A Novel Error Concealment Algorithm for H.264/AVC

Jinlei Zhang and Houqiang Li

Chinese Academy of Sciences Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, Hefei 230027, China
jlzhang@mail.ustc.edu.cn,lihq@ustc.edu.cn

Abstract. To benefit network transmission, the bit stream of the whole frame coded by H.264/AVC is usually grouped into one packet. However, the packet loss during transmission will lead to the distortion of the reconstructed video and the error propagation. To deal with this problem, error concealment (EC) strategy is widely used to recover the lost frames and to weaken the effect of error propagation. In this paper, we propose a new EC method which uses both the forward and backward motion information to recover the motion information of the current lost frame. Besides, for the pixels whose motion information is quite different compared with its neighboring pixels, we propose to use the spatial correlation to fill up the pixels by minimizing the total variation (TV) norm. With the help of our proposed algorithm, we can obtain the motion vectors of the lost frames and improve the accuracy of motion vectors derived by the unidirection recovery method. Besides, the optimization strategy, minimizing the TV-norm for the pixels with quite different motion vectors, can help the decoder to recover the reconstructed frames more accurately. Experimental results show that the proposed algorithm can achieve both better objective performance and subjective performance compared to well-known schemes.

Keywords: Video Coding, Error Concealment, Whole Frame Recovery, H.264/AVC.

1 Introduction

With the rapid development of digital technology, video communications over networks have received increasing attention from the communication applications. During the transmission process in the network environment, various channel errors may lead to damage or loss of compressed video data packets. Besides, as we known, video coding uses the temporal prediction which may cause the errors propagate easily to succeeding frames. As a consequence, it is necessary to design appropriate error resilience and concealment tools for video encoders and decoders in order to counteract such losses and alleviate the error propagation.

Many related works [1,2,3,4,5] have been proposed. However, most of these methods assume that only a few blocks or slices in a video frame are missing,

X. He et al. (Eds.): MMM 2015, Part I, LNCS 8935, pp. 466–476, 2015.

and they are incompetent for the case when the data of a whole frame is lost. Actually, in video coding, the temporal correlation is so strong that the compression performance of inter prediction is very effective and a whole frame can be encapsulated into one packet. Several algorithms have also been proposed to deal with whole frame loss. The simplest whole frame loss error concealment (EC) method[6] used in H.264 is the frame copy (FC) method which directly copies the previous reconstructed frame to recover the lost frame. However, the FC method may lead to the frame un-continuity especially for the high motion videos. Another method called motion copy (MC)[6] duplicates the motion vectors (MVs) of the previous reconstructed frame and recover the missing frame through motion compensation. However, the MC method is based on an assumption that the collocated blocks in successive frames have similar motion activity. Actually, the assumption is not suitable for all the blocks in the frame especially for the high motion videos.

Some other algorithms [7,8,9,10,11,12] try to generate a better motion vector for the lost frame. The main ideas can be classified into three typical categories.

1. Taking full advantage of the previous reconstructed frames' MV information instead of only one previous frame. [7] and [8] use the concept of optical flow to generate the MVs of each pixel in the missing frame. And based on the theory of multiple hypotheses, [9] devises an adaptive integration scheme to make full use of each hypothesis' strength, which fully exploits the correlation between consecutive frames.
2. Making use of both the forward and backward frames' MV information. [10] first proposes to generate the MV of the missing frame based on pixel level bi-direction frames. And [11] proposes another algorithm to recover the MV of each pixel in the missing frames based on pixel level bi-direction frames' MV information.
3. Classifying the frame into different types, and use different filtering method to recover the missing frame. [12] classifies the blocks into three types (i.e. covered state, conflict state, and non-covered state), and designs different EC method for each type.

All of the algorithms above focus on recovering the MV information using the temporal correlation, which could only provide limited information for frame recovery. In this paper, not only the temporal MV correlation, we also make use of the spatial correlation. Firstly, we classify the blocks into three classes (i.e., new scene, uniform linear motion and non-uniform linear motion), and using different MV recovery methods and motion compensation methods to recover the pixels in the missing frame. Secondly, we pick the pixels with incorrect MVs based on the previous estimated MV of each pixel, and recover them by minimizing the total variation (TV) norm.

