

An effective temporal error concealment in H.264 video sequences based on scene change detection-PCA model

Gwangmin Choe¹ · Cholman Nam² · Changgon Chu²

Received: 29 October 2017 / Revised: 9 April 2018 / Accepted: 22 May 2018 /
Published online: 14 June 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract This paper proposes a new temporal error concealment algorithm in H.264 video sequences based on scene change detection and PCA model. In order to detect scene change, dynamic threshold and image similarity metric are presented using coding prediction mode and DCT AC energy in H.264 baseline. UPCA (Updated PCA) model is presented by combining the scene change feature with Index transformation-Buffer updating approach. The lost images are concealed by Projection onto Convex Sets algorithm with UPCA model. Experimental results show that the proposed algorithm can achieve better error concealment performance for the higher motion and the frequent scene change, compared with the related method.

Keywords Error concealment · Scene change · H.264/AVC · PCA · Face recognition · Video sequence

1 Introduction

Wireless channel is error-prone due to its limit in bandwidth. In H264 which has high compression rate, one bit error results considerable decrease of PSNR of image for error propagation.

In [1], QoS specifications for LTE system is specified by service profile. For video stream service, packet error rate is specified as $10^{-3} \sim 10^{-6}$. However, specifications of QoS class

✉ Gwangmin Choe
ryongnam21@yahoo.com

¹ Visual Information Processing Laboratory, School of Computer Science and Technology, Kim Il Sung University, Pyongyang, Democratic People's Republic of Korea

² Information and Communication Laboratory, School of Computer Science and Technology, Kim Il Sung University, Pyongyang, Democratic People's Republic of Korea

5 and 9 do not guarantee bit error rate over radio channels, because they have non-GBR (non-Guaranteed Bit Rate).

In [18, 19], radio resources are allocated to users by using an optimal scheduling scheme for video transmission over multi-channel multi-radio multi-hop networks, and then transmission error rate is decreased whereas PSNR of HDTV video stream is increased. However, practical radio channels aggravate channel quality due to fast fading [2].

Error concealment is an effective way to recover lost information at the decoder (receiver) on condition bit error has already been occurred. Using error concealment, decoder can recover the lost block in current frame corrupted by bit error. It is clear that error concealment is not in accordance with an DMDS scheme proposed by [19], in which bit error rate of transmission channel is decreased using an effective radio resource scheduling to users. When a bit is inverted during transmission, PSNR of received video cannot be assured. In [2], the corrupted frames are filled by duplicating other frames, so that the received video has the same length as the sent video.

Video of QCIF and CIF formats are widely used in low display applications, such as 3G mobile systems and tactile dis-players. 2G and 3G mobile phones, which holds 80% of mobile subscribers [8], use CIF/QCIF videos more than HDTV video since 3G system only supports about 384kbps~2Mbit/s transmission bandwidth. In tactile displayer, display's resolution needs one receptor per mm^2 spacing of mechanoreceptors in human finger pads [14, 15].

Many researchers have already proposed effective temporal error concealment methods in which the temporal redundancy information is used to recover lost pixels in error image. Since motion vectors (MVs) play very important role in concealing the lost image, they studied how to recover a more accurate MV in depth discussion. In [10], the lost or erroneously received MVs are recovered with Boundary Matching Algorithm (BMA), in which the optimal MV is picked out from a set of candidates-MVs based on temporal continuity in video sequences, and then the lost pixels are interpolated by the MVs.

A framework of using various motion vectors for different regions within a damaged macro block was presented in [5]. In the early methods, discrete grotesqueries have been often appeared at boundary regions of the recovered MBs. In [3], the mosaic artifacts are decreased by combining motion compensation approach with BMA.

Different from BMA, in [12], an orientation adaptive interpolation scheme was derived from the pixel-wise statistical model instead of on a block-by-block basis. Although the motion vector or pixel-based methods are implemented with better performance over the other methods, yet, the methods show poor performance in concealing video frames with higher mobility or frequent scene changes.

Other methods, called as model or context-based error concealments, were proposed for spatial or temporal error concealment [6, 13]. A dynamic model Updating Principal Components (UPC) was proposed as a model for spatial error concealment [6]. When scene change occurs, UPC produces a relatively fine concealed results for loss of pixel data. However, the model does not show good performance, when there is no scene change.

The Principal Component Analysis (PCA) has long been used to model the visual content of images, for object recognition (especially faces) and reconstruction. In [13], a priori information for error concealment is obtained by training the context based PCA model for a Region of Interest (ROI) and uses a training model for reconstruction of lost data within the ROI. PCA model is not good in reconstruction, because the training model may be composed of dissimilar contents of current image, when scene change occurs. As can be seen from the related works, scene change influences the error concealment strategy.

In recent years, some studies have been carried out to detect video scene change in H.264 compressed domain, in which encoding features such as intra and inter prediction mode were used mainly as a basis for deciding scene change.

