Video Compression Based on Wavelet Transform and DBMA with Motion Compensation

Zhengxin Hou[1], Baochen Jiang[1,2], Yupeng Cao[1], Aiping Yang[1*], and Chengyou Wang[2]
[1] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, P. R. China
[2] School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, P. R. China
Email: zhengxinhou@163.com, jbc@sdu.edu.cn, cyp49025877@163.com, yangaiping@tju.edu.cn, wangchengyou@sdu.edu.cn

Abstract—In video compression, it can reduce the redundant information efficiently by motion estimation (ME) and motion compensation (MC), and less code is used to encode as much information as possible. In this paper, I frame encoding adopts wavelet transform and set partitioning in hierarchical trees (SPIHT) algorithm; for P frames, each frame sets the reconstructed frame of its previous frame as a reference frame, and then P frames proceed to code with ME and MC. In the step of ME, trading off between accuracy and computational complexity, adaptive fast search (AFS) algorithm combining with nodal search-based deformable block matching algorithm (NS-DBMA) is adopted to search for the matched block; and then in the MC process, wavelet transform combining with zerotree entropy (ZTE) algorithm is adopted according to the characteristics of residual image data. Meanwhile rate control is carried out in ZTE algorithm. Experimental results show that the proposed algorithm performs well for video sequences with complex motion and rich details, and reduces blocking artifacts obviously.

Index Terms—video compression, motion estimation (ME), motion compensation (MC), wavelet transform, nodal search-based deformable block matching algorithm (NS-DBMA), adaptive fast search (AFS)

I. Introduction

In frame sequences, the adjacent frame is correlated with each other to some extent. Dividing each frame into some blocks, each block can get its most matched block in its previous reconstructed frame through motion estimation (ME), so the motion vector (MV) is got. In many occasions, motion compensation (MC) is needed to make up the prediction difference caused by ME so as to get the prediction frame, which is as close as the original one. Therefore, the better the ME algorithm is, the more accurate the MV is. As a consequence, the amount of data that will be transferred will reduce obviously and the compression ratio of video can be improved.

The ME process is to find the block in reference frame that matches most closely the current block by block matching algorithm (BMA). In all block matching algorithms, the simplest and most efficient method is full search (FS) algorithm, while it is most useless since its huge computational amount. The FS algorithm searches the matching block in the reference frame pixel by pixel. If a frame in CIF format is divided into blocks of 16×16 and each block is searched ranging from 0 to 15 pixels, the total computational amount reaches 9.742×10^7. In addition, each matching operation includes a subtraction, an absolute value operation and an addition, so it consumes a large number of computational resources. When comparing different algorithms, the FS is seen as an ideal situation theoretically. That is, the closer the quality is to the FS, the better the algorithm is. To reduce the amount of computation and keep the search accuracy, many fast ME algorithms are proposed, such as three-step search [1], four-step search [2], block-based gradient descent search [3], diamond search [4], enhanced predictive diamond zonal search [5], and so on.

The above ME algorithms are mostly based on translational motion model, while it isn't fully consistent with the objective fact of objects motion. In many cases, objects usually have non-translational motion. Then if BMA is used to ME, it'll lead to inaccuracy that is not beneficial for video compression. In order to improve the performances of BMA, some famous algorithms are proposed, such as overlapped block motion compensation [6], [7], loop filter [8], unsymmetrical-cross multi-hexagon-grid search (UMHexagonS) [9], and so on. These algorithms do improve the performance of BMA to some extent, while they can't totally solve the prediction defects of non-translational motion. In this paper, nodal search-based deformable block matching algorithm (NS-DBMA) [10] is used to solve those problems. The NS-DBMA is also a ME method based on block matching. It uses bilinear motion model which is more complicated than translational model to represent the motion of object. In complicated motion cases, the NS-DBMA can further improve the accuracy of ME to make up the defect of BMA.

For better compression performance and minimizing blocking artifacts due to the existing video coding framework, the wavelet transform is used to I frame, and then the transformed data of I frame are coded with set partitioning in hierarchical trees (SPIHT) [11] algorithm. As a consequence, the video with good subjective effect can be achieved and no blocking artifact is felt at low bit

Manuscript received June 7, 2013; revised July 9, 2013.
This work was supported by the National Natural Science Foundation of China under Grant No. 61201371 and No. 61002027.
Corresponding author email: yangaiping@tju.edu.cn.
doi:10.12720/jcm.8.7.440-448

rate. For P frames, compared with the conventional BMA, which has obvious disadvantage to non-translational motion, adaptive fast search (AFS) [12] combining with NS-DBMA is adopted as an ideal method for ME. Firstly, AFS is used to get the MV of object motion; then if the precision can't meet the demand, NS-DBMA will be applied to ME and it can efficiently improve the precision of prediction. In the process of MC, wavelet transform is used to residual image at the beginning. Then zerotree entropy (ZTE) [13] is adopted to code the wavelet coefficients according to the characteristics of themselves. In addition, the ME of fractional precision adopts the interpolation method based on all phase DCT (APDCT) [14], and rate control is added into the quantization process of ZTE. To sum up, the video coding framework proposed in the paper is shown in Fig. 1.