The rest of the paper is organized as follows. In Section 2, the main idea of the MV generation and the minimizing TV-norm algorithm are elaborated. In Section 3, the objective and subjective results of the proposed algorithm are presented. At last, we conclude this paper in Section 4.

2 Proposed Error Concealment Algorithm

The existing methods which make use of the bi-directional MV information or multiple previous frames are able to provide a better performance compared with the MC method. However, they also have the shortcoming. The MVs of the lost blocks are generated based on the assumption that all the blocks are uniform linear moving. Actually, it is not always satisfied. Some MVs are very likely to be estimated inaccurately, especially in large motion activity. This shortcoming will damage the accuracy of the MV for the block, which will result in the degraded performance. In order to overcome this problem, we propose a hybrid EC method based on bi-directional MV information and the spatial correlation, which uses not only the bi-directional MVs, but also the spatial correlation to get rid of the inaccurate MV pixels. Thus, the proposed algorithm is able to discard the pixels with inaccurate MVs and recover them by exploiting spatial correlation, which is helpful to protect the spatial smoothness of the frame.

2.1 Blocks Classification

As we known, the content of the sequences is always changed. Sometimes new scenes come into the content, and sometimes cover phenomenon happens. Thus, the uniform linear motion assumption is always incorrect for these blocks. In this paper, we first classify the blocks into three types, i.e. new scene, uniform linear motion and non-uniform linear motion. The sorting rules are as follows:

1. If the collocated block in the forward frame is intra predicted, then we define the missing block as the new scene for the forward frames, and the forward MV information is neglected and marked as the unavailable MV; otherwise, the forward MV is marked as available MV which will be used to estimate the MV of the missing block.
2. If the collocated block in the backward frame is intra predicted, then we define the missing block as the new scene for the backward frames, and the backward MV information is neglected and marked as the unavailable MV; otherwise, the backward MV is marked as available MV.
3. We first scale the MV information according to the reference index as follows,

$$MV_x = \frac{MV_x}{ref+1}; MV_y = \frac{MV_y}{ref+1} \tag{1}$$

   where ref means the reference index of the block. And then according to the scaled MV information of the two directions, if the difference of these two scaled MVs is greater than a threshold, then we define the block as non-uniform linear motion block; otherwise, the block is marked as uniform linear motion block.
4. For the non-uniform linear motion blocks, if the difference of the residuals in the two directions is larger than a threshold, then the corresponding MV information of the block with larger residuals is ignored and marked as the unavailable MV, while the other MV information is marked as available MV.

2.2 Bi-directional Motion Compensation

According to the classification of the blocks in Subsection 2.1, we recover the missing pixels using bi-directional motion compensation. As discussed in [10], the block-based motion vector estimation (MVE) method provides similar performance as pixel-based MVE method in little motion scenes. However, for the scenes with high motion activity, pixel-based MVE has an obvious better performance than block-based MVE. Thus, in this paper, we also use the pixel-level motion compensation method to recover the missing pixels.

On one hand, if the forward MV is available, and it is $f = (f_x, f_y)$, then the lost pixel $p_f(x, y)$ can be recovered as follows,

$$p_f(x, y) = p_r(x + f_x, y + f_y) \tag{2}$$

where $p_r(x, y)$ means the pixels in position (x, y) of the previous frame.

On the other hand, similar to the forward recovery method, if the backward MV is available, and it is $b = (b_x, b_y)$, then the lost pixel recovered by backward prediction $p_b(x, y)$ is as follows,

$$p_b(x, y) = p_r(x + b_x, y + b_y) \tag{3}$$

where $p_r(x, y)$ is the pixels in the last correctly reconstructed frame, and the (b_x, b_y) is the scaled MV corresponding to the reference frame.

