Correlation-Based Motion Vector Processing
With Adaptive Interpolation Scheme for
Motion-Compensated Frame Interpolation

Ai-Mei Huang, Student Member, IEEE, and Truong Nguyen, Fellow, IEEE

Abstract—In this paper, we address the problems of unreliable motion vectors that cause visual artifacts but cannot be detected by high residual energy or bidirectional prediction difference in motion-compensated frame interpolation. A correlation-based motion vector processing method is proposed to detect and correct those unreliable motion vectors by explicitly considering motion vector correlation in the motion vector reliability classification, motion vector correction, and frame interpolation stages. Since our method gradually corrects unreliable motion vectors based on their reliability, we can effectively discover the areas where no motion is reliable to be used, such as occlusions and deformed structures. We also propose an adaptive frame interpolation scheme for the occlusion areas based on the analysis of their surrounding motion distribution. As a result, the interpolated frames using the proposed scheme have clearer structure edges and ghost artifacts are also greatly reduced. Experimental results show that our interpolated results have better visual quality than other methods. In addition, the proposed scheme is robust even for those video sequences that contain multiple and fast motions.

Index Terms—Frame rate up conversion, motion-compensated frame interpolation (MCFI), motion vector correlation, motion vector processing, video occlusions.

I. INTRODUCTION

The application of motion-compensated frame interpolation (MCFI) techniques to increase video frame rate at playback has gained significant attention in both academia and consumer electronics industries for the last decade. This is because MCFI improves temporal resolution by interpolating extra frames and can be used to reduce motion jerkiness for video applications that have low bandwidth requirement such as wireless video broadcasting, or to remove motion blurriness for LCD-panel display by increasing the refresh rate from 60 to 120 Hz. MCFI requires motion information between two frames, which can be either re-estimated at the decoder or retrieved directly from the received bitstreams, depending on available resources of the devices. For example, it may be preferred to use the received motion vectors (MVs) for those resource-limited hand-held devices, while TV applications can afford to apply a motion estimator to obtain better MVs. Unfortunately, the received MVs or re-estimated MVs simply using block matching algorithm (BMA) are often unreliable for frame interpolation in the sense that they fail to represent true motion. Directly employing these MVs usually results in unpleasant artifacts such as blockiness, ghost artifacts and deformed structures in the interpolated frames. Therefore, it is very challenging for MCFI to produce high quality, artifact-free interpolated frames without proper MV processing.

In the literature, many motion estimation algorithms performed at the decoder have been proposed for MCFI in order to obtain true motion. Along with the commonly used BMA, MVs can be estimated by further considering spatial and temporal correlations [1]–[3]. A hierarchical BMA was presented in [4], where three different window sizes are used to search for true MVs, based on the assumption that a large window is more suitable for finding global motions but a small window can find better local motions. A similar concept was proposed in [5], but the window sizes for motion estimation are determined by performing a similarity measure on the previously estimated MVs. By considering the motion distribution on object boundaries, image segmentation techniques are employed to further refine the estimated motion vector field (MVF) [3], [4], [6]–[8].

On the other hand, if motion estimation is not possible at the decoder, motion vector processing techniques can be used to obtain smoother motion from the received MVF. Vector median filter (VMF) was proposed to remove motion outliers [9]. An adaptive VMF was then presented in [10] by considering displaced frame difference. Sekiguchi et al. further used weighted averaging of neighboring MVs by exploiting the predicted error [11]. In addition to MV processing, MV reliability analysis was also addressed to help correct unreliable MVs. Sasai et al. determined the received MV reliability by counting and calculating the number of intra blocks, isolated MVs, and MV variance [12]. A frame that has an unreliable MVF is not used for interpolation. The works in [13] and [14] conducted a prior MV classification and only re-estimated those unreliable MVs to reduce computational complexity. We also proposed using residual energies to classify MV reliability from the received MVF and used this prior knowledge to correct unreliable MVs [15], [16]. As most of the MCFI approaches are block based, blocking artifacts are easily observed in the interpolated frames. In order to provide better visual quality, overlapped block motion compensation was adopted to smooth out the blockiness along the block boundaries [3], [17]. Alternat-
tively, instead of eliminating the blockiness artifacts in the pixel domain, the works in [18] and [19] resampled the received MVF into smaller blocks and minimized the difference among these finer MVs.

Algorithms for reducing visual artifacts in occlusion areas were also addressed in the literature. Lee et al. suggested weighted averaging interpolation by considering multiple motion trajectories and the corresponding prediction errors [20]. A pixel-wise interpolation approach using nonlinear filtering was addressed in [21]. Chen proposed adaptively choosing the forward and backward predictions based on the block boundary absolute difference [22]. Similar concepts were also presented in [2] and [23], but the prediction selection was performed at the encoder. The works in [4] and [7] embedded image segmentation information in the bitstream as side information so that decoder can chose better predictions for uncovered and covered areas.

In our previous work [24], a multistage MV processing method was proposed by gradually refining motion from the largest size of up to $32 \times 32$ to the smallest size of $4 \times 4$. The residual energy was used in MV reliability classification for the received MVF. Since these discovered unreliable MVs often occur at edges of moving objects, we suggested merging macroblocks (MBs) that are connected by high residual energy distribution and selecting a single best MV from the neighboring MVs for each merged group. Although the results showed that the interpolation artifacts can greatly be removed, we also noticed that not all of the unreliable MVs can be detected by their residual energy and coding types. This usually happens when there are repeated patterns in the background. The encoder may choose MVs that deviate from the real motion but still have smallest sum of absolute difference. In addition, we adopted the work in [19] as the last MV processing stage to smooth out the entire MVF to remove blocking artifacts, but it also created ghost artifact on the motion boundaries. Finally, our previous method did not explicitly consider the problem when occlusion happens, which usually can cause severe ghost artifacts.