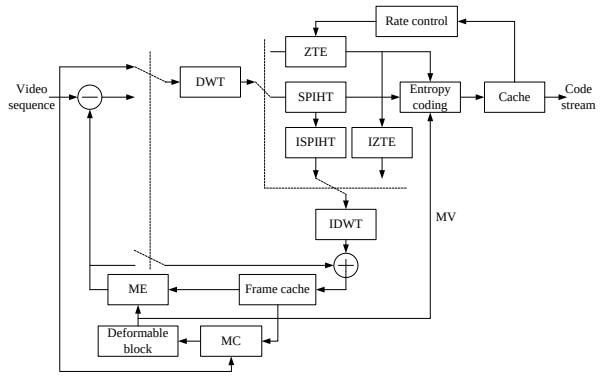


Figure 1. Proposed video coding framework.

The rest of this paper is organized as follows. Section II explains the code scheme of I frame. Section III describes the AFS and NS-DBMA used in ME, and introduces the APDCT interpolation used in NS-DBMA. Then Section IV presents the ZTE algorithm with rate control used in MC. Finally, experimental results and conclusion are given in Sections V and VI respectively.

II. I Frame Encoding Scheme

A video encoding unit consists of an I frame and some P frames. In the unit, the encoding process of I frame is the same with still color image and the reconstructed image is treated as the reference frame of next frame. For P frames, each frame uses the reconstructed image of its previous frame as reference when performing ME. Considering that the reconstructed frame is not totally consistent with the original one, the precision of ME will be affected and the following residual data must be considerable, resulting in the quality of reconstructed frame down. So it is obvious that the coding of I frame affects not only the reconstruction of itself, but also that of the subsequent P frames. In the following, two ways of I frame encoding are considered: DCT with run length coding and wavelet transform with SPIHT. If DCT is adopted in I frame, it will lead to obvious blocking artifacts at low bit rate, and then P frames will be affected by ME, lowering the performance of prediction and coding of P frames. By comparison, the wavelet method transforms the entire frame and it can concentrate the energy of image effectively. The blocking artifact is removed after decoding. Additionally, SPIHT is able to organize the wavelet coefficients efficiently, improving the efficiency of encoding. Therefore, wavelet transform with SPIHT is adopted for I frame encoding in this paper.

III. Motion Estimation

Motion estimation is the core of this paper and the block of 8×8 is adopted for ME. Firstly, AFS is used for each block and residual difference is got. Then, comparing the residual difference with a threshold value, if the difference value is bigger than the threshold, NS-DBMA is used for this block. Eventually, the optimal matched block is got. Considering that the value of quarter pixel is used in NS-DBMA, the fraction position of each frame needs interpolation. In the following parts, ME is introduced in detail.

A. Adaptive Fast Search (AFS)

UMHexagonS is accepted as fast ME with integer pixel by the video compression standard H.264 [15], [16], [17]. Compared with fast FS algorithm, UMHexagonS reduces computational amount by 90% and retains good rate distortion performance.

In this paper, the improved UMHexagonS called adaptive fast search (AFS) is used for ME. Based on UMHexagonS, the AFS makes improvement in two aspects: searching center prediction and the premature termination algorithm. It not only retains the quality of prediction, but also speeds up the prediction. Under kinds of conditions, it can save half of computational amount and improve the comprehensive performance obviously.

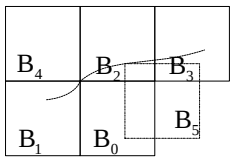


Figure 2. Current block and its adjacent blocks.

As shown in Fig. 2, B_0 is the block to be predicted. B_1, B_2, B_3 and B_4 are spatial adjacent blocks of B_0 in the same frame. And B_5 is the block, located in the previous frame, with the same spatial position of B_0. In UMHexagonS, the motion of these adjacent blocks is thought to be strong correlated with B_0. So the searching center of B_0 gets from these blocks' MVs. While over a large number of experiments, a phenomenon is found that B_1 and B_2 are most correlated with B_0, while B_5 is least correlated with B_0. This means that the MVs of B_3, B_4 and B_5 can't improve the prediction precision of B_0's MV, and even import some inaccuracy. So only the MVs of B_1 and B_2 are used in AFS to determine the

searching center. That is, the MV of the block which matches better with B_0 between B_1 and B_2 is set as the initial MV of B_0.

In UMHexagonS, different threshold TH is set to achieve premature termination. And it controls different searching procedure. Different threshold can trade off between the searching precision and computational amount. Setting the sum of absolute difference (SAD) as matching criterion, the optimal matching difference of FS for the block to predict is $\mathrm{SAD}_{\mathrm{best}}$, and the optimal matching difference of current algorithm is SAD_i. So these values satisfy:

$$\mathrm{SAD}_{\mathrm{best}} \le \mathrm{SAD}_i < TH \tag{1}$$

When TH is smaller than $\mathrm{SAD}_{\mathrm{best}}$, search won't achieve premature termination. In such a case, the accuracy is adequate, while the computational amount won't reduce. If TH is far bigger than $\mathrm{SAD}_{\mathrm{best}}$, the search will end quickly. But the accuracy is not so good that the quality of prediction will be bad. To sum up, it is a good choice that TH is a little bigger than $\mathrm{SAD}_{\mathrm{best}}$.

The improved UMHexagonS algorithm is shown as follows [10].

- Only the MVs of B_1 and B_2 are used in this part. To improve the prediction precision, the rectangular lattice searching around the searching center within one pixel is added.
- Skip the unsymmetrical-cross search, and use the Big_Diamond search in UMHexagonS to search. While in the Big_Diamond, the searching area is simplified from 5×5 to 3×3.

