block. In that case, applying a higher weighting to the candidate block and applying a lower weighting to the base block in the superimposition can form a lower distortion prediction block. In short, the best weightings that can form a better prediction block can be deduced from the distortions of the base signal block and a candidate block pointed by a MV.

However, the decoder side does not know the $SAD_{candidate}$ (SAD between a candidate block and the current block) and SAD_{base} (SAD between a candidate block and the base block) because current blocks only exist on the encoder side but not on the decoder side. We hypothesize that the neighboring pixels of the current block have similar motion and structure as the current block. Using the neighboring pixels of the current block, we try to estimate $SAD_{candidate}$ and SAD_{base}. The estimated $SAD_{candidate}$ and

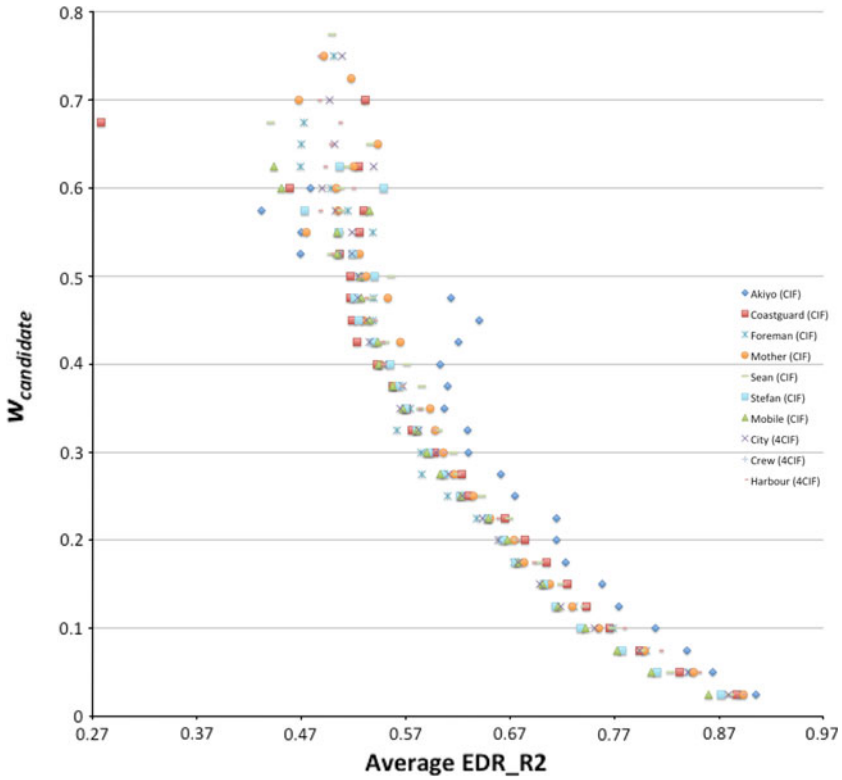

Fig. 12 Best $w_{candidate}$ selected versus EDR_R2 using Algorithm 1

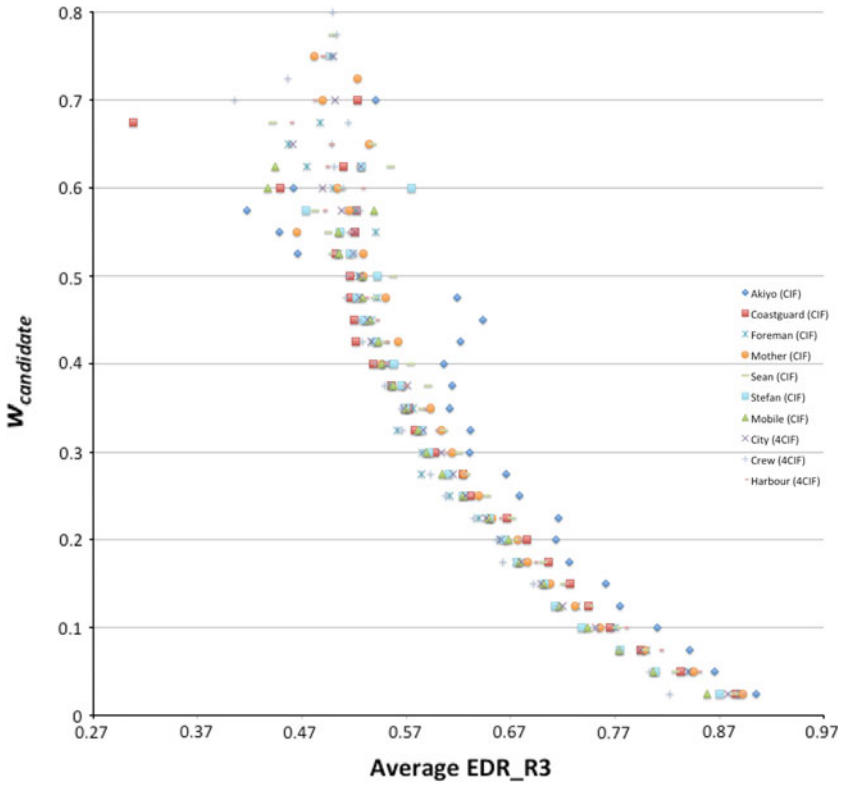

Fig. 13 Best $w_{candidate}$ selected versus ERD_R3 using Algorithm 1

estimated SAD_{base} are $ESAD_{candidate}$ and $ESAD_{base}$, respectively. Figure 9 illustrates how $ESAD_{candidate}$ and $ESAD_{base}$ are calculated.

The concept of estimated sum of differences (ESAD) calculation is a bit similar to template matching (TM) [6, 7]. In TM, the neighboring pixels of the current block are