In this paper, we further extend our previous method to solve the aforementioned problems. To detect unreliable MVs that have low residual energy, we propose classifying those MVs by calculating MV correlation in their local neighborhood. Unlike VMF that removes MV outliers one at a time and usually fails when irregular MVs occur in a cluster, we merge unreliable MVs due to low MV correlation and select a single best MV from their neighbors. In addition, based on the already calculated MV correlation information, we present an adaptive correlation-based MV averaging filter to remove blockiness artifacts. Both motion correlation and motion reliability are explicitly considered to determine the weights, so structure edges will not be blurred due to the MV smoothing process. Moreover, to minimize the ghost artifacts due to video occlusion, an adaptive frame interpolation scheme is proposed to dynamically determine to use forward or backward predictions for occlusion areas based on their surrounding MV distribution. In these occlusion areas, the MVs obtained by minimizing the bidirectional prediction difference are often unreliable once the occlusion area is larger than the unit block size of the received MVF. Therefore, we propose spirally assigning the neighboring corrected MVs to the occlusion areas and only forward or backward prediction is adopted for the frame interpolation. In this way, object motion and background motion can evenly distribute in large occlusion areas. Even for the partially deformed object, we also can reconstruct the object using adjacent corrected motion. Compared to the methods in [4], [7], and [25], which used image segmentation techniques and traced back several frames to determine occlusions, our occlusion areas are detected only based on the analysis of the current frame and the proposed occlusion processing is block based, which means we have lower complexity.

The rest of this paper is organized as follows. The challenges in MCFI are further described in Section II. In Section III, we demonstrate the classification process and the merging process using MV correlation for the received MVF. The proposed correlation-based MV processing method and frame interpolation scheme are presented in Sections IV and V, respectively. Experimental comparisons are given in Section VI.

II. CHALLENGES IN MOTION-COMPENSATED FRAME INTERPOLATION

In MCFI, the interpolated frame, $f_t$, is often obtained by one of the following two different methods:

$$f_t \left( i + \frac{1}{2} v_x, j + \frac{1}{2} v_y \right)$$

$$= \frac{1}{2} \cdot f_{t-1}(i + v_x, j + v_y)$$

$$+ \frac{1}{2} \cdot f_{t+1}(i, j)$$

(1)

or

$$f_t(i, j)$$

$$= \frac{1}{2} \cdot f_{t-1}(i + \frac{1}{2} v_x, j + \frac{1}{2} v_y)$$

$$+ \frac{1}{2} \cdot f_{t+1}(i - \frac{1}{2} v_x, j - \frac{1}{2} v_y)$$

(2)

where $f_{t-1}$ and $f_{t+1}$ denote the previous and current reconstructed frames, respectively. $v = (v_x, v_y)$ is the MV used to reconstruct the frame $f_{t+1}$, or it can be re-estimated at the decoder.

The first method using (1) assumes the interpolated frame can be produced by motion compensating from $f_{t-1}$ and $f_{t+1}$ along the motion trajectory. If a block-based MCFI is used, holes and overlapped regions frequently appear in the interpolated frame. As a result, a computationally expensive spatial interpolation is often adopted to fill holes [26]. In addition, tracking the motion trajectory and recording holes and overlapped regions can be a complicated process. Different from the complex motion-trajectory-based MCFI scheme, the second method in (2) simply takes the MVs of the co-located blocks and divides them by two to form forward and backward MVs. Then, the interpolated frame $f_t$ can be obtained by averaging forward and backward predictions. This method can also be referred to as bidirectional MCFI approach. Although there is no spatial interpolation required, ghost artifacts can easily occur in the interpolated frame if the received MV does not represent the real motion. Examples using these two schemes are demonstrated in Fig. 1(a) and (b), where black areas represent holes. In Fig. 1(b), we can observe that ghost artifacts occur at the motion boundary or the places where
MVFs are unreliable. In this paper, we use the second method in (2) to interpolate a frame because of its simplicity.

In [24], we have discussed the correlation between the received MV reliability and the associated residual energy. Our study suggested that a MV is often unreliable if its residual energy is high based on the observation that artifacts such as deformed structures often occur at the areas where high residual energies are decoded. However, there still exist some other potential issues for directly using the received MVF. In the following subsections, we further describe the challenges in MCFI that create visual artifacts in the interpolated frames.

A. Co-Located Motion Vectors

The MVFs in the received bitstreams are often obtained using block-based motion estimation at the encoder from $f_{t-1}$ to $f_{t+1}$. Even though these MVFs may represent true motion for blocks in $f_{t+1}$, they may not represent the motion of their co-located blocks in $f_t$. This is because the movement is estimated not based on the object position in the interpolated frame $f_t$ but based on the current reconstructed frame $f_{t+1}$. Once an object moves across more than one block from $f_{t-1}$ to $f_{t+1}$, those blocks between the object’s previous position and current position may have different motion or even different coding modes. Since the bidirectional MCFI takes the MVFs of the co-located blocks in $f_{t+1}$, those MVFs may fail to capture the object movement in $f_t$ and result in ghost artifacts.

We use Fig. 2 to illustrate this scenario, in which the moving object is composed of four blocks, $\{B_1, B_2, B_3, B_4\}$. This object moves across two blocks in both horizontal and vertical directions from $f_{t-1}$ to $f_{t+1}$. When $f_t$ is being interpolated, we take the MVFs of the co-located blocks in $f_{t+1}$. As a result, the gray blocks in $f_t$ cannot be interpolated successfully since the co-located blocks may not have the same motion as $B_1$, $B_2$, and $B_3$ in $f_{t+1}$. Only the block that uses the MV of $B_3$ happens to be interpolated accurately. In bidirectional MCFI, not only does the received MVF have this problem but any other block-based motion estimation algorithms that only consider unidirectional motion would encounter the same difficulty. To overcome this problem, bidirectional motion estimation can be used to find the best MV by minimizing absolute bidirectional prediction difference, instead of using only forward or backward motion estimation [3], [14], [17], [22]. Some other works suggested tracing motion trajectory projection on the interpolated frame [8], [11] but this approach assumes the received MVFs are accurate, which is not always true. Based on this discussion, we can, therefore, conclude that the received MVFs in $f_{t+1}$ should all be subject to further MV processing by checking their bidirectional prediction difference.

B. Irregular Motion Vectors

Generally speaking, the MVF between two frames is supposed to be smooth, except at the motion boundaries and occlusion areas, which can be detected by high residual energy. However, irregular MVFs with low residual energy can still occur, depending on how the motion estimation is performed at the encoder. This is because the encoder usually takes coding difficulties, such as prediction errors and the difference between the estimated MV and its neighboring causal MVs, as the cost function. As such, once an irregular MV has been estimated previously, it may affect the MV decision for the following blocks, especially in an area where repeated pattern occurs. That is, irregular MVs may appear in a cluster. VMF is often regarded as an efficient way to correct motion outliers, but it only works well for isolated irregular MVs [8], [10], [12], [23]. Once many irregular MVs occur in the same area, these unreliable MVs may dominate the performance of the VMF. Furthermore, these irregular MVs often cannot be detected by residual energy or bidirectional prediction difference.