According to the above analysis of threshold value, AFS defines the empirical threshold TH_{const} and adoptive threshold TH_{pred} of spatial adjacent blocks respectively. TH_{pred} is the smaller value between the prediction difference J_{B_1} and J_{B_2} for the spatial adjacent blocks. Then select the larger value between TH_{const} and TH_{pred}, shown as (2) and (3), as the premature termination value TH_{last}.

$$TH_{\mathrm{pred}} = \min(J_{B_1}, J_{B_2}) \tag{2}$$

$$TH_{\mathrm{last}} = \max(TH_{\mathrm{const}}, TH_{\mathrm{pred}}) \tag{3}$$

At the same time, three different search lattices' thresholds are defined as follows:

$$TH_{\mathrm{BD}} = TH_{\mathrm{last}} + 2 \cdot SIZE \tag{4}$$

$$TH_{\mathrm{HX}} = TH_{\mathrm{last}} + SIZE \tag{5}$$

$$TH_{\mathrm{SD}} = TH_{\mathrm{last}} \tag{6}$$

where $SIZE$ represents the size of block to predict; TH_{BD}, TH_{HX} and TH_{SD} are premature termination thresholds of the search lattices as Big_Diamond, Hexagon and Small_Diamond. During searching, if prediction difference is smaller than the corresponding threshold, skip this step and go to next step.

B. Nodal Search-based Deformable Block Matching Algorithm (NS-DBMA)

Suppose that the MV obtained by AFS is $\boldsymbol{V}_i^0 = [V_{x,i}^0, V_{y,i}^0]$, $(i = 1,2,3,4)$, where i represents the control node. Suppose that the searching range is R, so the initial step-size is $R_0 = \lfloor R/2 \rfloor$. Nine points are examined in every round, including the current point and the responding points in eight different directions. Each time when examining a point, interpolation is needed. Then the difference is calculated according to the SAD criterion. The point whose difference is the least is regarded as the optimal one. Then the optimal point is set as the starting point of next round. Modify its MV and the search of subsequent nodes is based on the modified MV. After updating all nodes, halve the search steps and perform another search with same method of the previous round. Until search step turns into one pixel, the search ends. And the point got in the last round is the final matched one.

Suppose that the optimal MV of each node is denoted as $[V_{x,i}^{\mathrm{op}}, V_{y,i}^{\mathrm{op}}]$, then we have

$$R = R_0 + R_1 + \cdots + R_{N-1},\ N = \lfloor \log_2 R \rfloor \tag{7}$$

$$\begin{cases} V_{x,i}^{\mathrm{op}} = V_{x,i}^0 + V_{x,i}^1 + \cdots + V_{x,i}^N = V_{x,i}^0 + \sum_{j=1}^{N} s_{x,i}^j R_j,\ s_{x,i}^j = \{-1,0,1\} \\ V_{y,i}^{\mathrm{op}} = V_{y,i}^0 + V_{y,i}^1 + \cdots + V_{y,i}^N = V_{y,i}^0 + \sum_{j=1}^{N} s_{y,i}^j R_j,\ s_{y,i}^j = \{-1,0,1\} \end{cases}. \tag{8}$$

In each round, the searching matrix to be examined is defined as follows:

$$Q_{3\times3}^j = \begin{Bmatrix} (-R_j,-R_j) & (0,-R_j) & (R_j,-R_j) \\ (-R_j,0) & (0,0) & (R_j,0) \\ (-R_j,R_j) & (0,R_j) & (R_j,R_j) \end{Bmatrix} \tag{9}$$

The search process of each node is shown as follows:

$$\begin{aligned} V_1^j &= V_1^{j-1} + \arg\left\{ Q_{3\times3}^j \middle| \min \mathrm{SAD}(V_1^j, V_2^{j-1}, V_3^{j-1}, V_4^{j-1}) \right\} \\ V_2^j &= V_2^{j-1} + \arg\left\{ Q_{3\times3}^j \middle| \min \mathrm{SAD}(V_1^j, V_2^j, V_3^{j-1}, V_4^{j-1}) \right\} \\ V_3^j &= V_3^{j-1} + \arg\left\{ Q_{3\times3}^j \middle| \min \mathrm{SAD}(V_1^j, V_2^j, V_3^j, V_4^{j-1}) \right\} \\ V_4^j &= V_4^{j-1} + \arg\left\{ Q_{3\times3}^j \middle| \min \mathrm{SAD}(V_1^j, V_2^j, V_3^j, V_4^j) \right\} \end{aligned} \tag{10}$$

Four individual searches are shown in the above instructions. Each time nine points are searched and each point needs to be calculated for the interpolation and difference.

A simple flow chart describing the NS-DBMA is shown in Fig. 3.

C. Fraction Position Interpolation

Since the continuity of actual movement of objects, the MVs of some blocks are not integers in ME processing. While fraction pixel doesn't exist, then the prediction

difference produces. It's necessary to get the value of fraction pixel using interpolation. The APDCT interpolation [14] is adopted in this paper.

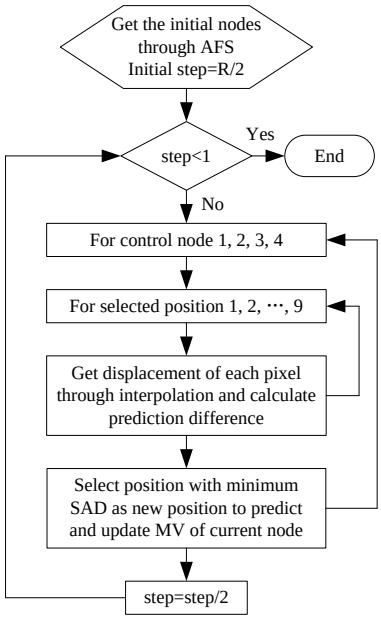


Figure 3. Flow chart of NS-DBMA.