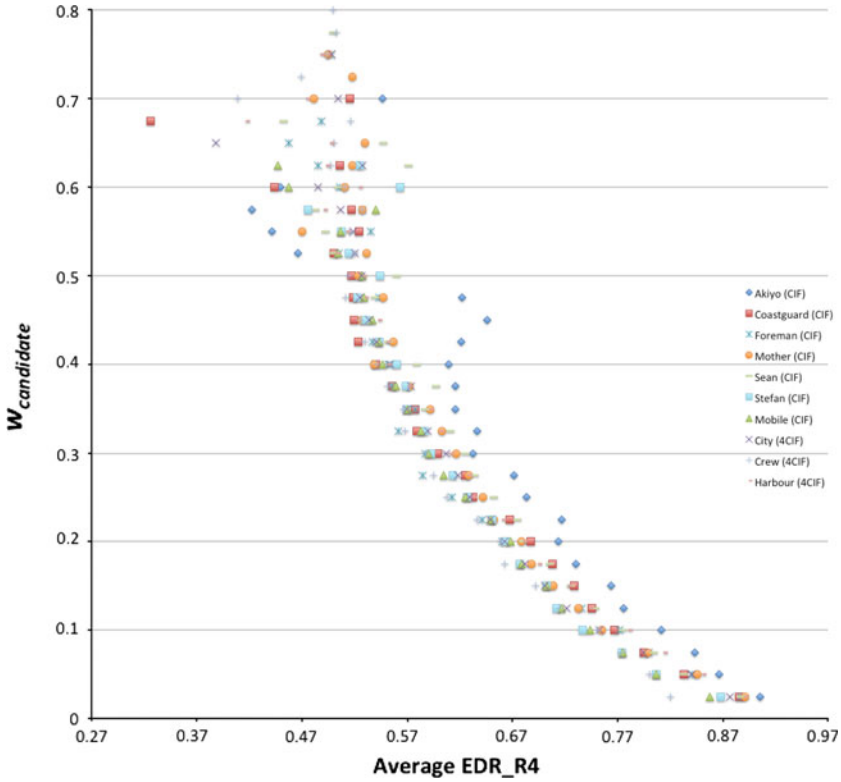

Fig. 14 Best $w_{\text{candidate}}$ selected versus EDR_R4 using Algorithm 1

used as a template to find a good prediction for the current block. Both ESAD and TM exploit the spatial correlation between the current block and its neighboring pixels. In Sect. 4, we will also use template matching to find the base block, instead of using full search as in Algorithm 1. We will discuss TM in more details in Sect. 4. Here, we first concentrate on the use of ESAD to estimate the distortion ratio (DR).

We calculate the EDR using $\text{ESAD}_{\text{candidate}}$ and $\text{ESAD}_{\text{base}}$ using Eq. 5:

$$\text{EDR} = \frac{\text{ESAD}_{\text{candidate}}}{\text{ESAD}_{\text{base}} + \text{ESAD}_{\text{candidate}}} \tag{5}$$

Algorithm 1 is repeated, with the average EDR recorded instead of the original distortion ratio. Figure 10 plots the best $w_{\text{candidate}}$ selected versus average EDR. It shows that the best weighting pair also has an inverse relationship with the average EDR similar to that inverse relationship with the distortion ratio. $\text{ESAD}_{\text{candidate}}$ and $\text{ESAD}_{\text{base}}$ can also be calculated using more neighboring pixels to the current block. Figure 11 shows that the number of neighboring pixels used in ESAD calculation can be increased. The fixed inverse-L shaped area increases with the number of rows of pixels used. If the current block size is $N + N$, the number of rows of pixels used in ESAD calculation is R, then the total number of pixels in ESAD calculation will be $R \times (N + N + R)$. For example, if $N = 16$, $R = 2$, the total number of pixels used is equal to $2 \times (16 + 16 + 2) = 68$.

For simplicity, ESAD calculated with $R = 2$ is named as ESAD_R2, ESAD calculated with $R = 3$ is named as ESAD_R3, ESAD calculated with $R = 4$ is named as ESAD_R4, and so on. ESAD_R2, ESAD_R3, and ESAD_R4 are shown in Fig. 11. They are calculated, respectively, and their plots against the best weighting are shown in Figs. 12, 13 and 14. From observation, the results using ESAD_R2, ESAD_R3, and ESAD_R4 are similar. In the coming sections, we will determine the optimum R number (that is the optimum number of pixels used in ESAD calculation) by analyzing their variances.

4 Base block calculation using lowest ESAD

From the analysis in the last section, we can see that prediction quality can be improved by superimposing two inter-frame prediction signals, and the best weighting pair selected has a relationship with the estimate distortion ratio (EDR). In the experiments with Algorithm 1 in the last section, the base signal block is found using conventional exhaustive search. A MV indicating the displacement of this base signal block in the reference frame is needed in the decoder side. Including the MV for the displacement of the candidate block, in total two MVs are required to be transmitted. To eliminate the use of two MVs, we select

Table 2 Average Base Block SAD using different numbers of pixels in $ESAD_{candidate}$ calculation

Num of rows (R)	Num of pixels in ESAD calculation	Akiyo (CIF)	Coastguard (CIF)	Foreman (CIF)	Mobile (CIF)	Mother (CIF)
1	33	170.01	1,314.20	791.52	1,597.16	480.96
2	68	165.40	1,250.30	771.94	1,480.15	459.76
3	105	**165.08**	1,239.33	**770.16**	1,453.42	456.09
4	144	165.48	1,236.54	773.58	1,436.14	449.34
5	185	166.86	**1,234.90**	777.81	1,423.74	445.04
6	228	167.71	1,236.20	783.34	1,421.05	442.26
7	273	168.26	1,236.05	789.18	1,418.94	441.85
8	320	168.69	1,239.20	795.27	1,418.56	442.43
9	369	168.97	1,242.76	804.84	**1,418.38**	441.64
10	420	169.13	1,246.92	812.46	1,419.44	**441.49**
Num of rows (R)	Num of pixels in ESAD calculation	Sean (CIF)	Stefan (CIF)	City (4CIF)	Crew (4CIF)	Harbor (4CIF)
1	33	484.98	1,327.70	1,153.48	1,304.74	1,403.67
2	68	465.27	1,276.80	1,083.99	1,262.12	1,328.98
3	105	456.36	1,264.19	1,064.11	1,241.63	1,303.64
4	144	449.98	**1,264.02**	1,055.07	1,231.50	1,290.20
5	185	449.87	1,265.87	**1,053.07**	1,224.85	1,282.83
6	228	447.75	1,274.64	1,053.19	1,220.82	1,277.68
7	273	446.93	1,286.64	1,055.30	1,218.90	1,276.90
8	320	**446.32**	1,292.87	1,058.49	**1,217.16**	1,274.52
9	369	446.70	1,303.05	1,063.96	1,217.57	**1,273.20**
10	420	446.74	1,318.06	1,069.42	1,217.58	1,273.89

Bold means the lowest value in that column

Table 3 Average SAD per block and average PSNR per frame of Algorithm 2

	Akiyo (CIF)	Coastguard (CIF)	Foreman (CIF)	Mobile (CIF)	Mother (CIF)
Algorithm 2					
Average SAD per block	114.37	1,045.27	546.68	1,197.13	298.86
Exhaustive search					
Average SAD per block	132.43	1,140.81	602.03	1,329.97	334.22
Algorithm 2					
Average PSNR (dB) per frame	46.06	31.47	35.43	28.04	42.49
Exhaustive search					
Average PSNR (dB) per frame	44.93	30.88	34.79	27.52	41.55
PSNR increment	1.13	0.59	0.64	0.52	0.94
	Sean (CIF)	Stefan (CIF)	City (4CIF)	Crew (4CIF)	Harbor (4CIF)
Algorithm 2					
Average SAD per block	335.73	1,097.86	855.80	896.92	1,023.51
Exhaustive search					
Average SAD per block	370.11	1,197.33	938.85	1,017.22	1,124.01
Algorithm 2					
Average PSNR (dB) per frame	42.22	28.08	33.47	35.34	32.27
Exhaustive search					
Average PSNR (dB) per frame	41.13	27.44	32.69	34.50	31.54
PSNR increment	1.09	0.64	0.78	0.84	0.73

another base block which has a displacement known by the decoder side.