Fig. 3 shows an example of failed interpolation due to irregular MVFs. Fig. 3(a) and (b) are reconstructed frame 181 and reconstructed frame 183 of FOREMAN sequence, respectively, and Fig. 3(d) is the interpolated result of frame 182 using direct MCFI. Since the areas of the building structure and the human face have repeated patterns and smooth contents, block-based motion estimation can choose any similar areas wherever the minimal absolute prediction difference occurs, instead of finding the true motion. Therefore, the motion distribution becomes very irregular at these two areas. These unreliable MVs cause severe visual artifacts such as pattern dislocation. Ha et al. suggested gathering statistics about the global motion from past few frames so that the motion estimation can favor the global motion to produce a smoother MVF for those frames that have repeated patterns [27]. However, this approach can fail easily when the video contains multiple object movement. Although we have proposed the method in [24] to discover unreliable MVs using residual energy, unfortunately, these irregular MVs often
have low residual energy as shown in Fig. 3(c) and are regarded as reliable MVs.

In order to effectively detect all unreliable MVs, the MV reliability classification process should explicitly consider both residual energy and MV correlation. This is because high residual energy is usually responsible for the unreliable MVs on the motion boundaries and deformed structures, while the MV correlation can reflect the motion reliability on the smooth areas and periodic scene.

C. Video Occlusions

Areas where new objects appear or existing objects disappear can be referred to as video occlusions. In these areas, the encoder may decide any MV that results in smallest prediction errors or an intracode mode is used. For bidirectional MCFI, the problem with the covered and uncovered regions is that the MVs of the co-located blocks in \( f_t \) cannot be decided as none of the received MVs in \( f_{t+1} \) are appropriate. For example, in Fig. 4, the object in the interpolated frame, \( f_t \), should be in the middle of the motion trajectory. We also observe that the covered areas in \( f_{t-1} \) and the uncovered areas in \( f_{t+1} \) (and vice versa) become occlusions since the perfect matches from \( f_{t-1} \) to \( f_{t+1} \) may not exist. For these occlusion areas, which are indicated by gray color, the visual artifacts caused by occlusion cannot be removed completely even though we have the correct MVs for the moving object. As a result, we can often observe ghost artifacts around the object contour. A commonly observed case is the frame boundary where the matched predictions do not exist if the video content is not static.

Conventionally, image segmentation techniques are often used to identify uncovered and covered regions. To reduce the ghost artifacts, the work in [20] imposed blurriness effect on the motion boundary. In order to remove the artifacts on the frame boundaries, we have proposed using unidirectional frame interpolation by examining whether the MVs point to outside or inside the frame boundary [24]. In this way, we can choose better motion compensation that is located inside the frame to avoid ghost artifacts. However, the occlusion can occur anywhere within the frame. To solve this problem in the block-based frame interpolation, the motion distribution around the occlusion region should be explicitly analyzed to determine which movements most likely cause the occlusion. Hence, based on this information, a different frame interpolation scenario that adaptively selects forward or backward prediction can be used to alleviate the ghost artifacts due to occlusion.

III. MOTION VECTOR ANALYSIS FOR MOTION-COMPENSATED FRAME INTERPOLATION

Based on the observation in the previous section, it is reasonable to argue that most of visual artifacts are caused by the MVs with high residual energies, the MVs with low correlations, and areas where no MVs are available. Hence, the received MVs should be further classified according to these different criteria. 8 × 8 block size is used for the MV classification process, so if an encoder uses the smaller block size for the motion estimation such as H.264, MVs will be averaged prior to the classification.

A. Motion Vector Classification

Let \( \mathbf{v}_{mn} \) denote the MV of each 8 × 8 block, \( \mathbf{b}_{mn} \), we classify \( \mathbf{v}_{mn} \) into three different reliability levels, unreliable due to high residual energy (\( L_1 \)), unreliable due to low inter-MV correlation (\( L_2 \)), and possibly unreliable (\( L_3 \)). The classification process is similar to the method in [24], with further analysis of those MVs with low residual energies. First, we calculate the residual energy for \( \mathbf{b}_{mn} \) by taking the sum of the absolute value of the reconstructed prediction errors

\[
E_{mn} = \sum_{(i,j) \in b_{mn}} |r_Y(i,j)| + 2 \cdot \left( \sum_{(i,j) \in b_{mn}} |r_{C_b}(i,j)| \right) + \sum_{(i,j) \in b_{mn}} |r_{C_r}(i,j)|
\]

(3)

where \( r_Y(i,j) \), \( r_{C_b}(i,j) \), and \( r_{C_r}(i,j) \) are the reconstructed residual signals of Y, Cb and Cr components, respectively. If the
residual energy, $E_{mn}$, is greater than a predefined threshold, $\varepsilon_1$, $v_{mn}$ will be classified as an unreliable MV and be put into the reliability set, $L_1$. We also consider intracoded blocks having unreliable MVs in $L_1$, since their residual energies are often too high to be encoded as intercoded mode. For the MVs whose residual energies are less than $\varepsilon_1$, we calculate their MV correlation with neighboring MVs to check if they are unreliable or possibly unreliable. MVs that are very dissimilar to adjacent MVs will be regarded as low correlated MVs and be placed into the reliability set, $L_2$. For the bidirectional MCFI scheme, since the received forward MVs cannot truly represent their co-located MBs in the interpolated frame, we have to examine the reliability of all received MVs. That is, for the MVs that are not classified yet, we will put them into the possibly unreliable set, $L_3$.