To achieve the APDCT interpolation easily, the interpolation kernel can be designed as an interpolation matrix $\boldsymbol{P}$. For the quarter-pixel interpolation of $N=4$ [14], [18], the matrix $\boldsymbol{P}$ can be expressed as (11).

$$\boldsymbol{P}_{3\times 6}=\begin{bmatrix} 0.0152 & -0.0890 & 0.8491 & 0.2768 & -0.0650 & 0.0128 \\ 0.0196 & -0.1052 & 0.5856 & 0.5856 & -0.1052 & 0.0196 \\ 0.0128 & -0.0650 & 0.2768 & 0.8491 & -0.0890 & 0.0152 \end{bmatrix}. \quad (11)$$

In matrix $\boldsymbol{P}$, its first row is the filter coefficients of one quarter pixel, its second row is the filter coefficients of half pixel and third row is the filter coefficients of three quarters pixel. Firstly, a one-dimensional horizontal filtering is conducted. Using the corresponding row filter of $\boldsymbol{P}$ to interpolate in the horizontal direction, and reference points are the known integer pixel. Then all the values of quarter-pixel interpolation in a row can be got one time. Until the operation of interpolation for rows finished, one-dimensional vertical column filtering is conducted in the following. This time the corresponding column filter of matrix $\boldsymbol{P}$ is used to interpolate in the vertical direction. Then the interpolation of quarter-pixel precision for the whole frame is got.

IV. Motion Compensation

For P frames, the precision is limited by ME and it's hard to get new changes of the frame to predict by ME. Under this condition, MC is an indispensable step to reconstruct the frame sequences exactly. The conventional MC with DCT firstly divides residual frame into blocks and has DCT transform for each block. Then the process for the transformed coefficients is completely same with that of grayscale image. The quality of reconstructed frame is acceptable for visual effect, while the code rate is high and isn't easy to control. Considering the shortcomings of DCT, the wavelet transform combining with ZTE coding for MC is used in this paper. In this method, the whole residual image is transformed with wavelet method. Then the wavelet coefficients are rearranged and quantized. At last, the coefficients of each block are encoded using ZTE algorithm. The following part introduces the algorithm used in MC.

A. ZTE Encoding

Zerotree root (ZTR) is a very important concept in ZTE coding. ZTR is a zero node, which locates in the lowest resolution sub-band of zero-tree. It satisfies the following requirements. First, this node is a zero node; second, its parent node isn't a zero node; third, this node doesn't locate in the highest resolution sub-band; last, all its child nodes and grandchild nodes are zero. Apparently, ZTR represents all the information of its corresponding zero-tree. If the position of ZTR is known, the position of all zero nodes on the zero-tree is known. The zero nodes that don't belong to any zero-tree are called isolate zero nodes. If nonzero nodes exist in the posterity of a zero-node, the zero-node is an isolate zero node. The residual frame which is got from ME is transformed by the wavelet with 3 layers. Then rearrange the wavelet coefficients using the method shown in Fig. 4. So the wavelet coefficients in different layer sub-bands which represent the same area of image are put into the same wavelet block to establish a direct link.

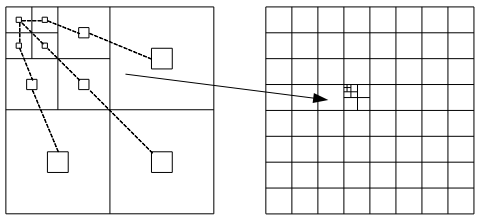


Figure 4. Wavelet block rearrange schema.

Then the wavelet coefficients block are quantized. Any efficient scheme can be used in this progress.

After quantization, the wavelet block is scanned from the lowest resolution sub-band to the highest one using ZTE algorithm. Then the essential node is distributed a symbol, which is one of the follows: zerotree root (ZTR), valued zerotree root (VZTR) and value (V).

ZTR represents that the coefficient is the root of zero-tree. In addition, its child nodes, grandchild nodes and the root coefficient are all zeros. In ZTE algorithm, further scanning is not needed.

VZTR represents that the coefficient's child nodes and grandchild nodes are all zeros, while the root coefficient itself isn't zero. In this situation, further scanning is not needed too.

V represents that no matter the coefficient is zero or nonzero, there is nonzero coefficient in its child nodes or grandchild nodes. In this case, further scanning is needed.

The ZTE algorithm simultaneously outputs three symbol sets: identification symbol set, the corresponding coefficients set of VZTR and the corresponding coefficients set of value. For any leaf-node without any child node being scanned, identification symbol is not needed to identify this leaf-node, achieving saving codes for identifiers. For these leaf-nodes, the corresponding coefficient values of them are put into the Value set to encode.

B. Quantization and Rate Control

As ZTE algorithm doesn't have embedded feature and can't control the code rate precisely, the quantization matrix becomes the most important factor that affects the efficiency of coding. If the value of quantization matrix is too small, it may not improve the subjective quality, while it will result in the increasing of nonzero value for the residual data. If the value of quantization matrix is too large, then many details in the image will be lost and the image becomes blurred so that it doesn't reach the goal of MC. Considering the circumstances above, adoptive quantization matrix varied with the complexity of image is very essential. If the image is with high quality, the values of quantization matrix are increased, or decreased. Then the quality and the consuming of code rate are balanced. Smooth allocation of rate makes the quality of video stable and the unacceptable situation for video sequences will not appear.