In [17], statistical constraint rule for scene change detection was proposed by intra mode feature, in which MB type is in conjunction with this rule to detect all scene change P frames, and intra mode histogram and the weighted distance metric are used to detect scene change I frames. Some researchers used the number of intra-coded macroblocks as decision criterion rule for detection of scene change [7]. In [16], three criteria, the difference of chroma prediction modes based on sub-block, the accumulative motion amount and the accumulative intra coding macroblock amount, were proposed. The algorithm is dependent on intra prediction modes, motion vectors and macroblock coding types that are extracted from H.264 baseline bitstream. Besides, there are a few MPEG Series-scene change algorithms using DCT coefficient, motion vector, macroblock coding mode [4] and the number of encoded bits [11].

Although these algorithms have better detection performance, but scene change in the H.264 QCIF (176×144) video streams is still a challenging research topic. No matter what decision criterion is used, the above algorithms are dependent on MB information, but if the number of MBs is not too much, the accuracy of detection is lowered. Since video temporal features are not fully considered in previous methods, we will propose a video characteristic training model based H.264 temporal error concealment algorithm in order to improve the error concealment performance. First, two decision criterions, dynamic threshold for the number of bits/frame and similarity metric for AC image, are presented separately based on the encoded prediction modes and DCT coefficients in H.264 compressed domain, and then a new H.264 QCIF video scene change detection algorithm is proposed. Next, PCA training model is updated with the scene change information. Finally, a new H.264 temporal error concealment is proposed based on POCS (Projection onto Convex Sets) with PCA algorithm.

2 Scene change based temporal error concealment

2.1 Scene change detection

In this section, we focus on dynamic threshold of inter frames and frame similarity of intra frames in H.264 baseline profile.

2.1.1 Intra MB and inter MB encoding

In H.264 standard, intra and inter prediction determine the coding mode for each MB. Intra prediction, for a MB, forms predictions of pixel values as linear interpolations of pixels from the adjacent edges of its neighboring MBs (or 4×4 block).

The intra predicted blocks are DCT transformed, quantized, and encoded by CAVLC to generate a compressed stream, if the real time application uses H.264 baseline profile. Inter MB encoding depends on whether the MB is a skipped MB or not. If the MB is a skipped MB, the MB 16×16 MV or Direct mode MV is calculated. If the MB is not a skipped MB, the direct mode MVs or differential MVs are encoded by CAVLC, according to whether direct mode is used or not. As a rule, I frame is encoded by intra mode and P frame is encoded by inter mode.

In addition, scene change P frames may be composed of intra or inter MBs, due to temporal continuity of frames. In this paper, if a particular frame is composed of intra MBs more than 70%, it is called intra frame or scene frame.

2.1.2 Scene change detection mechanism

The above review shows that intra prediction mode may be dominant, when scene change occurs.

Due to encoding mechanism, intra frames are compressed with lower efficiency than inter frames, and more bits, therefore, may be generated from them.

The difference between inter frame and intra frame helps to detect effectively scene changes over video streams. Let Y be a set of frames y_i and $Bits(y_i)$ be the number of bits of a frame. Then, the following equation yields intra frames, in terms of bits per frame.

$$\Gamma = \{y_i | Bits(y_i) > T(y_i)\} \tag{1}$$

$$T(y_i) = (Bits(y_i) + T(y_{i-1}))/2 + c \tag{2}$$

where, $T(y_i)$ depicts a dynamic threshold value for estimating prediction mode, c is an initial value depending on H.264 baseline profile.

Figure 1 shows the change of dynamic threshold for the number of bits. Also, it is obvious that intra frames can be automatically tracked by the dynamic threshold $T(y_i)$ in (2).

The plot has been measured on our edited video sequence (20 *carphone* frames + 50 *akiyo* frames + 30 *coastguard* frames) over that two scene changes occur at 21st and 71st frames. In (1), a set of intra frames, includes all I-, P- frames with scene change and I-frames with no scene change.

A frame similarity will be used to extract I-frames with no scene change out of it. Here, we adopt AC coefficient grouping scheme in DCT 4×4 block as shown in Fig. 2a. Each 4×4 block in MB includes 16 DCT coefficients which represents horizontal, vertical and diagonal AC features of the 4×4 block in pixel domain (time domain). For example, the 4×4 block is referred to horizontal block, if horizontal AC values are dominant AC values in vertical and diagonal directions.

For each direction of AC features, horizontal, vertical and diagonal AC energy- E_k^H , E_k^V and E_k^D of a DCT 4×4 block can be found as following