Inter-frame prediction using template matching (TM) has been studied in many research works, for example [7, 9–11]. A group of neighboring pixels of the current block is used as a template. Matchings between this template and the templates of the candidate blocks in the reference frame are performed. Certain matching criteria define the candidate block(s) suitable to be used for prediction. TM has the advantage that no additional MV or side information is

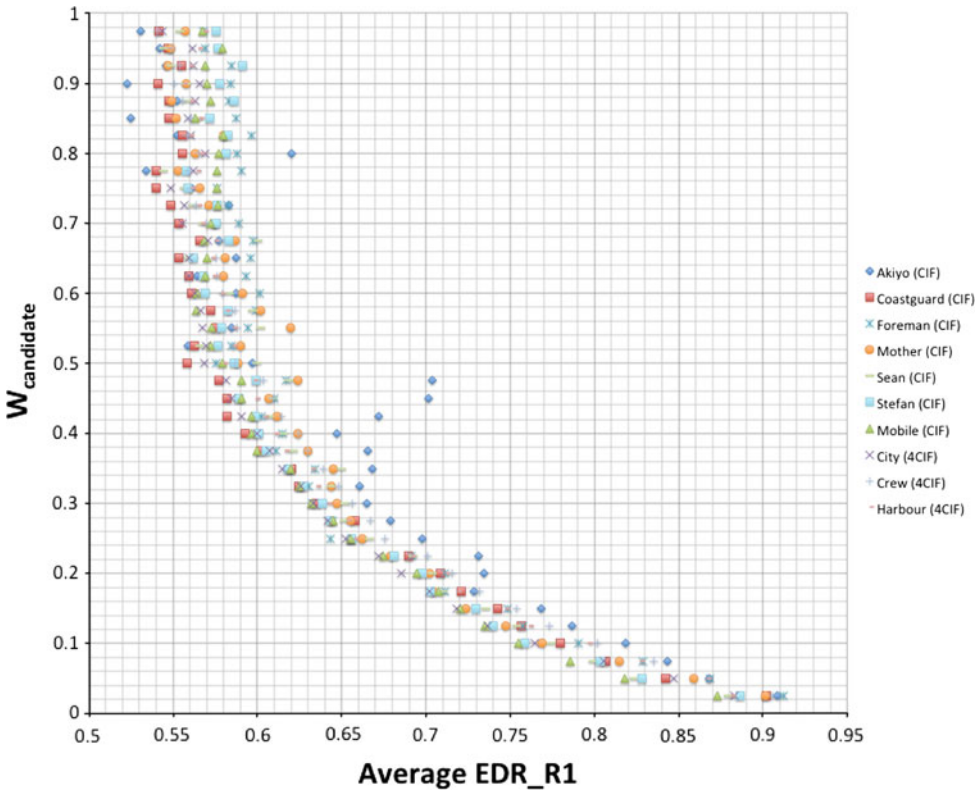

Fig. 15 Best $w_{\text{candidate}}$ selected versus average estimated distortion ratio EDR_R1 using Algorithm 2

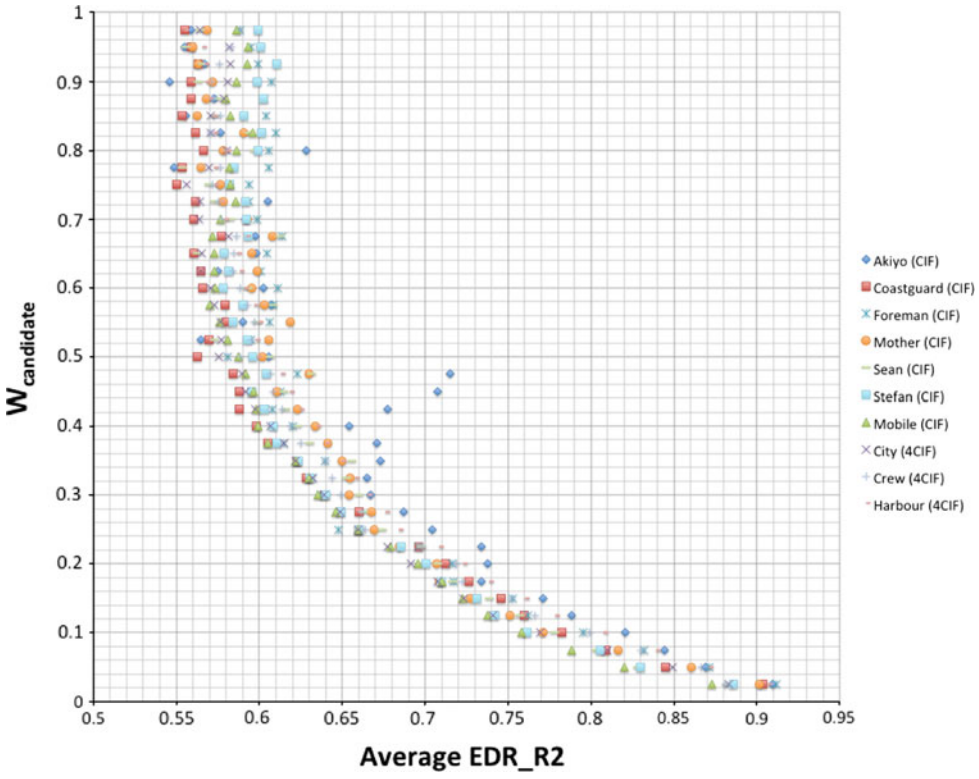

Fig. 16 Best $w_{\text{candidate}}$ selected versus average estimated distortion ratio EDR_R2 using Algorithm 2

needed to be sent to the decoder side. The disadvantage is that the prediction quality of the initial prediction block found by TM is not as good as that of the prediction block pointed by a MV found by full search or other search algorithms. Moreover, the computational complexity of the decoder side will also be increased. One of the methods of