A simple method is used to adjust the code rate in this paper. Firstly, according to the whole code rates of video series and the encoded I frame and P frames, residual rates are calculated. Then evenly distribute to the remaining P frames as the plan rates of current frame. Supposing the plan rates of current frame is M, quantize and encode with the current quantization matrix Q_i. If the coded rates $m_i > M$, it represents that current frame is with a high quality and the quantization matrix should decrease a little. So quantization matrix for the next frame should be $Q_{i+1} = Q_i + Q'$, or $Q_{i+1} = Q_i - Q'$, where Q' is the adjusted matrix. At the same time, update the plan rates of rest P frames. Repeat these steps until encoding ends. So it can ensure that the code rate of each frame is close to the plan rate realizing the goal of controlling code rate. The initial quantization matrix Q_0 and Q' are shown as follows.

$$Q_0 = \begin{bmatrix} 4 & 4 & 8 & 8 & 16 & 16 & 16 & 16 \\ 4 & 4 & 8 & 8 & 16 & 16 & 16 & 16 \\ 8 & 8 & 8 & 8 & 16 & 16 & 16 & 16 \\ 8 & 8 & 8 & 8 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \end{bmatrix}, \quad Q' = \frac{1}{4} Q_0 \qquad (12)$$

When the receiving terminal receives the residual data encoded by ZTE algorithm, it decodes the received code in a completely opposite order with the encoding terminal. Adding the frame data got from ME, and the frame of MC is got.

V. Experimental Results

All experimental results of the proposed algorithm in this paper are achieved by C language in the environment of VC++6.0. The test video sequences are in CIF format, with the size of 352×288, 30 frames per second (fps) and the interval between two I frames is 15 frames.

A. Effect of Different I Frame Encoding Schemes on P Frames

To compare the effect of different I frame encoding schemes on P frames, the "CDF 9/7" wavelet transform (also know as "bior4.4" in MATLAB) combining with SPIHT algorithm and DCT transform combining with run-length coding are used in I frame encoding. In addition, to avoid importing some other factors, P frames adopt AFS for ME and omit the procedure of MC. Now, I frame adopts the mentioned encoding schemes above and test the effect on the first 8 frames of Foreman and Mobile sequences. The experimental curves are shown in Fig. 5.

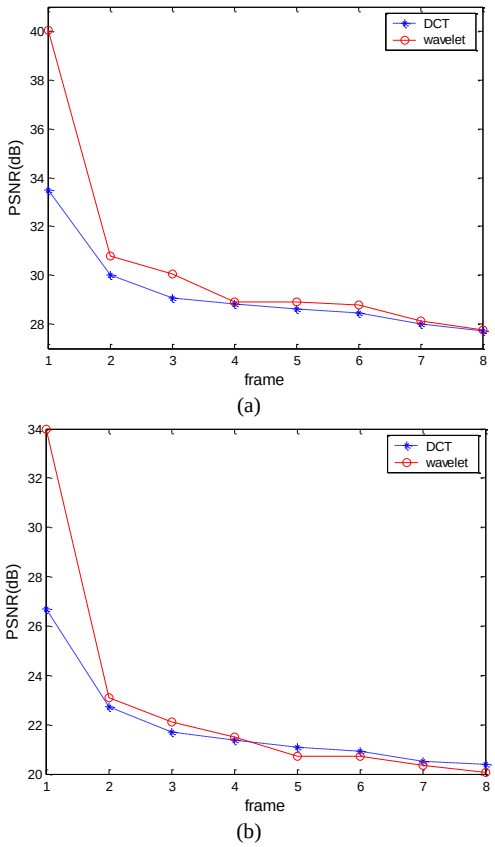


Figure 5. Effect of I frame on P frames: (a) Foreman, 1.0bpp, (b) Mobile, 2.0bpp.

As shown in Fig. 5, the quality of I frame affects that of P frames obviously for the video sequences with less

intense motion. While for the video sequences with complex or intense motion, the quality of I frame has little effect on P frames. To sum up, I frame affects the subsequent two or three frames obviously. And as the increasing of the distance between I frame and P frame, the effect decreases step by step.

(a) (b)

Figure 6. Performance of the first frame with different encoding scheme: (a) DCT+run length coding, PSNR=33.50dB, 23KB, (b) Wavelet+SPIHT, PSNR=40.03dB, 16KB.

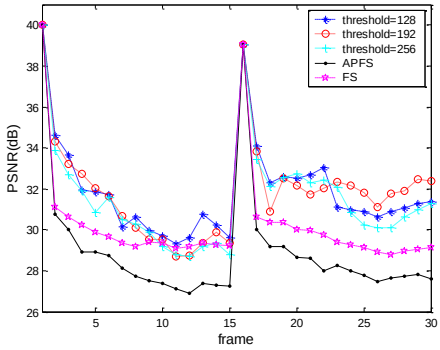
Comparing with DCT encoding, the wavelet coding can get a reconstructed image with higher quality at lower bit rate. This can lay a good foundation for the P-frame's ME. More importantly, wavelet transform solves the blocking artifacts well and improves the visual effect efficiently. The two different encoding methods are used on I frame. Fig. 6 shows the effect of first reconstructed frame for Foreman sequence. The decomposition layer is 4 and bit rate is 1.0 bits/pixel (bpp) in wavelet transform.