$$E_k^{H/V/D} = \sum_{m,n \in H/V/D} DCT^2(m, n) \tag{3}$$

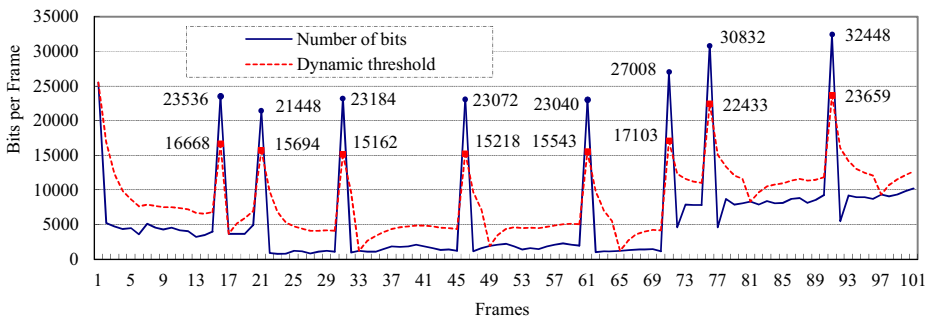


Fig. 1 Intra frames tracked by dynamic threshold

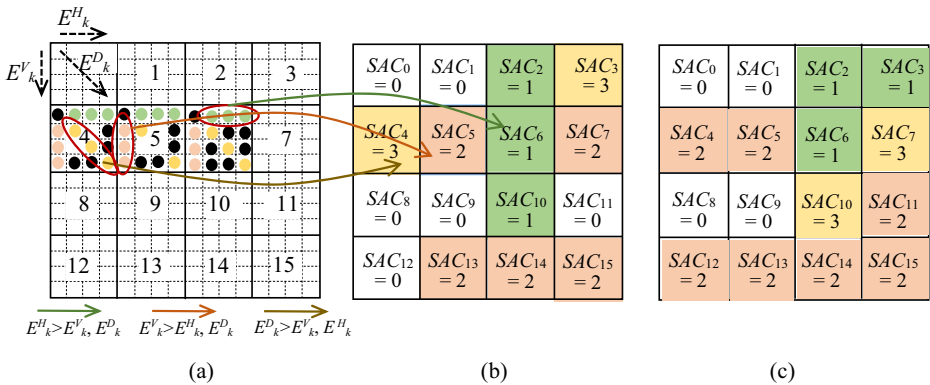


Fig. 2 Scheme grouping DCT coefficients of 4×4 sub-block (a), SAC of 49th MB in 6th *carphone* frame (b), SAC of 49th MB in 7th *carphone* frame (c)

where, k is an index of DCT 4×4 block and H, V and D are the sets of each direction coordinates within DCT 4×4 block.

Then, SAC (Statistical AC feature) of the DCT 4×4 block can be found as:

$$SAC_k = \begin{cases} D_H (= 1), (E_k^H > E_k^V, E_k^H > E_k^D) \\ D_V (= 2), (E_k^V > E_k^H, E_k^V > E_k^D) \\ D_D (= 3), (E_k^D > E_k^V, E_k^D > E_k^H) \\ 0 \end{cases} \quad (4)$$

SAC_k represents a dominant AC feature within DCT 4×4 block, in terms of AC energy.

We have analyzed the similarity of *carphone* frames using H.264 Visa Baseline Tool [7]. In Fig. 2, (b) and (c) show SAC values of 49th MBs (including man’s nose and lips) in 6th and 7th *carphone* frames, and it is called as SAC image. SAC images depict the same contents (blue region) being implied within two frames statistically.

Frame similarity Δf_i can be estimated as:

$$\Delta f_i = \sum_k \delta(SAC_k^{I_{i-1}} - SAC_k^{I_i}), I_i \in \Gamma \quad (5)$$

where Δf_i means the number of 4×4 blocks with the same feature of the same location in two frames. If $\Delta f_i < T_{AC} (= \sigma \times T_{max})$, I_i and I_{i-1} are similar, and if not, I_i and I_{i-1} are not similar, where σ is a similarity factor which is given as about 0.3 for H.264 QCIF video baseline profile. Then, scene detection is organized by frame similarity Δf_i as following: When I_i and I_{i-1} are not similar, if there are P-scene frames in between I_i and I_{i-1} , I_i is in non-scene frame, whereas if there is not a scene frame, I_i is in scene frame.

The scene change detection is done on H.264 encoder, and then the detected scene information is encapsulated in H.264 SEI Message Unit to transmit to decoder.

2.2 Temporal error concealment based on scene change detection-PCA model

In face recognition field, PCA has long been used to model the visual content of images for image reconstruction.

Let y be $(p \times q)$ image vector, and then image reconstruction \hat{y} can be obtained by combining linearly its individual components on an adequate eigenvector space as:

$$\hat{y} = m + \sum_k^p [(y - m)^T u_k] u_k \tag{6}$$

where m and u_k are mean image vector and k th component eigenvector in p -dimension eigenvector space respectively.

In PCA of face recognition, the mean image m is estimated by arithmetic mean over p images and u_k is given by covariance matrix that is estimated by subtraction of the p training images and the mean image. Image reconstruction error $e = y - \hat{y}$ depends on second term $y - m$ in (6), that is, whether mean image m is close to current image y or not. If there occurs a scene change in video streaming, a set of training images is composed of images with different visual contents, so it lowers the PSNR of the reconstructed image than image using the visual same content. The reconstruction image is unsuitable for error concealment, in which lost block is filled by close visual content. Our idea is to make the training images to be only the particular images having same contents related semantically, so that the arithmetic mean image m should be updated according to video characteristic.