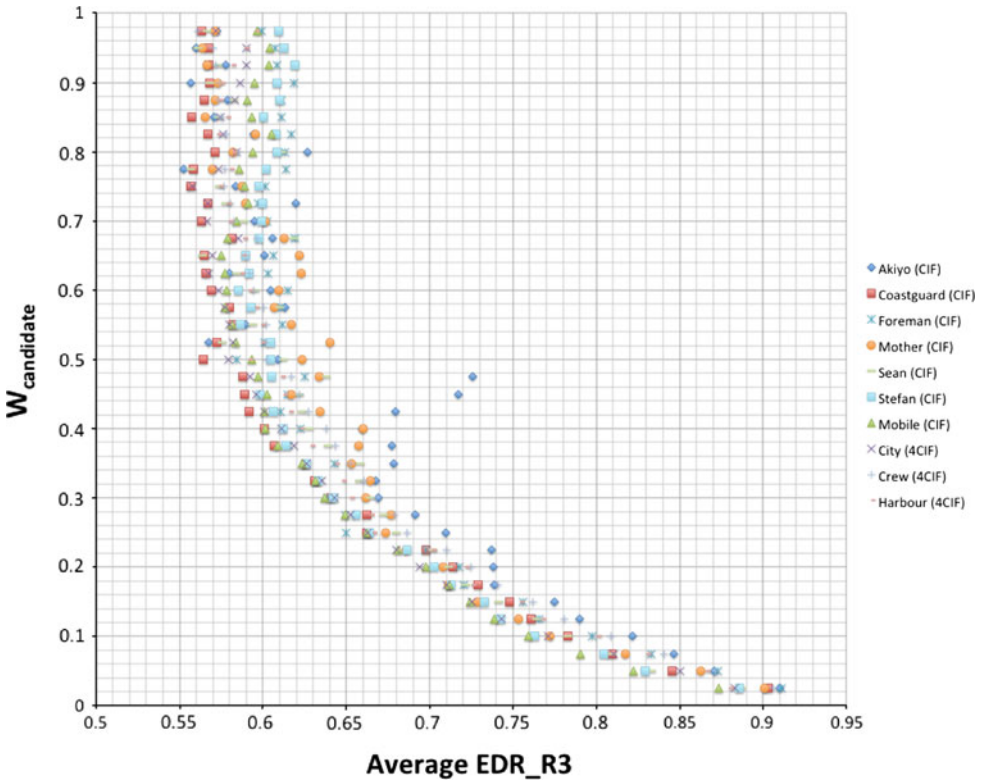

Fig. 17 Best $w_{\text{candidate}}$ selected versus average estimated distortion ratio EDR_R3 using Algorithm 2

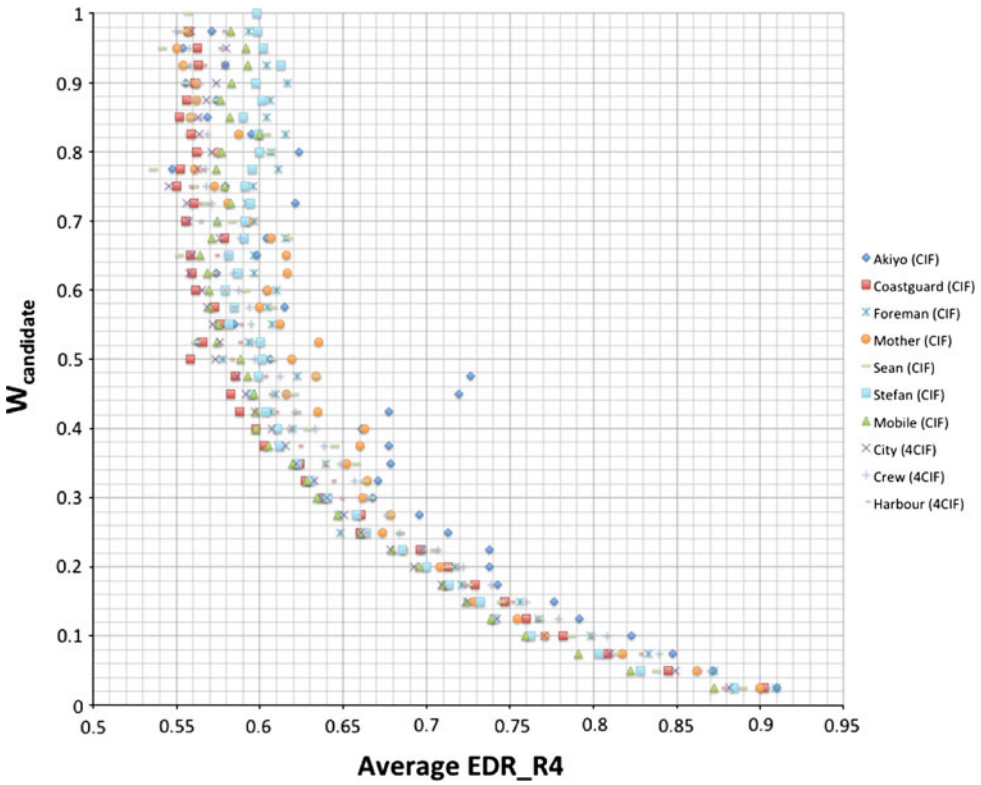

Fig. 18 Best $w_{\text{candidate}}$ selected versus average estimated distortion ratio EDR_R4 using Algorithm 2

performing TM is to find the lowest SAD value between the template of the current block and the templates of the candidate blocks. It is the same as finding the candidate block with the lowest $\text{ESAD}_{\text{candidate}}$. Because we assume that ESAD is a close approximation to the real SAD, we replace the first exhaustive search step in Algorithm 1 with

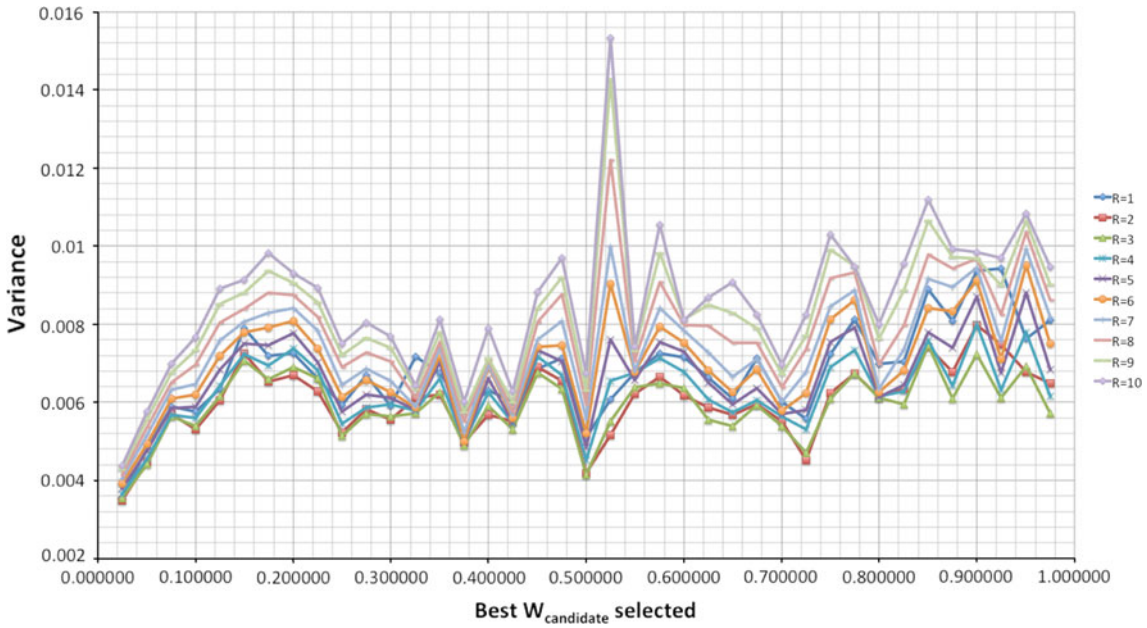

Fig. 19 Variance of EDR values using different numbers of rows of pixels in *Foreman*