B. Comparison of the Encoding Schemes for ME

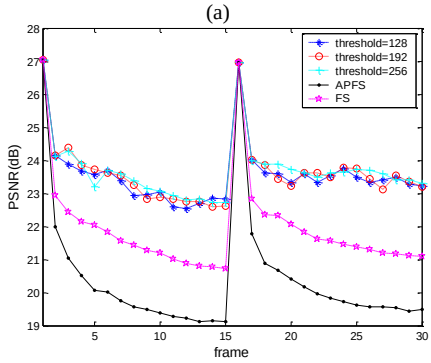
The 8×8 block is adopted for ME in this paper. For a block, AFS is first applied. If the precision doesn't meet the demand, then NS-DBMA is applied to ME. In this way, the precision increases obviously without increasing the amount of calculation and MV simultaneously. The fixed threshold T_{DBMA} will determine whether the NS-DBMA is used or not.

Six video sequences are tested in this paper and they are Foreman, Mobile, Bus, Football, Garden, and Stefan sequences. A variety of movements are involved in these sequences. I frame is inserted into the sequence every 15 frames and the rest are P frames. The rate of frame is 30fps. All the sequences are 30 frames. And then compare the quality of matching block by PSNR of luminance component. The complexity of calculation is measured by the average number of deformable search blocks (Pts_avg). I frame adopts 4 layers wavelet decomposition and SPIHT coding, and the bit rate is 1.0bpp. For P frames, the range of search is 15 pixels in four corresponding directions. Moreover, the BMA and FS algorithm are both with integer-pixel precision, while the NS-DBMA is with quarter-pixel precision.

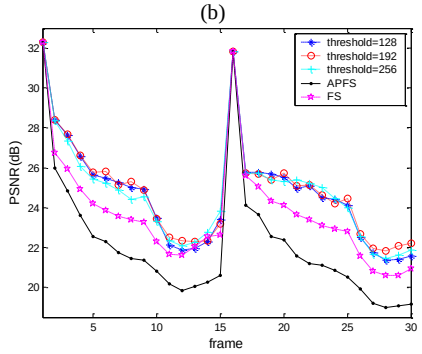
As shown in Fig. 7, the NS-DBMA does improve the precision of ME efficiently, especially for the sequence with complex motion, such as Mobile and Bus. The quality of recovery significantly improves, even is better than FS algorithm. So it's obvious that the motion model based on bilinear transform reflects the actual movement of objects better than the translational motion model. For the sequence with rich details, such as Garden and Stefan, NS-DBMA also performs well. While for the sequence with intense motion, such as Football, NS-DBMA doesn't have obvious advantages. This implies that NS-DBMA has a far better performance for deformation movement and zoom motion than AFS in prediction, while has limited increase in the prediction of translational motion. In addition, NS-DBMA also reduces the blocking artifacts raised by AFS, making the image smooth. Then the subjective feeling improves obviously. Fig. 8 shows the comparison of the seventh reconstructed frame from Foreman before and after using NS-DBMA.



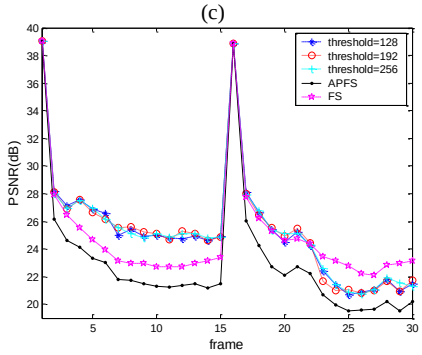
(a)



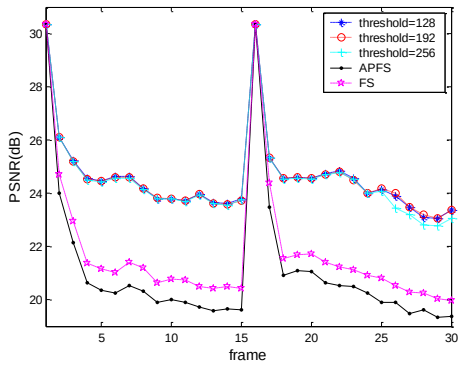
(b)



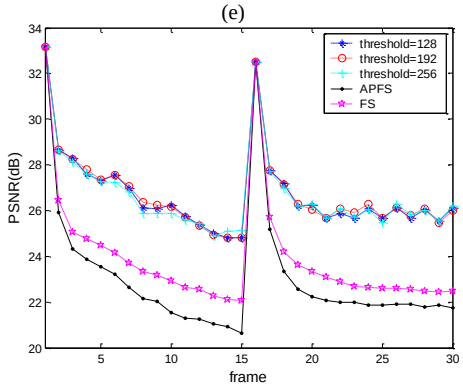
(c)



(d)



(e)



(f)

Figure 7. Performance of AFS, FS and NS-DBMA with different thresholds for different video sequences: (a) Foreman, (b) Mobile, (c) Bus, (d) Football, (e) Garden, (f) Stefan.

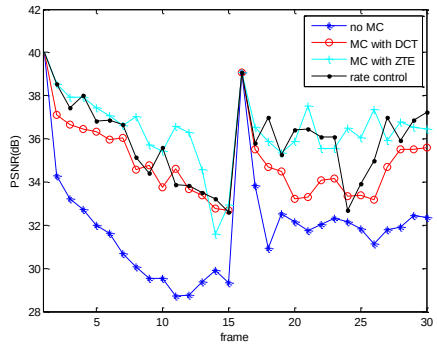
(a) (b)

Figure 8. Comparison of the seventh frame before and after using NS-DBMA: (a) Before using NS-DBMA, (b) After using NS-DBMA.