Now, note that a scene is defined as a succession of individual frames (images) semantically related with each other. It implies that a scene of frames should be used as a set of training images. Practically, after scene change, a few or lot of frames not related with current content may be still in buffer of training images. Hence, PCA model could not produce the fine results for video stream. Consequently, image buffer has to be updated to include the only related contents in accordance with scene change.

Let n be current instant of frame, and let N be a buffer size which can be refreshed. Then, Index Transformation, $n + i - N$ means the past location of frames within buffer of size N . Using the transformation, we can derive an adequate mean image m from image buffer as:

$$m(n) = 1/N \sum_{i=0}^{N-1} y(n + i - N) \tag{7}$$

where $y(n + i - N)$ is image vector at time $n + i - N$, where $i - N$, ($i < N$) represents how far an image is from the current time instant.

To distinguish the mean image from the traditional PCA, it will be called as UPCA training model. In UPCA model, it is a technical skill that the training set is updated by index transformation, that is to say, a training contents $y(n - N)$, $y(n + 1 - N)$ to $y(n - 1)$ is updated in accordance with the scene change, while its size N is dynamically altered with it.

We realize it simply by connecting buffer pointer with scene change output. If there is not a scene change, mean image $m(n)$ is estimated by average of images filled in image buffer. At this time, image pointer points max scale, it coincides with fact that training images should be in a scene. For example, $N = 6$. If there is a scene change, image pointer is refreshed, and points to 1, 26, according to the input frames. Updating the mean image only by scene change dynamically improves performance of image reconstruction for the video sequences with more complex motion or abrupt scene change. $m(n)$ and $y(n + i - N)$ are noted as $m(n)$ and y_{n+i-N} respectively. Now, we want to estimate k th component

eigenvector, u_k by applying $m^{(n)}$ to PCA. $(p \times q) \times (p \times q)$ covariance matrix at current time instant n is estimated as following:

$$C^n = 1/N \sum_{i=0}^{N-1} [(y_{n-i} - m^n)(y_{n-i} - m^n)^T] \tag{8}$$

In order to avoid computational complexity, PCA uses to calculate eigenvector $v_l (l = 0, , N - 1)$ of $(N \times N)$ dimensional matrix $L = A^T A$, where $A = [y_{n-N-1} - m^{(n)}, , y_n - m^{(n)}]$. Then, eigenvector $u_k^n (k = 0, , N - 1)$ of $(p \times q) \times (p \times q)$ covariance matrix $C^{(n)}$ is estimated by combining linearly variance $y_{n-N-1} - m^{(n)} (i = 0, , N - 1)$ and $v_k^{(n)} (k = 0, , N - 1)$ as follows:

$$U^{(n)} = [u_0^n, \dots, u_{N-1}^n] = [y_{n-N-1} - m^{(n)}, \dots, y_n - m^{(n)}][V_0^n, \dots, V_{N-1}^n] = AV \tag{9}$$

In this paper, the number of eigenvector is chosen as $3 \sim 6$ from experience:

2.3 Error Concealment Based on Iterative-POCS

Error concealment based on POCS gives constraints on each unknown convex set.

The optimal solution is obtained by iteratively projecting the previous solution onto each convex set. Temporal error concealment is performed by iterative-POCS with PCA, composes of three stages, such as PCA training, image reconstruction and error concealment stages.

Our POCS algorithm with PCA is shown in Fig. 3. PCA model is pretrained with correctly received or concealed frames, Where, PCA use the images from buffer whose content is refreshed by scene detection output. When there occurs a corrupted frame, in order to decrease the iterations, the corrupted area may be, first, copy-pasted from previous frame. Then, the first copy-pasted frame is projected onto the PCA model. This iteration keeps on until $e = y_i - \hat{y}_i = 0$.

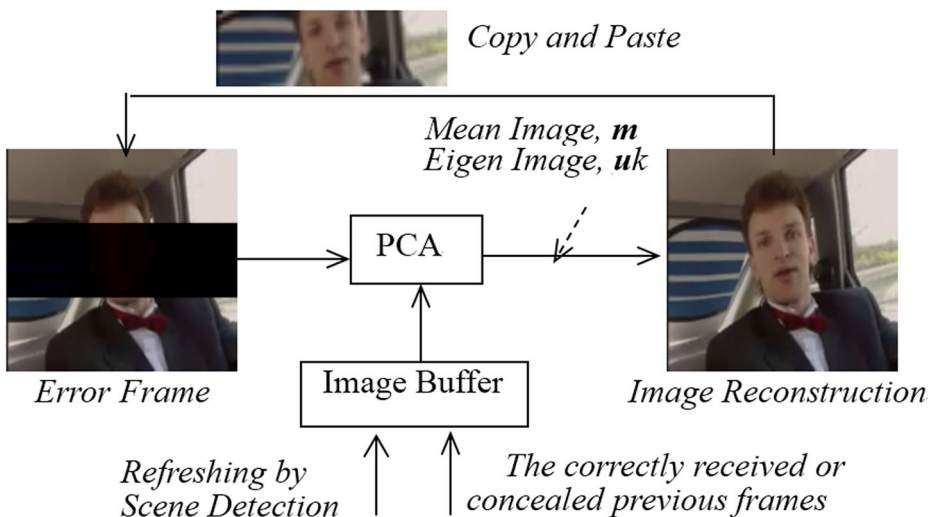


Fig. 3 Error concealment based on POCS with UPCA

3 Experimental results

In experiment, we use JM 15.0 reference software as the platform to simulate our algorithms. The standard video sequences in QCIF (176×144) format are analyzed. The encoder is based on the H.264 baseline profile and the QP is set to constant of 28,.