In order to detect the irregular MVs that have low residual energy, we calculate the correlation index of each MV to all its available adjacent MVs. Here, the correlation index is defined using Euclidian distance between $v_{mn}$ and its adjacent MVs. First, we calculate the motion magnitude distance as follows:

$$d(m, n, i, j) = ||v_{mn} - v_{m+n+j}||^2$$

(4)

where $v_{m+n+j}$ are the surrounding available MVs, i.e., inter-MVs. According to our observation, $d$ is usually higher than other areas if the local movement is relatively large. Therefore, to reduce the sensitivity from the motion magnitude values, the correlation index is defined as the magnitude variance in the local neighborhood

$$C_{mn} = \frac{1}{9} \frac{1}{9} \sum_{i=-1}^{1} \sum_{j=-1}^{1} d(m, n, i, j) .$$

(5)

To determine if the remaining MVs are unreliable or possibly unreliable, we compare $C_{mn}$ with the averaged MV correlation index in this neighborhood, which can be written in the following:

$$C_{mn}^{avg} = \frac{1}{9} \sum_{i=-1}^{1} \sum_{j=-1}^{1} C_{m+n+j} .$$

(6)

If $C_{mn}$ is greater than $C_{mn}^{avg}$ and the motion distance is greater than half of the averaged magnitude, $v_{mn}$ will be considered as an unreliable MV. After the residual energy and the correlation index classifications, if there are still MVs that are not classified yet, we place them into the reliability set $L_3$. The MV reliability map can, therefore, be created as follows:

$$\text{MVRM}_l(m, n) = \begin{cases} L_1, & \text{if } E_{mn} \geq \varepsilon_1 \text{ or } b_{mn} \text{ is intracoded} \\ L_2, & \text{if } C_{mn} > C_{mn}^{avg} \text{ and } C_{mn} > 0.5 \\ L_3, & \text{otherwise} \end{cases}$$

(7)

Based on this MV reliability analysis, the MV residual merging map and the MV correlation merging map can then be produced to assist the following MV correction processes.

B. Macroblock Merging Map for Motion Vector Processing

In [24], we have suggested that unreliable MVs should be grouped into larger blocks for MV correction according to the residual energy distribution and predefined merging shapes. We also apply this algorithm to all MVs in $L_3$ to create a MV residual merging map. However, since MVs of $L_1$ and $L_2$ are identified due to different reasons, they should not be merged together. For MVs in $L_2$, most of conventional methods prefer using VMF for the MV correction. However, as we mention in the previous section, this method can easily fail if low correlated MVs occur in a cluster. If we assume that the adjacent highly correlated MVs should belong to the same object, then we should not correct these irregular MVs separately but merge those MBs having similar irregular MVs as a merged group for further MV correction. From the observation of the received MVF, irregular MVs often come from the scenes composed of repeated or texture-like patterns, or smooth contents. As a merged group, the MV selection process, which will be described in the next section, is more likely to choose the right motion. This is because larger block size has more pixel references than smaller ones, and the estimation results are not easily to be affected by the scene content.

Based on the MV analysis, we then create two merging maps for the MV processing stage so that adjacent unreliable MVs can be corrected together and yield a single best MV. The merging process is performed on a MB basis, and all MBs that contain unreliable MVs will be examined in a raster scan order. The residual MB merging map, MBM$_e$, is first created based on residual energy distribution described in [24]. For unreliable MVs in $L_2$, we create a separate correlation MB merging map, MBM$_c$. If we find a MB containing unreliable MVs in $L_2$, adjacent MBs that have not yet been merged will be checked whether they have similar MVs to these unreliable MVs. If this is the case, these MBs will be merged together. If there are no similar MVs in the neighborhood, this MB will remain as a single $16 \times 16$ block and this unreliable MV is regarded as an isolated MV. MVs are considered similar if their angular distance, $d_\theta$, and Euclidian distance, $d$, are less than predefined thresholds, $\varepsilon_\theta$, and $\varepsilon_m$, respectively. Here, the angular distances $d_\theta$ can be represented as

$$d_\theta(m, n, i, j) = 1 - \frac{v_{mn} \cdot v_{ij}}{|v_{mn}| |v_{ij}|} = 1 - \cos \theta$$

(8)

where $\theta$ is the angle difference between $v_{ij}$ and $v_{mn}$. $d$ can be obtained using (4). We choose $32 \times 32$ as the maximum block size for the merging process. In both merging maps, each merged group will be assigned a unique index number to the MBs belonging to the same group. For the MV reliability classification process, the best value of $\varepsilon_1$ has been found empirically in [24] as 1100 and we also use the same value in the simulation.

IV. CORRELATION-BASED MOTION VECTOR PROCESSING USING BIDIRECTIONAL PREDICTION DIFFERENCE

According to the residual merging map (MBM$_e$), the correlation merging map (MBM$_c$), and the received MVF, we can select the best MV for each merged group from its own and neighboring MVs by minimizing the absolute bidirectional prediction difference (ABPD). As shown in Fig. 5, this proposed
MV selection process is along with an iterative threshold mechanism to decide when the process should be terminated. Initially, only the MV whose ABPD is less than the predefined threshold value, $\varepsilon_2$, will be selected to correct unreliable MVs within each merged group. If this is not the case, these unreliable MVs will remain the same and wait for the future correction. To trace the MV correction status, during each MV selection pass, only the index numbers for the corrected MVs and their associated merged MBs will be cleared from the merging maps and the MV reliability map. If there are still nonzero indices in the updated MBMM$_r$ and the MVF status is no longer changed due to the limitation on the threshold value, $\varepsilon_2$ will be increased for the next MV correction pass. Since image may consist of various contents such as objects with constant intensity (i.e., low ABPD value) and sharp edges (i.e., high ABPD values), by adaptively adjusting threshold values, we can gradually choose the best motion for each merged group. In this way, not only can better motion with lower ABPD value propagate to the neighborhood of unreliable MVs during the MV selection process, but unreliable MVs can be corrected according to their degree of MV reliability, i.e., ABPD values.

For the subsequent MV selection pass, since the correlation distribution has been changed, the irregular MVs that are classified in $L_3$ previously can be detected again. This is because these low-correlated MVs usually appear in a cluster and the initially detected unreliable MVs are probably located on the boundary where irregular MVs start to occur. As shown in Fig. 5, to correct these unreliable MVs, we, therefore, recursively examine the updated MV correlation distribution and modify the correlation merging map, MBMM$_r$, accordingly. The subsequent MV correlation classification will skip the unreliable MVs that are not corrected yet in the previous pass. This is because their merging status is determined by the residual energy distribution rather than the motion correlation. That is, for the unreliable MVs in $L_3$, if their reliability level is not changed, their merging status will be the same.