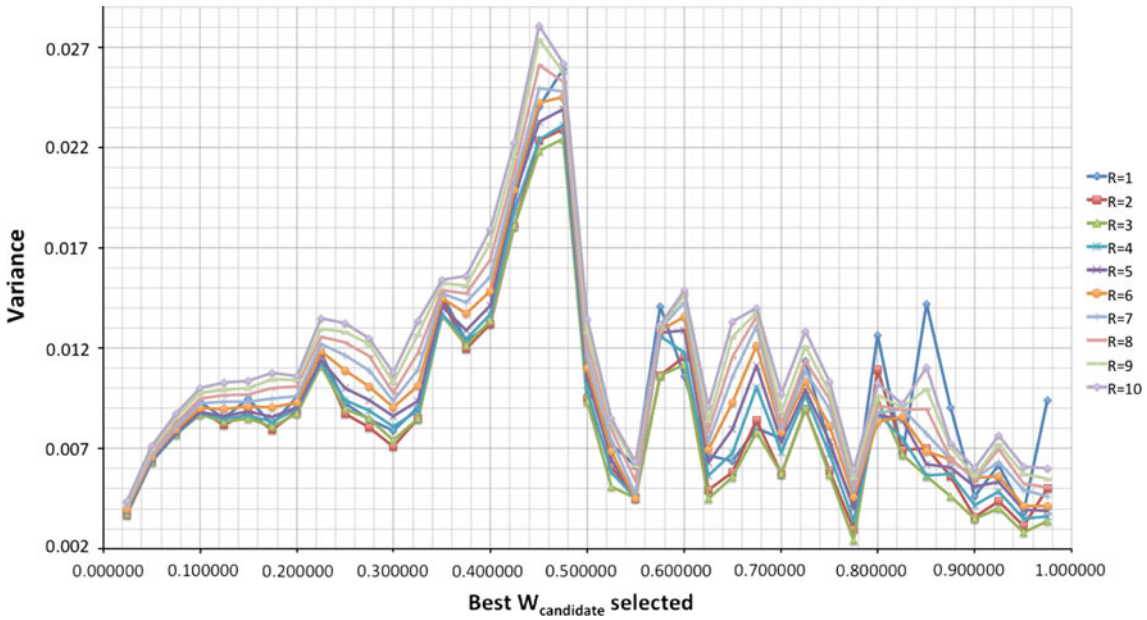

Fig. 20 Variance of EDR values using different numbers of rows of pixels in *Akiyo*

this TM, that is, finding the candidate block with the lowest ESAD.

We substitute the base block in Algorithm 1 with the candidate with the lowest ESAD. The flow of this Algorithm 2 is summarized as follows:

Step 1 For each candidate block, calculate $ESAD_{candidate}$ as depicted in Fig. 9

Step 2 Select the candidate block with the lowest $ESAD_{candidate}$ as base block

Step 3 Based on Eq. 3, perform block matching between A) the current block and B) the weighted combinations of candidate blocks in the search window and the base block found in Step 2. The superimposed block with the lowest distortion together with the weighting used and the MV are found

We need to determine the number of rows (R) of neighboring pixels to be used in the lowest $ESAD_{candidate}$ calculation in the first step of Algorithm 2. Experiment is

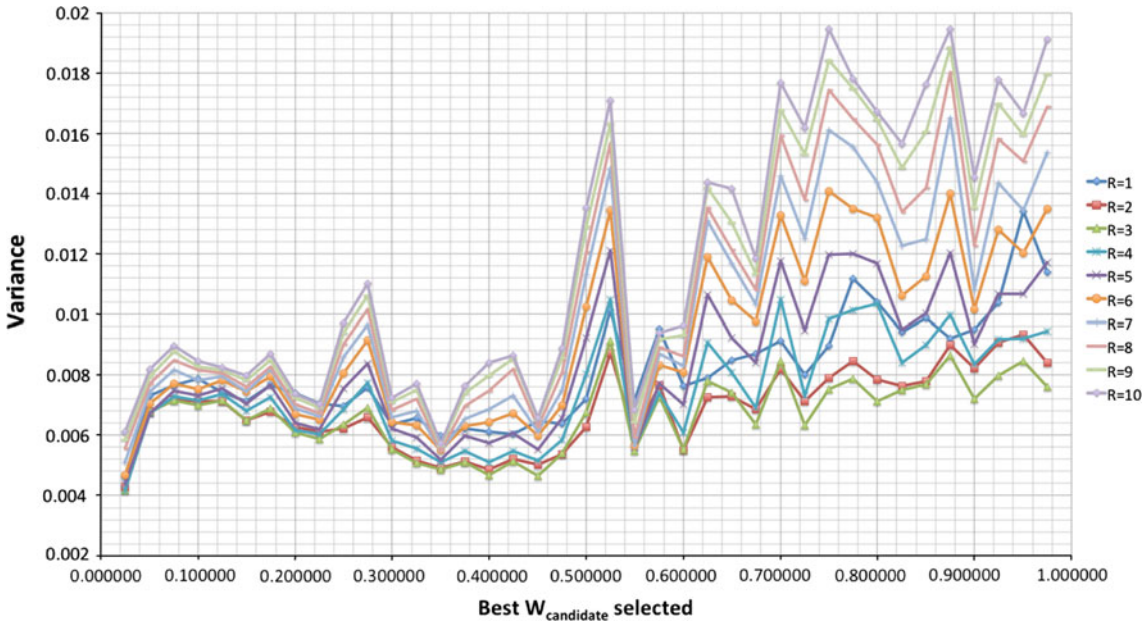

Fig. 21 Variance of EDR values using different numbers of rows of pixels in *Stefan*

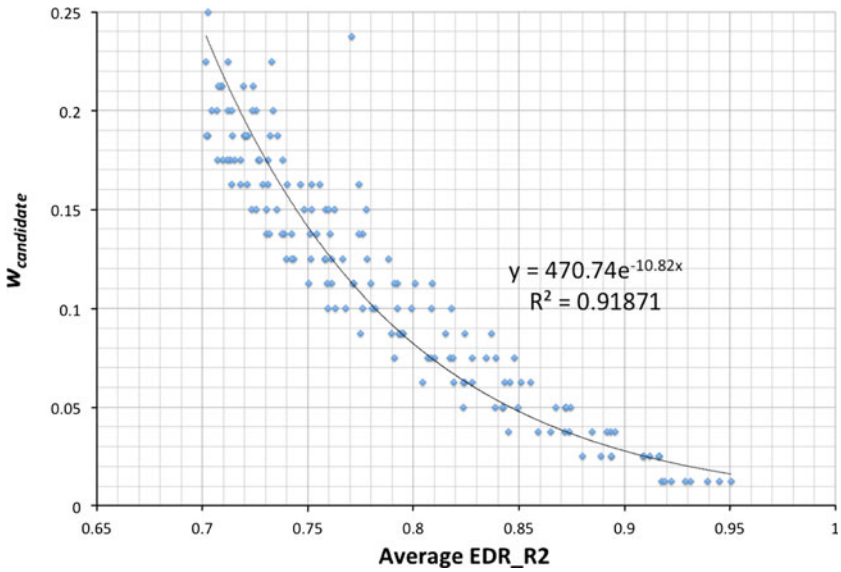

Fig. 22 Best fit exponential equation for the dataset of best $w_{\text{candidate}}$ selected versus average EDR_R2 using Algorithm 2 in the section. $0.7 <$ Average EDR_R2 < 1