3.1 Scene change detection results

3.1.1 Performance estimation and detection result

Scene detection performance is tested on our video sequence edited by following order; 20 *carphone* frames, 50 *akiyo* frames, 50 *coastguard* frames, 20 *carphone* frames, 50 *foreman* frames, 50 *container* frames, 50 *mobile* frames, 50 *daughter – mother* frames, 30 *news* frames, 50 *salesman* frames, 50 *silent* frames and 40 *trevor* frames.

Figure 4 displays the change of the number of bits for frames and the detected scene changes.

In figure, 211th to 260th frames are in a scene of 50 *mobile* frames and 261st ~ 310th frames are in a scene of 50 *daughter-mother* frames. The *mobile* frames have bits/frame close to intra frames in *daughter-mother*. It is difficult to detect intra frames with a fixed threshold. However, in the video scenes, intra and inter frames are distinguished by the proposed dynamic threshold correctly. As a result, all intra frames are detected by dynamic threshold (red plot). Blue circles (\circ) represent the resulting intra frames. Please note that GOP (=IPPPP...IPPPP...) length is 15 and scene frames are also encoded by intra prediction mode. Then, all scene change frames of both I-frame and P-frame are intra-frame due to its video compress effect. In addition, all I-frames are intra-frame in H.264 video sequences. Therefore, if there is a P-frame in a set of intra-frames, it is scene change frame. The re-sulting frames, in figure, are of the selected instants within GOP cycle (green triangle mark Δ). Except the P-frames, the remaining frames with scene change are 121st, 211th and 361st intra frames. In order to pick out three frames from intra frames with GOP length of 15, AC similarity critical algorithm are applied to all intra frames.

Table 1 shows AC similarity calculated on I-frames of video frames from 1st to 226th. These results are produced by AC energy similarity of 4×4 blocks of current I-frame and previous I-frame, that is, SACs. The bigger the similarity, the more I-frames will be in the same scene, and the smaller, in the other scene. In our decision, if the current similarity is more than 30% current I-frame is similar to previous frame, and if not, it is not similar. In Table 1, 76th, 121st, and 211th frames are picked as non-similar frames by the judgment value. Of the three non-similar frames, the only 121st and 211th frames are selected as scene

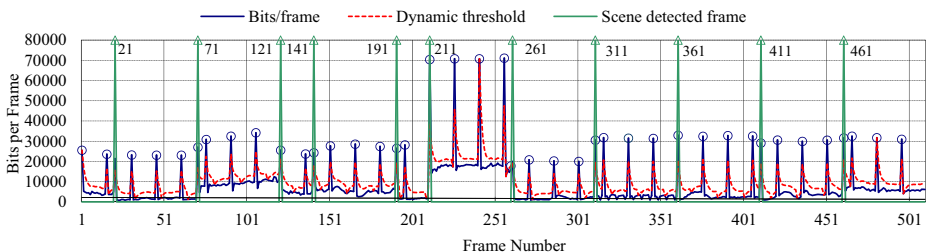


Fig. 4 Performance analyzes on our edited video sequence: 20 *carphone* + 50 *akiyo* + 50 *coastguard* + 20 *carphone* + 50 *foreman* + 50 *container* + 50 *mobile* + 50 *daughter – mother* + 30 *news* frames + 50 *salesman* + 50 *silent* + 40 *trevor* frames. Blue plot: bits/frame, red plot: dynamic threshold and green

Table 1 AC similarity of intra I-frames with GOP 15

Frame No	1	16	31	46	61	76	91	106
Similarity	881	1111	717	1351	1341	383	820	864
Frame No	121	131	151	166	181	196	211	226
Similarity	384	1118	592	627	745	634	421	551

change frame by our scene change estimation. In between 61st and 76th frames, 71st frame has been already selected as scene change frame by dynamic estimation. Meanwhile, while in between 106th and 121st frames, and in between 196th and 211th frames, there is no scene frame. Hence, 76th non-similar frame has been rejected from scene change frames. Similarly, 361st intra frame has been detected as scene change frame in Fig. 4. For whole scene changes, the resulting frames are shown as frame instant with green triangle marks (Δ), that is, all the 11 scene changes have been detected correctly over our tested video sequence.