At last, we combine the forward and backward pixel-based MVE methods. For the pixel in position (x, y) of the lost frame, its recovery value is estimated as follows,

$$p(x, y) = \omega(x, y) \times p_f(x, y) + (1 - \omega(x, y)) \times p_b(x, y) \tag{4}$$

where the pixel-based weight $\omega(x, y)$ is used to adjust the weights of the forward and backward methods. In this paper, we set $\omega(x, y)$ for all the pixels within the block according to the classification of the block as follows,

$$\omega(x, y) = \begin{cases} 1, & only\, f = (f_x, f_y)\, is\, available \\ 0, & only\, b = (b_x, b_y)\, is\, available \\ 0.5, & both\, f = (f_x, f_y)\, and\, b = (b_x, b_y)\, are\, available \end{cases} \tag{5}$$

2.3 Error Concealment Using Spatial Correlation

As the bi-directional MVE presented above, the pixels have the corresponding MV information. However, the MV of some pixels may be estimated inaccurately. In this subsection, we judge the accuracy of the estimated MV for each pixel as follows,

$$idx(x, y) = \begin{cases} 0, & if MV(x, y) - MV(x \pm 1, y \pm 1) > th \\ 1, & otherwise \end{cases} \tag{6}$$

where $MV(x \pm 1, y \pm 1)$ means the MV information of the pixels in the 8 neighboring area of current pixel. $idx(x,y) = 0$ means that the recovered pixel of (x,y) is inaccurate, and $idx(x,y) = 1$ means the recovered pixel is accurate.

According to the judgement method discussed above, there will be some holes in the missing frame since some of the recovered pixels are inaccurate. In other word, once the bi-directional MVE is completed, all elements in the missing frame are supposed to be trustworthy except for few remaining hole regions. In this subsection, we propose a method to recover the inaccurate pixels by making use of the spatial correlation. Let the current frame be denoted by $Y(x,y) \in R^{m \times n}$, and let Ω denote the set of trustworthy entries in $Y(x,y)$. Then the next task can be mathematically expressed as: given the set Ω, we shall recover $M \in R^{m \times n}$ from its incomplete observation $Y \in R^{m \times n}$. Since the temporal correlation has been exploited at subsection 2.2, we assume that the temporal smoothness has been satisfied. Thus, in this step, the spatial smoothness shall be enforced during the process of completion to reduce the unpleasant artifacts. Therefore, a 2-D TV regularized completion model is designed to achieve this objective.

We define the operator $O = [O_x^T, O_y^T]^T$, as the sub-operators of horizontal and vertical directions. The individual sub-operators are defined as follows,

$$\begin{aligned} O_x(M) &= vec[M(x+1,y) - M(x,y)] \\ O_y(M) &= vec[M(x,y+1) - M(x,y)] \end{aligned} \tag{7}$$

where $vec[\cdot]$ stands for the vectorization operation. The definition of 2-D TV norm of a space volume M is then expressed as,

$$||M||_{2-DTV} := \sum_i (|[O_x(M)]_i| + |[O_y(M)]_i|) \tag{8}$$

where $[\cdot]_i$ means the i^{th} element from the 2-D volume. Since the definition in Eq.(8) has a similar formulation with the l_1 norm, we express $||M||_{2-DTV}$ in a l_1 norm-like manner as,

$$||M||_{2-DTV} = ||O(M)||_1 \tag{9}$$

From the definition of 2-D TV norm, we can see that it provides a natural piecewise smoothness in space domain for the missing frame, with edge preserving property. Then, we formulate a convex program as follows,

$$\begin{aligned} &\min_M ||M||_{2-DTV} \\ &s.t.\ P_\Omega(M - Y) = 0 \end{aligned} \tag{10}$$

$P_\Omega(M - Y)$ shall project into $R^{m \times n}$ and the projection is the $(i,j)^{th}$ entry of $M - Y$ if $(i,j) \in \Omega$, otherwise 0.

To solve Eq.(10), we introduce an auxiliary variable $Z \in R^{m \times n}$ and reformulate Eq.(10) into

$$\begin{aligned} &\min_{M,Z} ||M||_{2-DTV} \\ &Y = Z + M \\ &P_\Omega(Z) = 0 \end{aligned} \tag{11}$$

Algorithm 1. ADM algorithm to solve Eq.(11)

Input: Y, tol
1: Initialize M, Z, Λ, β and p.
2: **while** $\frac{||M+Z-Y||_F}{||Y||_F} \geq tol$ **do**
3: Update M^{k+1}: solved the 2-D total variation regularized least square problem using Algorithm 2.
4: Update Z^{k+1}:$Z^{k+1} = P_\Omega(\frac{1}{\beta}\Lambda^k - M^{k+1} + Y)$.
5: Update Λ^{k+1}:$\Lambda^{k+1} = \Lambda^k - \beta(Y - Z^{k+1} - M^{k+1})$.
6: Update β:$\beta = p \times \beta$
7: **end while**
Output: M