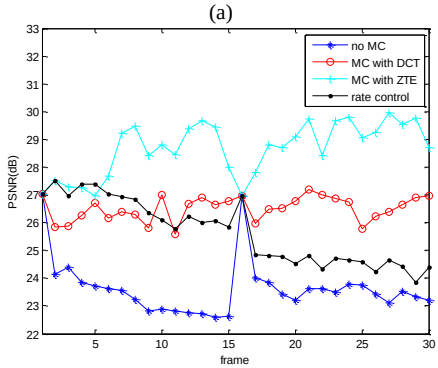
C. Comparison of Different MC Schemes

The precision of ME is limited. It is hard to get the new changes in the predicting frame by ME. To reconstruct the frame sequence accurately and reduce the difference of ME, MC is an essential step.

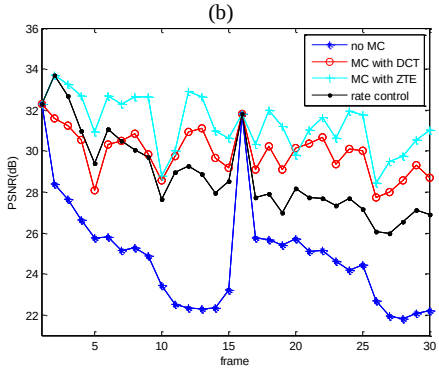
The Foreman, Mobile, Bus, Football, Garden and Stefan sequences are tested in this part. They are compared in four conditions: no MC, MC with DCT, MC with ZTE and MC with rate control. I frame adopts 4 layers wavelet transform and SPIHT coding. The bit rate is 1.0bpp. Then P frames adopt AFS combining with NS-DBMA. T_{DBMA} is set to 192. And the search range is 15 pixels in four corresponding directions. The whole codes in rate control are equal to the ones in DCT transform combining with run-length coding. Experimental results are shown in Fig. 9.



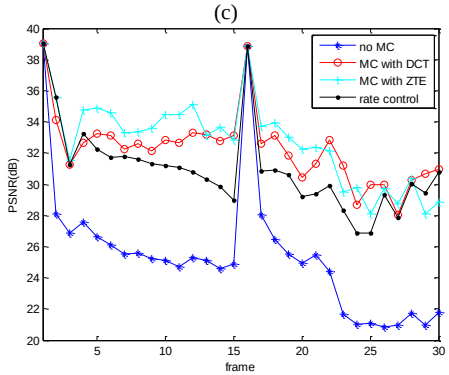
(a)



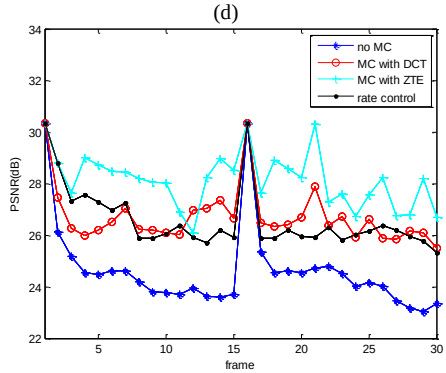
(b)



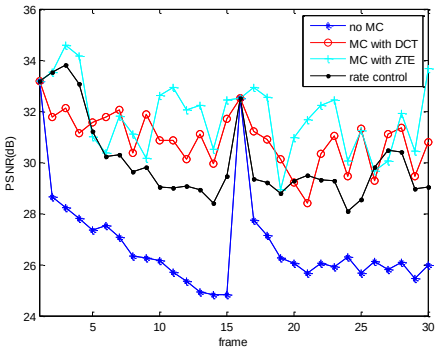
(c)



(d)



(e)



(f)

Figure 9. Comparison of MC with different schemes: (a) Foreman, (b) Mobile, (c) Bus, (d) Football, (e) Garden, (f) Stefan.

As shown in Fig. 9, MC can efficiently improve the quality of reconstructed frame. At the same bit rate, the wavelet combining with ZTE algorithm is more efficient than MC with DCT in most occasions. This is mainly because that DCT is appropriate for residual data got from BMA. While ZTE algorithm makes use of the relationship of wavelet coefficients got from the residual data. In the six tested sequences, ZTE doesn't perform better than DCT for Football sequence. This is because that the Football sequence includes intense motion. The NS-DBMA can't get a satisfied prediction for it. So the residual data is quite large. After wavelet transform, many high-frequency components exist. And these components will be discarded in the process of quantization. So the quality of reconstructed image decreases. While for the MC with DCT, the experimental quantization matrix can keep most high-frequency components. So the reconstructed image is rather good. To sum up, the wavelet transform combining with ZTE performs well for complex motion, translational motion and motion with rich details. For intense motion, the two methods have similar effect.

The rate control not only controls the output rate, but also makes the reconstructed frame smooth. In addition, it meets the requirements of human's subjective feeling well.

VI. Conclusion

In this paper, a new algorithm based on wavelet transform and NS-DBMA with motion compensation is proposed. Experiments are conducted on some typical video sequences. Firstly, comparisons are made with the conventional method in the following aspects: I frame coding, motion estimation and motion compensation. Experimental results are presented and the existing problems are analyzed. Finally, the overall performance of the proposed algorithm is compared with conventional algorithms. The experimental results demonstrate the efficiency of the proposed algorithm.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 61201371, No. 61002027). The authors would like to thank Chunxiao Zhang, Xiaoyan Wang, and Chunhui Guo for their help and valuable suggestions. The authors also thank the anonymous reviewers and the editor for their valuable comments to improve the presentation of the paper.