3.1.2 Comparison with related works

In order to choose the best method for a given error concealment, several scene detection methods have to be compared. However, it is not easy to fairly compare the previous methods with each other, because there is not yet a standard video sequence for testing scene change. Hence, scene changes have been still tested in coinciding with their particular applications. In experiment, scene detection output is applied to error concealment for H.264 QCIF video sequences, in which the only scene cut or high motion rate is required. Therefore, we have compared a few related methods with our method. Of course, the comparison is not fair, because test sequence are edited sequences and are not well-known standard sequence. However, we can have a trend for the given methods. The proposed algorithm has been compared to other scene cut detection solutions, by the application of the MSU-Video Quality Measurement Tool [16], which is able to implement four different scene change methods, such as 1) SAD based Pixel-Level Comparison, 2) Global Histogram based SAD Estimation, 3) intra MB type-Based Detection, and 4) Motion Vector based Similarity Measurement. We use standard individual video sequences to edit a large video sequence, whose length is 5100 frames and in which 110 scene changes occur, in order to include as many different affects as possible, such as low and high mobility, panning, zooming, light variations, scene changes, and so on.

The performance measurements are implemented with the well-known Recall and Precision parameters [5], which are defined as follows.

$$\begin{aligned}
 \text{Recall} &= \frac{\text{Correct}}{\text{Correct} + \text{Missed}} \\
 \text{Precision} &= \frac{\text{Correct}}{\text{Correct} + \text{False}}
 \end{aligned}
 \tag{10}$$

The results are given in Table 2. It can be seen that the proposed method achieves on average 99% recall rate with a 98% precision rate for all scene changes. In comparison with the related methods, the proposed method achieves superior performances in terms of both recall and precision rates. Additional advantages with the proposed method can be highlighted the following as: 1) Algorithm is implemented for the abrupt scene change and the higher mobility in compressed domain, 2) thus, quite suitable for real-time error concealment in real time video sequences, 3) only two initial parameters are required for scene change detection.

Table 2 Performance comparison of scene detection methods

Detection algorithms	Correct	False	Missed	Recall	Precision
Proposed algorithm	110	2	1	0.99	0.98
MSU-1 [16]	107	31	3	0.97	0.76
MSU-2 [16]	105	8	5	0.96	0.93
MSU-3 [16]	91	24	19	0.83	0.79
MSU-4 [16]	99	12	11	0.9	0.89

3.2 Error concealment experiment

In this experiment, standard test QCIF (176×144) sequences, *carphone*, *akiyo*, *foreman*, *coastguard*, *news*, *trevor* and *mobile* are tested. In exception of *news* and *trevor*, the other sequences are composed of a scene, that is, scene change does not occur in the sequences, but there is a few difference in detailed video characteristics. (refer to [14]). In error concealment experiment, one MB scanning scheme, FMO0 have been used to order scanning MBs at encoder. In FMO0 scheme, missing MBs concentrate on a particular region of frame.

3.2.1 Image reconstruction

As mentioned above, if more than one frames are received at decoder successfully, error concealment with UPCA model can be applied to pixel domain in current image. When decoder receives a frame with no error, it can be used to update the existing PCA model. When a corrupted frame with missing MBs is received at decoder, it uses PCA method to reconstruct the missing pixels in the image. We present the reconstruction results using FMO0, in order to verify error concealment effects.

Figure 5 shows the reconstruction images using five previous frames. As can be seen, there includes no visual artifact in the reconstruction images. The satisfactory results are



Fig. 5 Error images (top) and the reconstructed images with UPCA model (bellow)

Table 3 PSNR of images reconstructed with UPCA using 1 to 5 frames after scene cha

Frames	Second	Third	Fourth	Fifth	Sixth
Foreman	38.6275	37.6930	37.2030	36.4364	37.1019
<i>Akiyo</i>	50.6786	50.6788	50.5679	41.3495	58.9952
<i>Carphone</i>	42.3100	36.500	41.2058	41.2491	34.8666
<i>Mobile</i>	41.2462	41.4494	42.0299	41.4392	42.4744

still obtained even when less than five previous frames are used to reconstruct the corrupted image with UPCA. The results have been analyzed with PSNR of the reconstruction images at Table 3. The all evaluations in PSNR have been calculated inside a whole frame. In table, second frames are reconstructed by UPCA model composing of one correctly received frame, 3rd frames by two frames, 4th frames by three frames, and so on. Regardless of using one or two frame, all concealed images show very fine PSNR of about 40 dB. Please note that UPCA model is proposed in order to reconstruct the corrupted images when scene change occurs. For second frame after scene change, the resulting reconstructions are 38.6275 dB, 50.6786 dB, 42.3100 dB and 41.2462 dB for foreman, *akiyo*, *carphone* and *mobile* frames respectively. It is seen that error concealment can be included in the image reconstruction process as soon as scene change occurs.

Rationality of UPCA model is analyzed in contrast to previous works [8, 14]. In Table 4, the reconstructed PSNRs of foreman have been produced by UPCA, PCA and UPC model respectively. In [8], UPC model was assigned to spatial error concealment. However, in our experiment, the UPC model could be composed of previous frames. Hence, the model is applied to temporal concealment and then is compared with UPCA model. The traditional PCA model yields the satisfactory result for reconstructing with 4 frames after scene change. However, for second frame with scene change, the result is poor PSNR of 2.7411 dB. For 3rd and 4th frames, the PSNRs are low similar to the previous. In UPC model, the serious artifact is improved more than PCA. However, when 4 frames are used to reconstruct the corrupted image, that is, when being supposed that there is no scene change, the method with decay factor $\alpha = 0.1$ has PSNR of 2dB lower than PCA model.