To solve Eq.(11), we apply the alternating direction method (ADM) [13], and we consider the augmented Lagrangian function of Eq.(11) as follows,

$$\begin{gathered} L_A(M, Z, \Lambda, \beta) = ||M||_{2-D\,TV} - \\ < \Lambda, Y - Z - M > + \frac{\beta}{2}||Y - Z - M||_F^2 \end{gathered} \tag{12}$$

where $\Lambda \in R^{m\times n}$ is the Lagrange multiplier attached to the equality constraint $Y = Z + M$, β is a positive value either constant or varying with the optimization process, $< \cdot >$ denotes the matrix inner product and $|| \cdot ||_F$ denotes the Frobenius norm. Each iteration of the ADM optimizes the augmented Lagrangian function with respect to M and Z in a coordinate descent manner, and updates the Lagrange multiplier accordingly. Therefore, the original problem is now decomposed into several simpler subproblems. Assume that after the k^{th} iteration, the parameters are (M^k, Z^k, Λ^k), then the ADM generates the $(k+1)^{th}$ iterate $(M^{k+1}, Z^{k+1}, \Lambda^{k+1})$ as

$$M^{k+1} = arg \min_M ||M||_{2-D\,TV} + \frac{\beta}{2}||M - Y + Z^k + \frac{1}{\beta}\Lambda^k||_F^2 \tag{13}$$

$$Z^{k+1} = arg \min_{Z:P_\Omega(Z)=0} \frac{\beta}{2}||Z - Y + M^{k+1} + \frac{1}{\beta}\Lambda^k||_F^2 \tag{14}$$

$$\Lambda^{k+1} = \Lambda^k - \beta(Y - Z^{k+1} - M^{k+1}) \tag{15}$$

The detailed ADM algorithm to solve the problem is shown as Algorithm 1.

Each subproblem is solved as follows:

1. M-subproblem: The first subproblem to get the new iteration of M^{k+1} is a 2-D TV regularized least squares problem. A similar problem has been investigated in [14], in which a 3-D TV problem is formulated. Different to the problem in [14], in this paper, it is a 2-D TV problem, and we only need to compute the matrices on the spatial dimension. The detailed algorithm is shown as Algorithm 2.

Algorithm 2. The algorithm to update M^{k+1}

Input: $Q^k = Y - Z^k - \frac{1}{\beta}\Lambda^k$, tol_M

Initialize M^0, u^0, P^0, μ, q and $i = 0$.
Compute the matrices $\mathcal{F}[O_x]$, $\mathcal{F}[O_y]$ and $\mathcal{F}[I]$
while $\frac{||M_{i+1}-M_i||_F}{||M_i||_F} \geq tol_M$ **do**
  Update M_{i+1}: $M_{i+1} = \mathcal{F}^{-1}[\frac{\mathcal{F}[\beta Q^k+\mu O^T u - O^T P]}{\beta|\mathcal{F}[I]|^2+\mu(|\mathcal{F}[O_x]|^2+|\mathcal{F}[O_y]|^2)}]$.
  Update $u_{i+1} = (u_x, u_y)$:$u_x = max(|O_x M_{i+1} + \frac{1}{\mu}P_x| - \frac{1}{\mu}, 0) \cdot sign(O_x M_{i+1} + \frac{1}{\mu}P_x)$. u_y is derived similar to u_x.
  Update the Lagrange multiplier P_{i+1}:$P_{i+1} = P_i - \mu(u_{i+1} - OM_{i+1})$.
  Update μ:$\mu = q \cdot \mu$.
  $i = i + 1$.
end while
$M^{k+1} = M_i$