The MV selection process stops whenever the merged groups in MBMM$_r$ are all assigned the single best motion or $\varepsilon_2$ is
greater than a predefined maximum threshold value, $\varepsilon_{2,\text{max}}$. This maximum threshold value should be designed to find appropriate motion for all merged groups, so if there are merged groups of MBMM$_{r}$ that are not assigned any motion due to high ABPD values, apparently, they are occlusions. We will leave their MVF blank, since forcing them to have new MVs using a very high $\varepsilon_{2}$ value still cannot obtain reliable MVs. For these occlusion areas, a parallel occlusion classification is undertaken for further processing. The reason why the iteration process is only defined based on MBMM$_{r}$ is that most of occlusions occur in high residual energy and intracoded areas. In [24], we obtained the best $\varepsilon_{2}$ value, 45, based on the experimental analysis of FOREMAN sequence. Therefore, for the unreliable MVs in $L_{2}$, we simply use same $\varepsilon_{2}$ value, but for unreliable MVs in $L_{1}$, we start the $\varepsilon_{2}$ value from 30 with the step size of 15, and set the $\varepsilon_{2,\text{max}}$ value to be 60 for all test sequences.

As shown in Fig. 5, to ensure the remaining unchecked MVs, i.e., possibly unreliable MVs, are truly reliable, we simply apply the MV selection process with fixed block size of 16 x 16 to examine if their MVs do have smaller ABPD than others. Since the MVF is corrected and regular at this stage, even with small fixed block size, the possibility to select inaccurate MVs for the MVs in $L_{3}$ is relatively low. As the MV selection always prefers the major motion for each merged group, once the selected motion cannot well represent the details such as the motion boundary, the areas with different motion usually have higher ABPD values than other areas. Therefore, we classify the new obtained MVF based on the smaller 8 x 8 block size using the ABPD energy distribution, and the reclassification method and the refinement process in [24]. After the new identified unreliable MVs in MVRM$_{2}$ are further refined using surrounding dissimilar MVs, to minimize the blockiness artifacts and keep the object edge sharp at the same time, we resample the 8 x 8 MVF into finer 4 x 4 MVF with consideration on both MV correlation and ABPD distribution. The following subsections will present in details the MV selection process and the MVF interpolation process. In Fig. 5, we also demonstrate how the received MVF is gradually corrected and improved based on the interpolated results for each stage.

### A. Motion Vector Selection

We take MBMM$_{r}$ and MBMM$_{l}$ as initial reference maps for the MV correction process. In [24], the MV selection process chooses the best MV, $v_{b}^{*}$, for each merged group based on minimum absolute difference between forward and backward predictions, which can be written as follows:

$$v_{b}^{*} = \arg \min_{v \in S} (\text{ABPD}(v))$$  \hspace{1cm} (9)

where

$$\text{ABPD}(v) = \frac{1}{N_{G}} \sum_{x,y \in G} \left[ f_{t-1} \left( x + \frac{1}{2} v_{x}, y + \frac{1}{2} v_{y} \right) - f_{t+1} \left( x - \frac{1}{2} v_{x}, y - \frac{1}{2} v_{y} \right) \right].$$

$S$ denotes the set of the MV candidates, which is composed of neighboring MVs and MVs within the merged group. $G$ denotes the current merged group and $N_{G}$ is the corresponding size. For the merged groups indicated by MBMM$_{l}$, we simply select the best motion using (9) as described in [24].

However, those unreliable MVs in $L_{2}$ are identified due to irregular MV distribution. Depending on what the scene is composed of and how the MV estimation is performed at the encoder, the selected MVs may tend to distribute randomly if we merely consider the minimum ABPD. Therefore, we take both minimum ABPD and MV correlation into account for merged groups of MBMM$_{r}$. That is, we choose MVs that have minimum ABPD among adjacent MVs that have higher correlations than the original MV. Hence, the candidate set, $S$, can be re-written as follows:

$$S = \begin{cases} v_{b,i,j} & \text{if } C(v_{b,i,j}) < C(v_{m,n}) \\ \emptyset & \text{otherwise}, \end{cases}$$

Slightly different from (5), the correlation index, $C(v)$, is calculated based on the boundary MVs of the merged group and its neighboring MVs. That is, each merged group is considered as a unit block, and only the motion distances between the merged group and its neighboring available MVs are used to select MV candidates. This is because the MVs inside the merged group are similar and directly calculating the motion distance may not truly reflect the motion correlation for the merged group. If the correlation index of $v_{b,i,j}$ is less than the original correlation index, $v_{m,n}$, $v_{b,i,j}$ will then be considered as MV candidates for $S$. Once the best MV exists, we assign it to all MBs within the merged group. If not, it means that there are no other MVs having higher correlation and better representing the local motion than the original MVs. In such case, these unreliable MVs might belong to the area where the motion starts to differ or the different moving object. We, therefore, skip this MV selection process and keep the MVs and the MBMM$_{r}$ unchanged.

Please note that the MV selection processes for both types of unreliable MVs are performed in the same pass. Hence, if the MV assignment for the merged groups of MBMM$_{r}$ is not completed due to the threshold mechanism, we will have to update MBMM$_{l}$ according to the current MV correlation distribution for the next pass. Therefore, the updated MV map can be created as follows:

$$\text{MVRM}_{l}^{*}(m,n) = \begin{cases} L_{1}, & \text{if } \text{MVRM}_{l}(m,n) = L_{1} \text{ and } C_{m,n} \geq C_{m,n}^{\text{avg}} \text{ and } C_{m,n} > 0.5 \\ L_{2}, & \text{if } C_{m,n}^{\text{avg}} > C_{m,n} \text{ and } C_{m,n} > 0.5 \\ L_{3}, & \text{otherwise}, \end{cases}$$  \hspace{1cm} (10)

If the unreliable MVs of $L_{1}$ are not correct yet, they will stay in $L_{1}$ level for the next correction. Based on the updated correlation indices and averaged correlation indices, $C_{m,n}^{\text{avg}}$ and $C_{m,n}^{\text{avg}}$, we can discover more low-correlated unreliable MVs that have not been detected in the first place. According to the updated MVRM$_{l}$, the corresponding MBMM$_{l}$ can be recreated as well. For MBMM$_{r}$, the merging status will be the same except the blocks whose MVs have been corrected. In the end of each pass, we will check if unreliable MVs in MBMM$_{r}$ are all corrected, if the threshold value is still within the predefined range, and if the occlusion caused by unreliable MVs of $L_{1}$ has reasonable size, to decide when the MV selection should be completed.
Fig. 5, due to the consideration on MV correlation, the unreliable motion in the direct interpolated result has been corrected. The ghost artifacts around the shirt area do not appear in the interpolated result using MVF\(^2\).