References

[1] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 4, no. 4, pp. 438-442, Aug 1994.

[2] L. M. Po and W. C. Ma, "A novel four-step search algorithm for fast block motion estimation," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 6, no. 3, pp. 313-317, Jun 1996.

[3] L. K. Liu and E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 6, no. 4, pp. 419-422, Aug 1996.

[4] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block matching motion estimation," in *Proc. Int. Conf. on Inform., Commun., Signal Process.*, Singapore, Sep 9-12, 1997, pp. 292-296.

[5] A. M. Tourapis, O. C. Au, and M. L. Liou, "New results on zonal based motion estimation algorithms - advanced predictive diamond zonal search," in *Proc. IEEE Int. Symp. on Circuits and Syst.*, Sydney, Australia, vol. 5, May 6-9, 2001, pp. 183-186.

[6] S. Nogaki and M. Ohta, "An overlapped block motion compensation for high quality motion picture coding," in *Proc. IEEE Int. Symp. on Circuits and Syst.*, New York, USA, vol. 1, May 10-13, 1992, pp. 184-187.

[7] M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach," *IEEE Trans. on Image Process.*, vol. 3, no. 5, pp. 693-699, Sep 1994.

[8] B. Girod, "Motion compensation: visual aspects, accuracy, and limitations," in *Motion Analysis and Image Sequence Process.*, M. I. Sezan and R. L. Lagendijk, Eds., Boston: Kluwer Academic Publishers, Mar 1993, pp. 125-152.

[9] Z. B. Chen, P. Zhou, Y. He, and G. Z. Wang, *Fast Motion Estimation for JVT, Documents JVT-G016, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)*, Pattaya, Thailand, March 2003.

[10] W. Wei, Z. X. Hou, and Y. C. Guo, "Motion compensation using nodal search-based deformable block matching," *Acta Electronica Sinica*, vol. 33, no. 8, pp. 1421-1424, Aug 2005.

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

[12] W. Wei and Z. X. Hou, "A fast motion estimation algorithm with adaptive threshold," *Journal of Optoelectronics.Laser*, vol. 19, no. 9, pp. 1254-1257, Sep 2008.

[13] S. A. Martucci, I. Sodagar, T. H. Chiang, and Y. Q. Zhang, "A zerotree wavelet video coder," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 7, no. 1, pp. 109-118, Feb 1997.

[14] C. Y. Wang, Z. X. Hou, K. He, and A. P. Yang, "JPEG-based image coding algorithm at low bit rates with down-sampling and interpolation," in *Proc. 4th Int. Conf. on Wireless Commun., Networking and Mobile Computing*, Dalian, China, vol. 3, Oct 12-14, 2008, pp. 3226-3230.

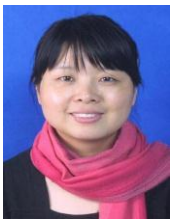
[15] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification, *ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC*, Joint Video Team, Mar. 2003.

[16] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Syst. for Video Technol*., vol. 13, no. 7, pp. 560-576, Jul 2003.
[17] I. E. G. Richardson, "H.264 and MPEG-4 video compression: Video coding for next-generation multimedia," in *Southern Gate, Chichester*, UK: John Wiley & Sons Ltd., Sep. 2003.
[18] L. L. Zhao, "All phase digital filtering and image interpolation based on IDCT/DCT," Ph.D. dissertation, School of Electron. Inform. Eng., Tianjin Univ., Tianjin, China, Jun. 2006, pp. 98-101.

Zhengxin Hou was born in Tianjin, China in 1945. He received his B.S. degree in radio electronics from Peking University, China, in 1969 and his M.E. degree in communication and electronic systems from Tianjin University, China, in 1982. Now he is a professor and supervisor of the Ph.D. students in the School of Electronic Information Engineering, Tianjin University, China. His current research interests include digital filtering theory, digital image and video processing, and wavelet analysis.

Baochen Jiang was born in Shandong province, China in 1962. He received his B.S. degree in radio electronics from Shandong University, China, in 1983 and his M.E. degree in communication and electronic systems from Tsinghua University, China, in 1990. Now he is a professor in the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China. His current research interests include signal and information processing, image and video compression, and smart grid technology.

Yupeng Cao was born in Liaoning province, China in 1985. He received his B.E. degree in electronic information engineering from Dalian University of Technology, China, in 2007 and his M.E. degree in signal and information processing from Tianjin University, China, in 2009. His research interests include digital image and video processing.



Aiping Yang was born in Shandong province, China in 1976. She received her B.Ed. degree in mathematics education from Shandong Normal University, China, in 2000 and her M.S. degree in operations research and control theory from Tianjin University, China, in 2003 and her Ph.D. degree in signal and information processing from Tianjin University, China, in 2008. Now she is an associate professor in the School of Electronic Information Engineering, Tianjin University, China. Her current research interests include digital image and video processing, compressive sensing theory, and wavelet analysis.

Chengyou Wang was born in Shandong province, China in 1979. He received his B.E. degree in electronic information science and technology from Yantai University, China, in 2004 and his M.E. and Ph.D. degree in signal and information processing from Tianjin University, China, in 2007 and 2010 respectively. Now he is a lecturer in the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China. His current research interests include digital image and video processing, wavelet analysis and its applications, and smart grid technology.