conducted to measure the average SAD value of the base block found using different numbers of pixels in $\text{ESAD}_{\text{candidate}}$ calculation. Block size is 16×16 pixels. Ten different numbers of rows are tested. The results are shown in Table 2. The numbers in bold font highlight the lowest average SAD in each test sequence. Using more number of rows (R) in the lowest $\text{ESAD}_{\text{candidate}}$ calculation does not yield a lower distortion base block in the most test cases. We can observe that $R = 3$, that is the number of pixels in $\text{ESAD}_{\text{candidate}}$ calculation for the base block equals 105, yields a lower distortion base block for most sequences. Consider that a lower distortion base block is more favorable because a lower distortion base block resembles the one found by exhaustive full search as in Algorithm 1, we select $R = 3$ in the $\text{ESAD}_{\text{candidate}}$ calculation for finding the base block.

The prediction quality achieved using Algorithm 2 is shown in Table 3. We can observe that the quality improvement is lower than that achieved by Algorithm 1. This is reasonable because in Algorithm 2, the base block is calculated using TM, where in Algorithm 1 the base block is found using conventional exhaustive search. However, Algorithm 2 only requires the transmission of one MV to the decoder side.

Figure 15 plots the best $w_{\text{candidate}}$ selected versus average EDR, using Algorithm 2. We can see that there is a clear trend of increasing EDR with decreasing $w_{\text{candidate}}$ selected.

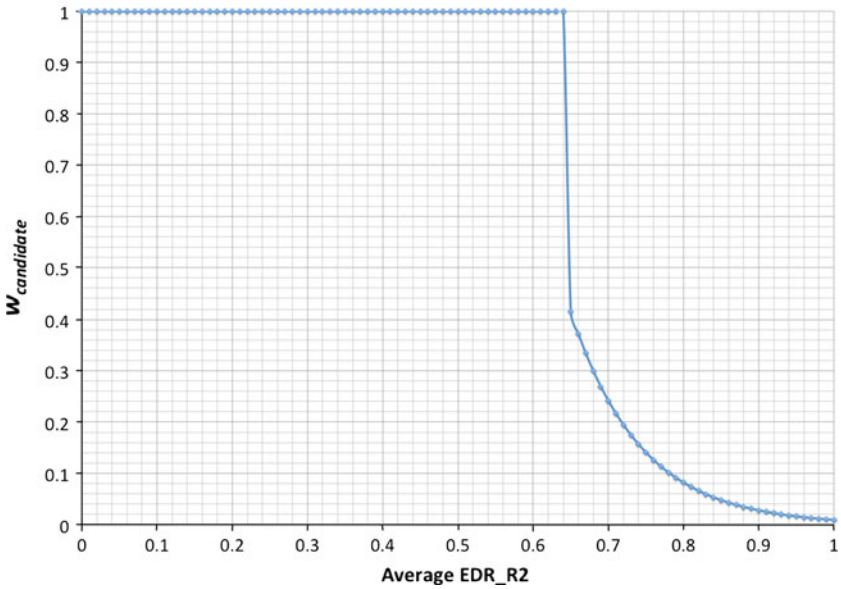

Fig. 23 Optimum $w_{\text{candidate}}$ prediction function

Experiments using different numbers of rows (R) of neighboring pixels in $\text{ESAD}_{\text{candidate}}$ and $\text{ESAD}_{\text{base}}$ in the EDR calculation are conducted. Average EDR_R2, EDR_R3, and EDR_R4 are the average EDR values of using 2 rows, 3 rows, and 4 rows of neighboring pixels in $\text{ESAD}_{\text{candidate}}$ and $\text{ESAD}_{\text{base}}$, respectively. They are plotted against the best $w_{\text{candidate}}$ selected in Figs. 16, 17, 18. We can see that all of them have a clear trend of increasing EDR with decreasing $w_{\text{candidate}}$ selected.

We have to determine which number of rows (R) of neighboring pixels is most suitable for the EDR calculation. We do this by measuring the variances of the EDR values using different R numbers. Figure 19 plots the variances of the EDR values using R from 1 to 10, in test sequence *Foreman*. Figures 20 and 21 are the results in test sequences *Akiyo* and *Stefan*. From Figs. 19–21, it can be seen that the variances are the lowest with $R = 2$ or $R = 3$. This result is consistent in sequences with different motion contents. Consider that $R = 2$ has lower computational complexity because it uses fewer numbers of neighboring pixels in the EDR calculation, we select $R = 2$, that is using two rows of neighboring pixels of the base block and the candidate blocks, to calculate the EDR values.

5 Adaptive superimposed inter-frame search algorithm

In Sect. 3, we propose the use of neighboring pixels of the current block to calculate the EDR. It is found that there is a strong correlation between the EDR and the best weighting in the superimposition of two signal blocks. In Sect. 4, we further reduced the use of two MVs into one. We find that $R = 2$, that is using two rows of adjacent pixels to calculate EDR, can provide a robust prediction of the optimum weightings to be used in the superimposition of the base block and the candidate block.

In a close analysis of the best $w_{\text{candidate}}$ selected versus average estimated distortion ratio EDR_R2 using Algorithm 2 as shown in Fig. 22, we find that in the section average EDR_R2 larger than 0.7, there is a clear exponentially tread of decreasing best $w_{\text{candidate}}$ selected versus average EDR_R2. We can obtain a best fit exponential function for this section. The exponential function has a R-squared value of 0.919, which shows that the correlation is quite high. To remove the sample noise, samples formed by <0.5 % of the total blocks are removed. In the section average EDR_R2 smaller than 0.7, although we can also find best fit functions for the sample points, they cannot help us to predict an optimum $w_{\text{candidate}}$ from a EDR_R2 because the sample points are too much vertically lined-up. Based on this analysis, we obtain an optimum $w_{\text{candidate}}$ function, which is shown in Fig. 23.

For our experiments using CIF and 4CIF sequences, with prediction block size in 16 × 16 pixels, the optimum $w_{\text{candidate}}$ prediction function is:

$$w_{\text{candidate}} = \begin{cases} 1, & \text{EDR} \in [0, 0.7] \\ 470.74e^{-10.82*EDR}, & \text{EDR} \in (0.7, 1.0] \end{cases} \quad (6)$$

When EDR_R2 is smaller than or equal to 0.7, $w_{\text{candidate}}$ is equal to 1. That means w_{base} equals to 0. The base block is not superimposed with the candidate block. This is original block matching. For EDR_R2 larger than 0.7, $w_{\text{candidate}}$ is calculated using the exponential function. Weighted superimposition is performed according to the weighting pair obtained. Based on this model, we propose an adaptive superimposed inter-frame search algorithm (ASIS) algorithm.