3.2.2 Comparison with several error concealments

Error concealments produce several artifacts due to difference place of error bits, that is, according to whether loss areas occur in a scene or in scene change.

First, Error concealment performance is tested on *foreman*, *akiyo*, *carphone* and *mobile* of QCIF sequences in terms of PSNR. Tables 5 and 6 show PSNRs for concealed frames.

Error frames have the distribution of corrupted pixels as in Fig. 5. As can be seen from Table 5, most of the methods yield high PSNR values on video sequences with low mobility of object or background, in except of PCA method. Especially, BMA has better PSNR over

Table 4 PSNR of *foreman* frames reconstructed with individual training

Algorithms	Foreman frames after scene change				
	Second	Third	Fourth	Fifth	Sixth
Updated PCA	38.6275	37.6930	37.2030	36.4364	37.1019
UPC	37.3978	35.5463	36.1020	33.4223	34.1312
Traditional PCA	2.7411	8.7497	18.1826	36.4364	37.1019

Table 5 Comparison of PSNRs of images concealed with individual temporal error concealment(1)

Algorithms	4th frames after scene change on Video sequences			
	Foreman	<i>Akiyo</i>	<i>Carphone</i>	<i>Mobile</i>
Copy-paste [4]	31.9589	45.2904	35.1217	31.7787
BMA	25.1264	52.4953	40.5904	28.3196
PCA	18.1826	15.0603	22.1474	19.9393
UPC	36.1020	42.7706	39.8456	36.9826
UPCA	37.203	50.5679	41.2058	42.0229
Corrupted frame	8.9408	13.9429	11	8.6137

UPCA, if two boundaries (top and bellow) of loss area are available and if the motions are linear and homogeneous (all parts moving in the same direction), for example, *akiyo*. However, the video quality reconstructed over foreman and *mobile* sequences, that include ROI and background of complex motions, are decreased greatly compared with Copy-Paste or UPCA. Context-based methods, such as PCA, UPC and UPCA, can have better flexibility by training models in ROI with non-linear and inhomogeneous movement.

UPCA produces PSNRs of 37.203 dB, 50.5679 dB, 41.2058 dB and 42.0229 dB over four test sequences, respectively. The values are more than 1.101 dB, 7.7973 dB, 1.36 dB and 5.0403 dB on foreman, *akiyo*, *carphone* and *mobile* sequences respectively compared with UPC model. It is obvious that UPCA represents video characteristics, such as high motion, various moving directions and zoom in/out, more adequately than UPC mechanism.

In Table 6, the resulting PSNRs are also similar to the above testing. In Table 5, PCA model yields low PSNR with the severe artifacts, because there are not yet enough frames needed for training, whereas, in Table 6, on which six frames are used for training, there are no artifacts. PCA model yields the same result as UPCA.

It shows that the idea, using Index Transform and image buffer for training structure to be close to scene contents, is right.

Second, the performance of UPCA is tested on video sequences with CIF format. Figure 6 shows corrupted frames and concealed frames on car-phone sequence with QP=28, display resolution of CIF format and GOP=30. In during transmission, bit error has appeared in 6th frame, so first corrupted frame is the 6th frame. When there is no bit error in car-phone sequence, PSNRs of 9th and 16th frames are 41.6232dB and 40.2022dB respectively, but when there occurs bit error with error rate of 10⁻³ in transmitting over radio channel, the PSNRs

Table 6 Comparison of PSNRs of images concealed with individual temporal error concealments(2)

Algorithms	8th frames after scene change on Video sequences			
	Foreman	<i>Akiyo</i>	<i>Carphone</i>	<i>Mobile</i>
Copy-paste[4]	32.6615	45.2132	28.6641	32.6615
BMA	30.0909	46.8989	37.3345	27.4456
PCA model	38.2261	45.5549	33.5883	38.2261
UPC model	38.0112	39.7979	32.4545	36.2546
UPCA	38.2261	45.5549	33.5883	38.2261
Corrupted frame	8.8915	13.9481	10.9054	8.8915

	Original frame	Corrupted frame	Concealed frame
9 th			
	41.6232dB	19.239dB	29.1423dB
16 th			
	40.2022dB	20.345dB	30.2032dB

Fig. 6 Lost frames caused by a bit error and frames concealed by UPCA

are fallen to 19.239dB and 20.345dB. As it can be seen from concealed frames, UPCA not only yields the result of satisfaction visually, but it also has PSNRs as high as QCIF video error con-cealment. CIF video concealment, however, needs calculation time longer than QCIF video concealment. In our test of car-phone, the time is almost 300ms per frame.