Although the proposed MV selection process needs to recursively calculate ABPD for each merged group, the MV candidates are actually almost the same to the MVs in the previous pass except the new MVs that have propagated to the neighborhood. In order to reduce the computation load, we can create two tables to save the MVs that have occurred before and their corresponding ABPD values. In this way, the computation complexity required for unreliable MVs in \(L_1\) is similar to the method in [24]. For the identified low-correlated MVs, we can also use same approach to avoid the repeated calculation of ABPD values.

### B. Motion Vector Averaging Based on MV Correlation

To reduce the blockiness artifacts, we further resample the MVF from one MV with 8 \(\times\) 8 block size, \(v_{m,n}\), into four MVs, \(\{v_{m,n}^1, v_{m,n}^2, v_{m,n}^3, v_{m,n}^4\}\). In general, the vector averaging filter can always provide desirable MV smoothing effect for reducing blockiness artifacts. However, the visual quality of the motion sensitive areas such as sharp object edges and striped textures are often distorted by the MV smoothing process. This is because in these areas, unpleasant artifacts can easily show up even when the motion is only modified slightly. Therefore, the motion smoothing should be performed with consideration of the motion correlation and the scene contents so that we can reduce the MV smoothing impact from neighboring MVs on these motion sensitive areas. Based on this argument, we propose an adaptively weighted vector averaging process as follows:

\[
\begin{align*}
\mathbf{v}_{m,n}^k &= \left[ \frac{\sum_{i,j} f \left( d_{m,n}^k, \epsilon_{m,n}^k \right) v_{x,i,j}}{\sum_{i,j} f \left( d_{m,n}^k, \epsilon_{m,n}^k \right) v_{y,i,j}} \right] \\
&= \left[ \frac{\sum_{i,j} f \left( d_{m,n}^k, \epsilon_{m,n}^k \right) v_{x,i,j} \epsilon_{m,n}^k}{\sum_{i,j} f \left( d_{m,n}^k, \epsilon_{m,n}^k \right) v_{y,i,j} \epsilon_{m,n}^k} \right]
\end{align*}
\]

(11)

where \(v_{x,i,j}\) and \(v_{y,i,j}\) are horizontal and vertical components of \(v_{i,j}\), respectively. Here, \(v_{m,n}\) and ABPD\(_{m,n}\) are partitioned into four sub-blocks, \(v_{m,n}^k\) and \(\epsilon_{m,n}^k\) individually. \(\epsilon_{m,n}^k\), which is the same as ABPD\(_{m,n}\), is used to roughly measure the interpolation difficulty. That is, if the moving object is not the same in two consecutive decoded frames, the areas where object is distorted should have high ABPD values. In such case, motion smoothing can help to minimize the difference between block boundaries. When the scene consists of simple textures or the scenes between two decoded frames are the same, the weights of adjacent MVs should be decreased since the scene content might be very sensitive to MV adjustment. \(d_{m,n}^k\) is the corresponding Euclidian distance between \(v_{m,n}^k\) and adjacent MVs, \(v_{i,j}\). If the distance is large, which usually happens when motion has sudden change, the corresponding weights should be reduced to reserve sharp object edges. From this discussion, we, therefore, choose \(f\) function as an inverse mapping function for both vector distance and ABPD energy.

Initially, we assign \(v_{m,n}\) to \(v_{m,n}^k\), \(k = 1, 2, 3, 4\). Then, we will set the weight for the centered MV to be one and the weights of neighboring \(v_{m,n}^k\) will be updated individually using (11). We choose sigmoidally shaped function for (11) with two input parameters, \(\epsilon_{m,n}^k\) and \(d_{m,n}^k\) to adaptively adjust the weights for MV averaging. Thus, the inverse mapping function can be then written as follows:

\[
f \left( d_{m,n}^k, Q \left( \epsilon_{m,n}^k \right) \right) = \frac{1}{1 + e^{-d_{m,n}^k - Q(\epsilon_{m,n}^k)}}
\]

(12)

where \(d_{m,n}^k\) is the corresponding distance between \(v_{m,n}^k\) and surrounding MVs, and \(Q(\epsilon_{m,n}^k)\) is the step function of the ABPD energy. In (12), the sigmoidal function opens to the right, so as \(d_{m,n}^k\) increases, the weight value decreases accordingly. \(Q(\epsilon_{m,n}^k)\) is used to decide the center of the sigmoidal function in which the weight value reduces to half. Based on the previous discussion, we should shift the sigmoidal center rightward when the ABPD energy is large. Similarly, as ABPD decreases, the center will be moved leftward until ABPD\(_{m,n}\) = 0 and only same MVs can have nonzero weights, i.e., 1. In Fig. 6, we use four different sigmoidal functions for MV averaging in the experimental simulations. As observed, MVs whose distances are similar within a certain range can have same or similar impacts during vector averaging. Likewise, once the distance goes beyond a certain range, we can also reduce its weight immediately.

In our implementation, we sample the sigmoidal functions with four different \(Q(\epsilon_{m,n}^k)\) values in advance and save these sampled values in look-up tables. Hence, the proposed MVF interpolation method can simply obtain corresponding weights from the tables according to the input \(d_{m,n}^k\) value. The performance gain for the adaptively weighted MV averaging is shown in Fig. 5. As observed, the blockiness artifacts around the area where ABPD energy is high are removed and the object contour such as the face also looks sharp.
V. ADAPTIVE FRAME INTERPOLATION SCHEME FOR OCCLUSION AREAS

In the proposed MV processing, we do not assign any MVs to occlusion areas since the motion is only reliable when appropriate predictions can be found from both forward and backward frames. Hence, if the unreliable MVs of \( L_1 \) have not been corrected until the MV selection process terminated, the MBs that still have nonzero indices in MBMM\(_m\) will be regarded as occlusions. In addition, the MBs whose MVs still cannot be corrected during the MV refinement stage are also considered as occlusions. In order to assist the subsequent frame interpolation for the occlusion areas, an occlusion map (OCM) is created to indicate the location and the range of occlusion:

\[
OCM_{m,n} = \begin{cases} 
1, & \text{if MBMM}_{m,n} \neq 0 \\
2, & \text{MB that has } ABPD > \varepsilon_3 \\
0, & \text{otherwise}
\end{cases}
\]  

(13)

where \( \varepsilon_3 \) is the same as the threshold value in [24]. In OCM, the first type of occlusion often has larger size and the resulting artifacts are more visible, such as deformed structures and the occlusion caused by very large movement. The second type of occlusion is usually the regional occlusion such as the motion boundary and the surrounding MBs of type 1 occlusions. As shown in Fig. 5, we will calculate the occlusion size to determine whether we are capable of recovering the occlusion areas. If the occlusion region is larger than a predefined threshold, \( \varepsilon_{OCM} \), we will skip the interpolation process and repeat the current decoded frame.