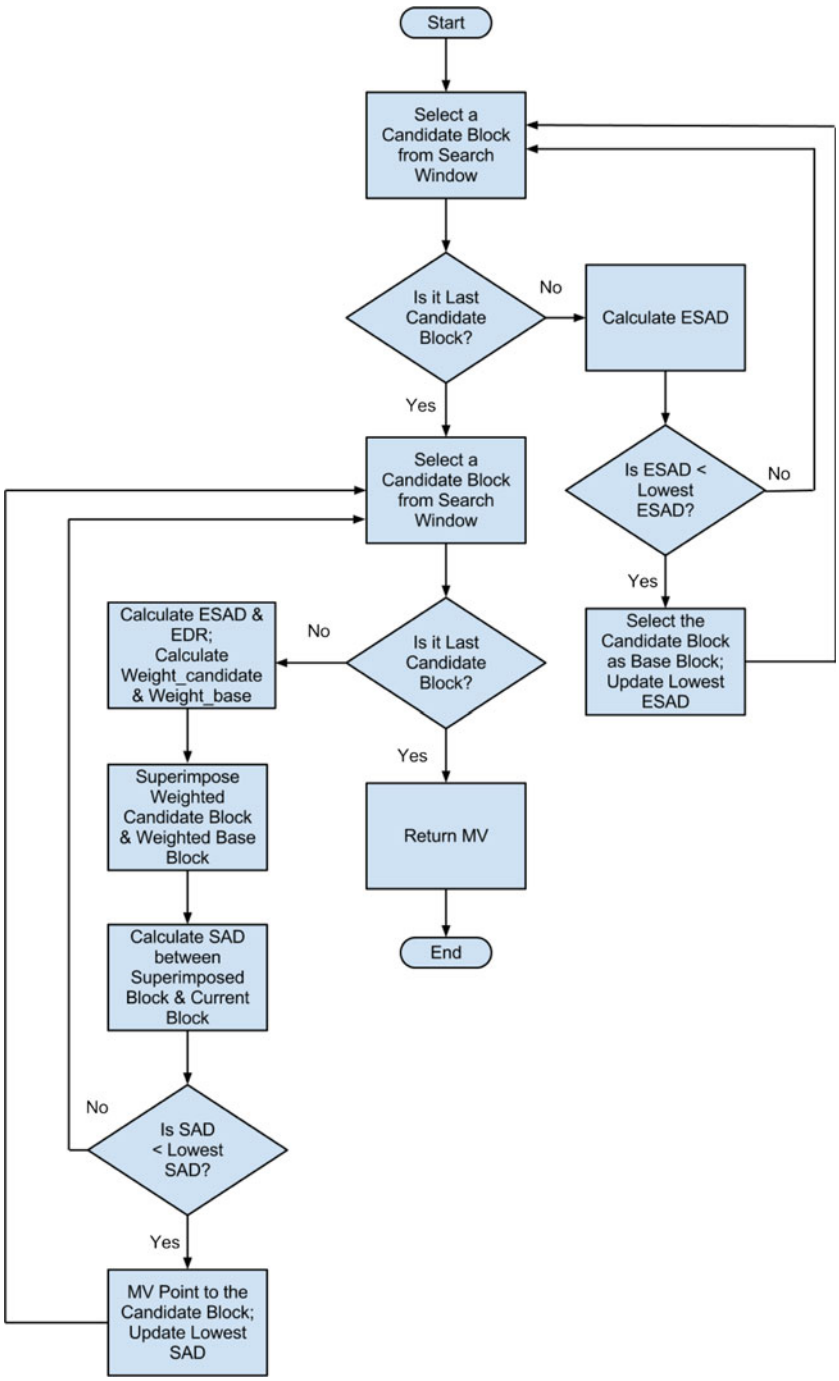

Fig. 24 Block scheme of ASIS

ASIS combines two methods: (1) use the lowest $ESAD_{candidate}$ to find the base block, and (2) use optimum $w_{candidate}$ prediction function in Eq. 6 to estimate the weighting to be used in the superimposition of the base block and the candidate block. The flow of ASIS is summarized as follows:

Step 1 For each candidate block, calculate $ESAD_{candidate}$

Step 2 Select the candidate block with the lowest $ESAD_{candidate}$ as base block. $ESAD_{base}$ is set as this lowest $ESAD_{candidate}$

Step 3 At each candidate position, calculate the EDR using $ESAD_{candidate}$ and $ESAD_{base}$. Calculate the weighting pair $w_{candidate}$ and w_{base} based on the EDR and the function in Eq. 6. Superimpose each candidate block with the base block to form the

Table 4 Average SAD per block and average PSNR per frame of proposed ASIS

	Akiyo (CIF)	Coastguard (CIF)	Foreman (CIF)	Mobile (CIF)	Mother (CIF)
ASIS					
Average SAD per block	121.95	1,080.59	580.76	1,261.79	318.38
Exhaustive search					
Average SAD per block	132.43	1,140.81	602.03	1,329.97	334.22
ASIS					
Average PSNR (dB) per frame	45.44	31.19	34.88	27.73	41.91
Exhaustive search					
Average PSNR (dB) per frame	44.93	30.88	34.79	27.52	41.55
PSNR increment	0.51	0.31	0.09	0.21	0.36
	Sean (CIF)	Stefan (CIF)	City (4CIF)	Crew (4CIF)	Harbor (4CIF)
ASIS					
Average SAD per block	352.52	1,148.14	894.97	967.80	1,065.83
Exhaustive search					
Average SAD per block	370.11	1,197.33	938.85	1,017.22	1,124.01
ASIS					
Average PSNR (dB) per frame	41.61	27.66	32.95	34.80	31.87
Exhaustive search					
Average PSNR (dB) per frame	41.13	27.44	32.69	34.50	31.54
PSNR increment	0.48	0.22	0.26	0.31	0.33

Table 5 Comparison of prediction quality of ASIS with conventional exhaustive search

	Akiyo (CIF)	Coastguard (CIF)	Foreman (CIF)	Mobile (CIF)	Mother (CIF)
% of blocks with better prediction quality than exhaustive search	40.7	79.8	63.3	74.0	65.1
% of blocks with same prediction quality as exhaustive search	58.9	19.8	34.5	24.6	33.4
% of blocks with worse prediction quality than exhaustive search	0.4	0.4	2.1	1.4	1.5
	Sean (CIF)	Stefan (CIF)	City (4CIF)	Crew (4CIF)	Harbor (4CIF)
% of blocks with better prediction quality than exhaustive search	75.5	71.4	78.8	47.7	73.4
% of blocks with same prediction quality as exhaustive search	23.8	26.2	20.4	50.6	25.8
% of blocks with worse prediction quality than exhaustive search	0.7	2.4	0.8	1.8	0.8

superimposed candidate blocks, using the weighting pairs calculated.

Step 4 Perform SAD block matching between A) the current block and B) the superimposed candidate blocks. The superimposed candidate block with the lowest distortion and the MV are found.