4 Conclusion

In this paper, we proposed a new error concealment algorithm using scene change detection in H.264 QCIF video sequences. First, the effective dynamic threshold and AC similarity metric are presented using the statistical characteristic of H.264 video image. With these parameters, a novel scene detection algorithm is built in JM15.0 encoder. The scene detection output is used to improve the performance of the satisfactory video quality and higher PSNR. In temporal error concealment, the updated PCA is exploited by using Index transformation and image buffer refreshing. UPCA is very effective for reconstructing the corrupted frames that are located at 1st to 6th frames after scene change. The lost images are concealed by using POCS algorithm with UPCA model. Experiments show that the proposed method is apparently superior to the related works.

References

1. Ameigeiras P, Navarro-Ortiz J, Andres-Maldonado P et al (2016) 3GPP QoS-based scheduling framework for LTE. *EURASIP J Wirel Commun Netw* 78:78
2. Behar R, Samet Y, Nossenson R (2015) Session management of variable video rate streaming session over multi-channel networks. In: 7th international conference on evolving internet, pp 26–33
3. Bo Y, Ng KW (2004) A novel motion vector recovery algorithm for error concealment in video transmission. *Consumer Commun Netw Conference*:621–623
4. Brandt J, Troitzky J, Wolf L (2008) Fast frame-based scene change detection in the compressed domain for MPEG-4 video. In: Second international conference on next generation *Mobile* applications, services, and technologies NGMAST'09[c], pp 514–520

5. Chen T, Zhang X, Shi Y-Q (2003) Error concealment using refined boundary matching algorithm, information technology. *Res Education*:55–59
6. Chen TP-C, Chen T (2002) Second-generation error concealment for video transport over error-prone channels. *Wiley Wireless Commun Mobile Comput* 6(2):607–624
7. De Bruyne S, De Neve W, De Wolf K, De Schrijver D, Van de Verhoeve P (2007) Temporal video segmentation on H.264 compressed bit streams. *Advances in Multimedia Modeling. MMM2007*. Springer, Singapore, pp 1–12
8. Dehghani M, Arshad K, MacKenzie R (2015) LTE-advanced radio access enhancements, a survey. *Wireless Pers Commun* 80:891–921
9. H.264 Software Coordination JM Software, ver. 15.0, 2009,1: <http://iphone.hhi.de/suehring/tml>
10. Lam WM, Reibman AR, Liu B (1993) Recovery of lost or erroneously received motion vectors. *ICASSP-93* 4:417–420
11. Li H, Liu G, Zhang Z, Li Y (2004) Adaptive scene detection algorithm for VBR video stream. *IEEE Trans Multimed* 6(4):624–633
12. Li X, Orchard MT (2002) Novel sequential error-concealment techniques using orientation adaptive interpolation. *IEEE Trans Circuit Syst Video Technol* 10(12):854–864
13. Nemethova O (2007) Error resilient transmission of video streaming over wireless mobile networks. *Technische Universit at Wien Osterreich* 3:77–88
14. Xie X, Zaitsev Y (2014) Scalable, MEMS-enabled, vibrational tactile actuators for high resolution tactile displays. *JMM* 24(12):11
15. Xie X, Zaitsev Y et al (2014) Compact, scalable, high-resolution, MEMS-enabled tactile displays. In: *Proceedings of solid-state sensors, actuators, and microsystems workshop*, June 8–12, pp 127–130
16. Yu G, Li Z, Wang S-Y, Shen L-S (2010) A GOP level video scene change detection algorithm in H.264 compression domain. *Chinese J Electron* 38(2):382–386
17. Zeng W, Gao W (2005) Shot change detection on H.264 compressed video. In: *IEEE international symposium oil circuits and systems, Kobe, Japan*. IEEE, pp 3459–3462
18. Zhou L (2015) Mobile device-to-device video distribution: theory and application. *Journal ACM Transactions on Multi-media Computing, Communications and Applications (ToMM)* 12(3):1253–1271
19. Zhou L, Wang XB, Wei T, Muntean G-M, Geller B (2010) Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks. *IEEE J Selected Areas in Commun* 28(3):409–419



Gwangmin Choe received the M.Sc.degree in automatic control in 2003 from Kim Il Sung University, Pyongyang, D.P.R.K., the Ph.D. degree in computer science in 2015 from Huazhong University of Science and Technology (HUST), Wuhan, China and Ph.D.degree in intelligent information processing in 2016 from Kim Il Sung University, Pyongyang, D.P.R.K.

His research interests include visual tracking, computer vision, image processing, computer graphics and virtual reality.



Cholman Nam received the B.S. and M.S. degree in information and communication engineering from Kim Il Sung University, D.P.R.K, in 1991 and 1994, respectively. In 2011, he received Ph. D. degree in information engineering from Institute of Information Science at Beijing Jiaotong University, China, where he joined the institute in 2007. In 2011, he received Ph. D. degree in multimedia communication engineering from Kim Il Sung University, D.P.R.K. He is currently in image processing, video compression, multimedia communication, 3G mobile communication research.



Changgon Chu received the B.S. and M.S. degree in information and communication engineering from Kim Il Sung University, D.P.R.K, in 1998 and 2003, respectively. He is currently in image processing, 3G mobile communication research.