Output: M^{k+1}

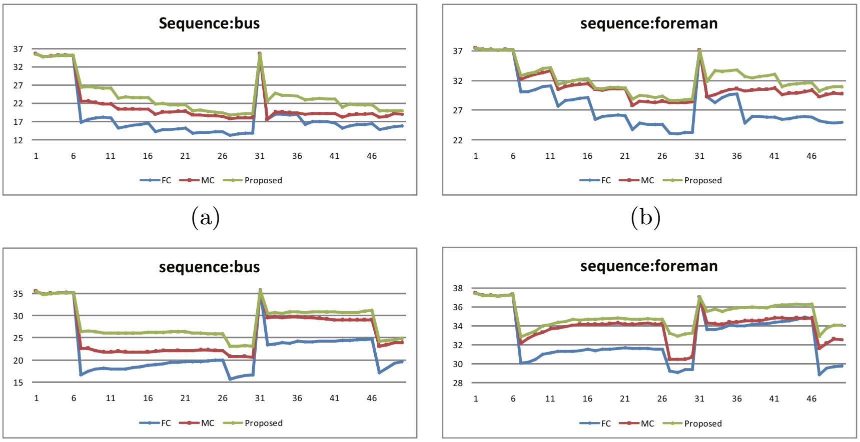
2. Z-subproblem: The second subproblem is to update Z^{k+1}, it is obtained as,

$$Z_{i,j}^{k+1} = \begin{cases} \frac{1}{\beta}\Lambda^k - M^{k+1} + Y, & if(i,j) \notin \Omega \\ 0, & otherwise \end{cases} \tag{16}$$

3 Experimental Results

To prove the effectiveness of the proposed algorithm, we first compare the algorithm with FC and MC methods. The proposed algorithm is implemented on the H.264/AVC reference software JM18.5. Five test sequences $foreman$($30Hz$, 300 frames), bus($15Hz$, 75 frames), $mobile$($30Hz$, 300 frames), $basketballdrive$($50Hz$, 500 frames) and $soccer$($30Hz$, 300 frames) are chosen to evaluate the performance of the proposed algorithm. All of the test sequences are in QCIF size. The period of intra frame reset is 30 and the number of reference frames is 5. A constant QP of 28 is maintained for all the frames.

In this simulation, we test two cases: 1) one frame is dropped in every 5 pictures; 2) one frame is dropped in every 20 pictures. And the dropped frames are concealed with EC algorithms. The peak signal-to-noise ratio(PSNR) is chosen as the objective measurement, which is computed using the original video sequence as the reference. The PSNR values of each method for the first 50 frames of bus and $foreman$ sequences are plotted in Fig. 1. It is obvious that the proposed algorithm outperforms the existing method under different dropping cases. Besides, the average PSNR performances over all the erroneous frames are shown in Table 1 and Table2, from which we can see that the proposed algorithm yields higher PSNR performances than the other methods. For the first case, the proposed algorithm achieves 4.662dB and 1.336dB in average performance gain compared with FC and MC, and for the second case, it achieves about 4.334dB



(a) (b) (c) (d)

Fig. 1. PSNR results of different EC algorithms: the top two figures correspond to one frame dropped every 5 frames, and the bottom two figures correspond to one frame dropped every 20 frames