The block scheme of ASIS is depicted in Fig. 24. It can be observed that the algorithmic overhead of ASIS is not very complex.

6 Experimental results

The rate-distortion improvement brought by MHMCP is studied in-depth in many research works, for example [4–6], and thus will not be repeated in this paper. We do not want to restrict our proposed MHMCP in any particular coding scheme. Instead, we want to show that our method can estimate the optimum weighting pair adaptively and efficiently. Therefore, we measure the prediction

performance of our proposed optimal weighting estimation method using the simplest objective measurements, namely the average SAD per block and the average PSNR per frame. Because no side information will be transmitted in our proposed algorithm, the gain will be pure prediction improvement.

Table 4 shows the average SAD per block and average PSNR per frame achieved by ASIS and conventional exhaustive search using sequences of different motion contents and of different resolutions. We can see that in all sequences ASIS has lower SAD per block and higher PSNR per frame than conventional exhaustive search. The average PSNR improvement is 0.31 dB. In our algorithm, SAD is used as the distortion measurement and the superimposed block with the lowest SAD is selected because SAD has a lower computational complexity. If mean square error (MSE) is used, the PSNR improvement can be enhanced because PSNR is calculated in MSE. Although the prediction improvement is lower than that in Algorithms 1 and 2, ASIS does not require the transmission of extra side information like most other MHMCP algorithms. This prediction improvement will be a pure gain.

We also measure the number of blocks (excluding frame-boundary blocks) that ASIS achieves better, same, or worse prediction (lower SAD) than conventional exhaustive search. The percentages are shown in Table 5. We can see that a much higher percentage of blocks can be better predicted using the proposed ASIS algorithm. In average, 67 % of the blocks achieve better prediction quality using ASIS, while in average only 1.2 % of the blocks have worse prediction quality than exhaustive search. The phenomenon that ASIS has a worse prediction than conventional exhaustive search in some few cases is reasonable. First, the weighting used in the superimpositions is estimated using the EDR, which may not be the best weighting for the superimposition. Second, the EDR is calculated using the neighboring pixels. When the neighboring pixels are not correlated with the current block, the prediction accuracy of our proposed algorithm will be affected.

7 Conclusion

In this research work, we discovered that distortion ratio is highly correlated to the best weighting used in the multi-hypothesis motion compensation prediction. A method using the neighboring pixels to estimate the distortion ratio is developed. A novel motion compensation prediction algorithm called adaptive superimposed inter-frame search algorithm (ASIS) is proposed. This algorithm has better prediction quality than conventional exhaustive search without additional MV or side information. Therefore, the prediction quality gain will be a pure gain.

The algorithm is also very robust as it works well in video sequences of different motion contents and resolutions. Moreover, the algorithmic decision overhead of the proposed method is very low compared with other multi-hypothesis motion compensation prediction methods. It can be implemented in hardware for real time high-quality video coding applications. We believe this work opens a new and high potential path in motion prediction research.

References

1. Sullivan, G.J.: Multi-hypothesis motion compensation for low bit-rate video coding. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis, Apr 1993, vol. 5, pp. 437–440
2. Girod, B.: Efficiency analysis of multihypothesis motion-compensated prediction for video coding. IEEE Trans. Image Process. **9**(2), 173–183 (2000)
3. Flierl, M., Wiegand, T., Girod, B.: A locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction. In: Proceedings of the Data Compression Conference, Snowbird, Utah, Apr 1998, pp. 239–248
4. Flierl, M., Wiegand, T., Girod, B.: Rate-constrained multihypothesis prediction for motion-compensated video compression. IEEE Trans. Circuits Syst. Video Technol. **12**(11), 957–969 (2002)
5. Kim, C.-S., Kim, R.-C., Lee, S.-U.: Robust transmission of video sequence using double-vector motion compensation. IEEE Trans. Circuits Syst. Video Technol. **11**(9), 1011–1021 (2001)
6. Kung, W.-Y., Kim, C.-S., Jay, C.-C.: Analysis of multihypothesis motion compensated prediction (MHMCP) for robust visual communication. IEEE Trans. Circuits Syst. Video Technol. **16**(1), 146–153 (2006)
7. Suzuki, Y., Boon, C.-S., Tan, T.-K.: Inter frame coding with template matching averaging. In: IEEE International Conference on Image Processing, vol. 3, pp. III-409–III-412 (2007)
8. Haroun, T., Labeau, F.: Robust multiple hypothesis motion compensated prediction within the H.264/AVC standard. In: 2nd International Conference on Image Processing Theory Tools and Applications (IPTA), pp. 149–153, 7–10 July 2010
9. Yang, J., Won, K., Joen, B.: Motion vector coding with selection of an optimal predictive motion vector. In: Optical Engineering, vol. 48, no. 1 (2009)
10. Kamp, S., Evertz, M., Wien, M.: Decoder side motion vector derivation for inter frame video coding. In: IEEE International Conference on Image Processing, pp. 1120–1123 (2008)
11. Yang, W., Au, O.-C., Pang, C., Dai, J., Zou, F., Wen, X., Liu, Y.: An efficient motion vector coding algorithm based on adaptive predictor selection. In: Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2175–2178, May 30 2010–June 2 2010

Author Biographies

Lai-Man Po (M'92, SM'09) received the B.Sc. and Ph.D. degrees, both in electronic engineering, from the City University of Hong Kong in 1988 and 1991, respectively. Since November 1991, he has been with the Department of Electronic Engineering, City University of Hong Kong, and is currently Associate Professor. His research

interests are in vector quantization, motion estimation for video compression and H.264/AVC fast mode decision algorithms.

Ka-Ho Ng (S'10) received his B.E. and M.Phil. degrees, both in electronic engineering, from City University of Hong Kong in 2005 and 2008, respectively. His research interests include video coding and motion estimation.

Xu-Yuan Xu (S'11) received his B.E. degree in information engineering in 2010 in City University of Hong Kong. His final year project "Stereoscopic Video Generation From Monoscopic Video" won the Best Tertiary Student Project of Asia Pacific International And Communication Awards (APICTA) in 2010. He is currently a Ph.D. student in the Department of Electronic Engineering in City University of Hong Kong. His research interests include 3D video coding and 3D view synthesis.

Kwok-Wai Cheung (M'10) received the beng, M.Sc. and Ph.D. degrees from City University of Hong Kong in 1990, 1994 and 2001, all in Electronic Engineering. He worked in Hong Kong Telecom as trainee engineer/project engineer from 1990 to 1995. He was a research student/research assistant at the Department of Electronic Engineering, City University of Hong Kong, from 1996 to 2002. He joined Chu Hai College of Higher Education, Hong Kong, in 2002. Currently, he is Associate Professor, Department of Computer Science, Chu Hai College of Higher Education. Dr. Cheung's research interests are in the areas of image/video coding, multimedia database.

Li-Tong Feng (S'12) received his B.E. degree in electronic science and technology from Harbin Institute of Technology in 2008. His research interests include depth map processing and optical design.