Excluding the appearance of new large objects and large-scale object distortion, most occlusion cases are commonly caused by existing moving objects. These occlusions usually occur around the object contour or the frame boundary, so their sizes are often within a reasonable range and we can easily recover them by adaptively selecting the forward or backward prediction. By analyzing the motion distribution around the occlusion area, a prediction reference map can be further created to determine whether forward prediction or backward prediction is better for this occlusion area.

A. Adaptive Frame Interpolation Scheme

To reduce the possible visual artifacts in the bidirectional interpolation scheme, (2) is modified as follows:

\[
f_k(i, j) = w_f \cdot f_{i-1} \left( i + \frac{1}{2} v_x, j + \frac{1}{2} v_y \right) + w_b \cdot f_{i+1} \left( i - \frac{1}{2} v_x, j - \frac{1}{2} v_y \right)
\]  

(14)

where \( w_f \) and \( w_b \) are the weights for forward and backward predictions, respectively. For the identified occlusion areas in OCM, \( w_f \) and \( w_b \) should be adaptively adjusted to obtain the best visual experience.

By analyzing the corrected MVF, we can learn about how these movements cause the occlusion. We again use Fig. 4(b) as an example, in which the upper occlusion (uncovered region) is induced by \( f_{i-1} \), so the better prediction can only be obtained from the backward frame, \( f_{i+1} \). Similarly, the lower occlusion (covered region) can only find correct prediction in
the forward frame, \( f_{t+1} \). That is, for the interpolated frame, the co-located blocks of the moving object’s previous position should have backward predictions, and the co-located blocks of the moving object’s current position should have forward predictions. By examining the MV directions and MV magnitudes, a block-based prediction reference map (PRM) can be further derived to indicate the prediction selection:

\[
PRM_{m,n} = \begin{cases} 
  w_f = \frac{1}{2}, w_b = \frac{1}{2}, & \text{if } OCM_{m,n} = 0, \\
  w_f = 1, w_b = 0, & \text{if } b_{m,n} \text{ of } f_{t-1} \text{ is pointed by MVs} \\
  w_f = 0, w_b = 1, & \text{otherwise}, 
\end{cases}
\]

In our implementation, we examine the neighboring MVs for each occlusion area in the forward direction. In OCM, adjacent occlusion blocks are considered as one occlusion area. If co-located blocks of the occlusion region are not pointed by any neighboring MVs, this occlusion region will only have backward predictions for the interpolation. Otherwise, we assign forward predictions. Same criteria can also be used to explain the occlusion occurrence on the frame boundary. If a frame boundary MB does not have any MVs pointing to it, this means that it is the initial block of the whole movement. In such case, the backward prediction should be considered.

In Fig. 7, we use a synthesized figure to further demonstrate this occlusion process. In Fig. 7(a), there are three moving objects, which are indicated using different structure patterns, and each grid represents \(8 \times 8\) block size. As observed, the received MVF has been corrected for nonocclusion areas, but for the occlusion areas, which is indicated by gray color, there is no motion since it is very difficult to find appropriate predictions bidirectionally. Based on the analysis of the motion distribution in the nonocclusion areas, the resulting prediction map is shown in Fig. 7(b), where red color, blue color, and green color denote using backward predictions, using forward predictions, using bidirectional predictions, respectively. Since the proposed interpolation scheme uses three different prediction modes for the interpolated frame, to make the pixel values do not change abruptly on the boundary between forward and backward predictions, we apply low-pass filtering on the boundary pixels afterward.

**B. Motion Vector Processing in Occlusion Areas**

In Fig. 5, as the object major motions and detailed motions are determined, we start to assign neighboring corrected MVs to the type 1 occlusion regions. Conventional methods usually assume only one or two movements around the occlusion region, but various movements may occur once the occlusion size is very large or more than one moving objects in the same area. Therefore, it is very difficult to tell which movements the occlusion should have when multiple movements occur. In order to avoid the occlusion motion distribution being dominated by either motion, we spirally assign neighboring MVs to occlusion regions. Along the spiral trace, we start from the left-top and clockwise assign the MV to each occlusion using VMF based on the \(8 \times 8\) block size. If there are no MVs available for VMF processing, we will check this block in the future iteration. The assignment stops until all occlusion sub-blocks have MVs. In this way, once the occlusion size is large, we can still ensure these interpolated blocks can follow the neighboring movement and the occlusion can also be evenly divided by surrounding major movements.

We use Fig. 7(c) for the further demonstration, in which the occlusion blocks have been assigned MVs unidirectionally from the neighboring corrected MVs. In Fig. 7(d), the MVF has been further interpolated using the proposed weighted MV averaging filter, and the resulting MVF becomes smoother and will be used for the frame interpolation. Please note that this MV processing does not include the type 2 occlusions since they already have nearly correct motions. The reason why these MVs still have high ABPD energies after the MV refinement is that they may be located on the edges of the occlusion area or the occlusion is so small that we still can find correct motion. In Fig. 5, the interpolated result using the proposed adaptive interpolation scheme...
Fig. 10. Interpolated results of frame 20 of FOOTBALL sequence using (a) original frame, (b) direct interpolation (PSNR: 22.60 dB, SSIM: 0.6960), (c) bidirectional BMA (PSNR: 22.55 dB, SSIM: 0.7237), (d) the method in [24] (PSNR: 22.70 dB, SSIM: 0.7209), (e) the proposed MV processing with the bidirectional MCFI scheme (PSNR: 22.99 dB, SSIM: 0.7405), and (f) the proposed MV processing method with the proposed MCFI scheme (PSNR: 22.28 dB, SSIM: 0.7172), respectively.