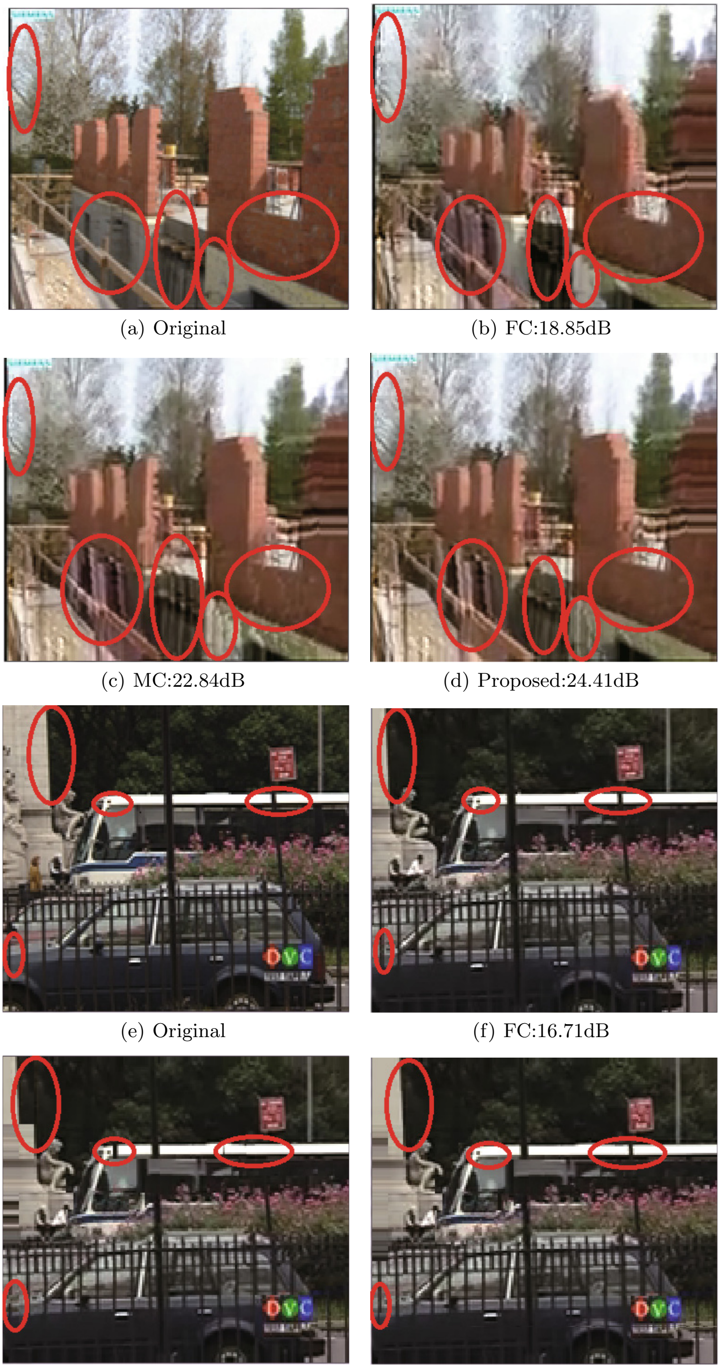
Table 1. Compared the proposed algorithm with FC and MC, one frame is dropped every 5 frames

Sequences	FC	MC	Proposed	Proposed-FC	Proposed-MC
foreman	23.06	25.50	27.09	4.03	1.59
bus	15.63	18.99	21.02	5.39	2.03
mobile	20.68	25.96	26.60	5.92	0.64
basketballdrive	19.75	23.04	24.88	3.29	1.84
soccer	17.44	19.70	20.28	2.26	0.58
average	19.312	22.638	23.974	4.662	1.336

and 1.466dB in average. For some sequences with new scene coming into the content, such as *bus*, the performance gain achieves to 5.39dB and 6.05dB compared with FC under two test cases.

For subjective evaluation, one original frame and three recovered frames reconstructed by FC, MC and the proposed algorithm are demonstrated in Fig.2. The 6^{th} frame of *bus* and the 239^{th} frame of $foreman$ are chosen. From Fig.2, we can observe that the proposed algorithm has an obvious subjective quality improvement compared with FC and MC.

Since the proposed algorithm can be easily transplanted to other versions of the codec and achieve an equally good performance, we transplant the codec into JM10.2 to compare with the result reported in [11]. All the test conditions are the same as that used in [11]. The number of reference frames is 1, QP is 22, the sequence size is QCIF, intra period is 15, and one frame is dropped in



(a) Original (b) FC:18.85dB

(c) MC:22.84dB (d) Proposed:24.41dB

(e) Original (f) FC:16.71dB

(g) MC:22.64dB (h) Proposed:26.42dB

Fig. 2. Subjective quality results of different EC algorithms: the top four figures are sequence *foreman*, and the bottom four figures are sequence *bus*

Table 2. Compared the proposed algorithm with FC and MC, one frame is dropped every 20 frames

Sequences	FC	MC	Proposed	Proposed-FC	Proposed-MC
foreman	27.75	31.80	32.51	4.76	0.71
bus	19.07	22.86	25.12	6.05	2.26
mobile	27.00	29.26	30.37	3.37	1.11
basketballdrive	23.59	26.44	29.00	5.41	2.56
soccer	22.15	23.54	24.23	2.08	0.69
average	23.912	26.78	28.246	4.334	1.466

Table 3. Compared the proposed algorithm with the method proposed in [11]

Sequences	Error Free	FC	MC	[11]	Proposed	Proposed-[11]
mobile	38.80	26.69	30.68	31.82	32.71	0.89
bus	39.08	18.97	25.84	26.43	27.15	0.72

every 15 frames. The results are shown as Table 3, from which we can see that the proposed algorithm also outperforms the method proposed in [11].

4 Conclusion

In this paper, we propose a whole frame loss error concealment algorithm to recover the lost frames. Although the follow-up frames in temporal layer can not be decoded correctly, the motion information and the residuals can be decoded, Thus, we propose to use the bi-directional motion information to recover the MVs of the missing frames. Besides, we use the MV and residuals information to classify the blocks into three types, and we design different methods to deal with each type. Therefore, the proposed algorithm can achieve better MV information for the missing frames. Besides, by exploiting the spatial correlation, we propose to improve the recovered frames by minimizing the TV-norm of some incorrect pixels, which enhances the spatial smoothness of the recovered frames and provides a better subjective visual quality. Experimental results show that our proposed whole frame loss error concealment algorithm can achieve significant PSNR improvement compared with previous works and the subjective quality also outperforms previous works.

Acknowledgments. This work was supported in part by 973 Program under Contract 2013CB329004, the Fundamental Research Funds for the Central Universities under Contract WK2100060011, and the open project of Beijing Multimedia and Intelligent Software Key laboratory in Beijing University of Technology.

References

1. Kumwilaisak, W., Kuo, C.J.: Spatial error concealment with sequence-aligned texture modeling and adaptive directional recovery. Journal of Visual Communication and Image Representation 22(2), 164–177 (2011)
2. Zhang, R.F., Zhou, Y.H., Huang, X.D.: Content-adaptive spatial error concealment for video communication. IEEE Transactions on Consumer Electronics 50(1), 335–341 (2004)
3. Zhai, G., Cai, J., Lin, W., Yang, X., Zhang, W.: Image Error-concealment via Block-based Bilateral Filtering. In: Proceeding of IEEE International ICME, pp. 621–624 (2008)
4. Zhang, Y.B., Xiang, X.G., Ma, S.W., Zhao, D.B., Gao, W.: Aoto Regressive Model and Weighted Least Squares based Packet Video Error Concealment. In: Proceeding of IEEE DCC, pp. 455–464 (2010)
5. Zhou, J., Yan, B., Gharavi, H.: Efficient Motion Vector Interpolation for Error Concealment of H.264/AVC. IEEE Transactions on Broadcasting 57(1), 75–80 (2011)
6. Wu, Z., Boyce, J.M.: An Error Concealment Scheme for Entire Frame Losses Based on H.264. In: Proceeding IEEE ISCAS (2006)
7. Belfiore, S., Grangetto, M., Magli, E., Olmo, G.: Concealment of Whole-Frame Losses for Wireless Low Bit-Rate Video Based on Multiframe Optical Flow Estimation. IEEE Transactions on Multimedia 7(2), 316–329 (2005)
8. Baccichet, P., Bagni, D., Chimienti, A., Pezzoni, L., Rovati, F.S.: Frame Concealment for H.264/AVC Decoders. IEEE Transactions on Consumer Electronics 51(1), 227–233 (2005)
9. Zhao, C., Ma, S.W., Zhang, J., Gao, W.: A Highly Effective Error Concealment Method for Whole Frame Loss. In: Preceeding of IEEE ISCAS, pp. 2135–2138 (2013)
10. Chen, Y., Yu, K., Li, J., Li, S.: An Error Concealment Algorithm for Entire Frame Loss in Video Transmission. In: Preceeding of IEEE PCS (2004)
11. Yan, B., Gharavi, H.: A Hybrid Frame Concealment Algorithm for H.264/AVC. IEEE Transactions on Image Processing 19(1), 98–107 (2010)
12. Wu, Z.Y., Boyce, J.M.: An Error Concealment Scheme for Entire Frame Losses Based on H.264/AVC. In: Preceeding of IEEE ISCAS, pp. 4463–4466 (2006)
13. Bertsekas, D.: Nonlinear Programming. Athena Scientific (1999)
14. Chan, S.H., Khoshabeh, R., Gibson, K.B., Gill, P.E., Nguyen, T.Q.: An Augmented Lagrangian Method for Total Variation Video Restoration. IEEE Transactions on Image Processing 20(11), 3097–3111 (2011)