Fig. 11. Interpolated results of frame 52 of FORMULA 1 sequence using (a) original frame, (b) direct interpolation (PSNR: 30.53 dB, SSIM: 0.9430), (c) bidirectional BMA (PSNR: 26.83 dB, SSIM: 0.8278), (d) the method in [24] (PSNR: 29.02 dB, SSIM: 0.9520), (e) the proposed MV processing with the bidirectional MCFI scheme (PSNR: 33.52 dB, SSIM: 0.9681), and (f) the proposed MV processing method with the proposed MCFI scheme (PSNR: 32.66 dB, SSIM: 0.9632), respectively.

looks much better than the general bidirectional scheme. The occlusion artifacts around the sleeves, the shoulder, and the name tag are removed.

VI. SIMULATIONS

In this section, simulation results are demonstrated to evaluate the performance of the proposed method. We compare our method with direct interpolation, VMF, MV selection in [28], the multistage MV processing method in [24], bidirectional BMA with consideration on MV correlation based on 8 × 8 and 16 × 16 block size, respectively, and the proposed MV processing method with forward prediction only(Fw-Pred), backward prediction only(Bw-Pred), bidirectional(Bi-Pred), and adaptively selective prediction(Adaptive-Pred). For the bidirectional BMA, we set the motion search range to be −16 to 16 and the estimated MV can be written as

$$v_{mn} = \arg \min_{v \in \mathcal{V}} \left( \text{SAD}(v) + \lambda \sum_{v_N} |v - v_N| \right)$$

where SAD(v) is the sum of absolute difference of forward and backward predictions using v · v_N are the neighboring causal MVs of b_{mn}. We set λ to be 40, which is the multiplier for the influence of MV difference. Six video sequences, FOREMAN, FORMULA 1, WALK, FAST FOOD, STEPHAN, and FOOTBALL of CIF frame resolution are used with the original frame rate of 30 frames per second (fps). They are encoded using H.264 with even frames skipped to generate video bitstreams of 15 fps. The skipped frames are interpolated at the decoder for the evaluation. The rate-distortion control function and the RD optimization function are enabled during the encoding process. The averaged bit rates for these test sequences are set to be 384 kbps or 512 kbps according to the contents.

The visual comparisons are presented in Figs. 8–11. In Fig. 8, blockiness can easily be seen in the interpolated results using direct interpolation, bidirectional BMA, and the method in [24]. In Fig. 8(e) and (f), since the received MVF has been corrected and been interpolated by the proposed adaptively weighted MV averaging, most of blockiness artifacts are removed. For the mouth and nose areas, there are no matched bidirectional predictions from two consecutive decoded frames, so a lot of ghost artifacts can be observed around the face as shown in Fig. 8(b)–(e). By adopting the proposed adaptive frame interpolation scheme,
these ghost artifacts are completely removed. Although our interpolated frame is slightly different from the original one, the mouth and the nose areas still follow the motion trajectory from frame 93 to frame 95.

In addition to using PSNR to measure the signal fidelity between the interpolated result and its original frame, we adopt another objective measurement, structure similarity (SSIM), for quality assessment [29], which was also used in [23], [13]. In Fig. 8, since the unreliable MVs with high residual energies and the unreliable low-correlated MVs are corrected, our result using the general bidirectional MCFI scheme performs the best. The reason why the final result, Fig. 8(f), performs slightly worse than Fig. 8(e) is that the proposed interpolation scheme may increase the difference between the original frame and interpolated frame in the occlusion regions and object deformed areas.

In Figs. 9(f) and 10(f), we can observe that the ghost effects are greatly reduced compared to other results using conventional methods. The shapes of moving objects are sharper than bidirectional scheme and the boundary between moving objects looks smoother. In addition, conventional methods cannot maintain the wall and the shirt areas well due to the appearance of the similar patterns. Even when the MVs are reestimated with consideration on correlation, the wall still cannot be reconstructed well as shown in Fig. 9(c). By using the proposed correlation-based MV processing method, these irregular received MVs are corrected and those dislocation artifacts do not show in our interpolated result as shown in Fig. 9(e) and (f). Fig. 10 shows that even when the frame contains various movements, the proposed correlation-based MV processing method and the adaptive frame interpolation scheme still can perform well and has clearer object contours than other conventional methods. In Fig. 11, the occlusion occurs in the cross section of the white lines and the number area. The proposed MCFI scheme removes the ghost artifacts by adjusting the weights of predictions in this region, so the ghost artifacts do not appear in Fig. 11(f). Compared to Fig. 11(d), since the MV correlation and the MV reliability are considered during the MVF interpolation, the static numbers are not blurry in Fig. 11(e) and (f).

For better comparison, we list averaged PSNR and SSIM values for these six video sequences in Tables I and II. As observed, our SSIM performance using bidirectional MCFI scheme is consistently better than others but PSNR performance performs worse in FORMULA 1. Comparing the proposed MV processing method to the method in [24], we perform much better in FORMULA 1, FAST FOOD, and FOREMAN than other video sequences. This is because these three sequences contain stripe patterns or patterns with similar density. Since the proposed MV processing method skips frames if the occlusion size is greater than $E_{occlusion}$, for these conventional methods, the skipped frames are also not counted in the PSNR and SSIM computation. The frame rates of the interpolated videos are 29.44, 29.18, 28.50, 29.50, 28.30, and 25.50 for WALK, FORMULA 1, FAST FOOD, FORMEN, STEPHAN, and FOOTBALL, respectively. More interpolated results can be viewed at http://videoprocessing.ucsd.edu/~aihuang/jnl2008.htm.

VII. CONCLUSION

Based on our previous work, we further propose using MV correlation to discover unreliable MVs that cannot be detected using the received residual energies. Considering those similar irregular MVs should belong to the same object, we, therefore, create a MB merging map for the correlation-based MV selection process. In this way, more pixels can be referenced for the MV selection so that correct motion can be found even in smooth areas or repeat pattern areas. In addition, to solve the occlusion issues, we correct unreliable MVs recursively based on the different levels of the ABPD energy so the occlusion areas or the deformed objects can be discovered effectively. These identified occlusion areas will be further analyzed and be assigned appropriate MVs from neighboring corrected MVs. After the MVF has been determined, a prediction reference map will be created to denote which interpolation modes for the current occlusion should be employed. As a result, the interpolated results using the proposed method can remove most of the ghost artifacts and obtain clearer object contours.
REFERENCES


