



# Gaussian mixture-based generative approach for H.265/HEVC error concealment

Ali Radmehr<sup>1</sup> · Ali Aghagolzadeh<sup>1</sup> · Seyed Mehdi Hosseini Andargoli<sup>1</sup>

Received: 9 December 2022 / Revised: 4 July 2023 / Accepted: 10 August 2023 / Published online: 26 December 2023  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

## Abstract

Loss of data in video transmission over an error-prone channel is inevitable. Video error concealment is a powerful tool for resolving the effects of these errors. In this paper, we conducted several experiments with various video sequences to estimate the distribution of motion vectors (MVs) surrounding the corrupted area. As a result of the experiments, the proposed method suggests an innovative generative clustering approach using the Gaussian mixture model (GMM). The proposed algorithm first measures the GMM's parameters based on the available surrounding MVs. Then each MV is assigned to exactly one cluster. Next, each cluster's likelihood is calculated, and the one is chosen based on the maximum likelihood criteria. Finally, new MVs are generated for the chosen cluster, and the MV that can minimize the boundary distortion is selected for the corrupted block. Comparison concerning recent state-of-the-art techniques shows progress in PSNR and SSIM for H.265/HEVC. The proposed method improves the average PSNR by up to 5.67 dB and an average increase of 0.1135 in SSIM. Moreover, the computational complexity of the proposed algorithm is context-adaptive and shows better performance for the videos with relatively uniform motions over the missing areas.

**Keywords** Video error concealment · Gaussian mixture model · Motion vector clustering · H.265/HEVC

## 1 Introduction

Channel errors are unavoidable in communication networks. Thus, video transmissions cannot be robust. Moreover, losing a small fraction of a bitstream could cause severe quality degradation since the video sequences are highly correlated [1–3]. In other words, the propagated errors result in low-quality received video data because of the predictive coding and motion estimation [4–6].

Various methods are proposed in the literature to address the effect of channel errors, which can improve the robustness of video communication. The two main categories of these

methods are error resilience (ER) and error concealment (EC) techniques.

Forward error correction (FEC) and automatic repeat request (ARQ) procedures are among ER methods that attempt to guarantee a certain level of quality of service (QoS) in a communication network. However, the former raises the bandwidth of video data, while the latter cannot deal with multicasting or broadcasting scenarios. Therefore, the ER techniques are unsuitable for many video communications [5, 7–9].

EC technique is a post-processing method trying to recover the erroneous parts of a video frame using correctly decoded spatial/temporal information. These methods are categorized into three classes: spatial, temporal, and hybrid. This paper proposed a temporal error concealment (TEC) method to conceal the impaired blocks. TEC algorithms exploit the correlation among motion vectors (MVs) of adjacent blocks located in the current or previous frames [10, 11].

In H.264/AVC standard, EC methods conceal the corrupted macroblocks (MBs) in each erroneous slice. In H.265/HEVC, the damaged area of a frame is relatively large. Thus, most of the EC methods are not very effective [12].

✉ Ali Aghagolzadeh  
aghagol@nit.ac.ir

Ali Radmehr  
a.radmehr.ir@ieee.org

Seyed Mehdi Hosseini Andargoli  
smh\_andargoli@nit.ac.ir

<sup>1</sup> Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Mazandaran, Iran

Furthermore, the substantial progress in H.265/HEVC video coding has further raised several challenging issues that must be addressed. These issues are listed below.

- Traditional EC techniques only rely on MV's correlation, which is not necessarily valid, especially for frames with non-uniform motions. In H.265/HEVC, the number of neighboring MVs is not often sufficient to recover the lossy area, mainly when there is no flexible macroblock ordering (FMO) slicing available [12, 13].
- Many EC techniques are suggested for H.264/AVC but cannot be applied to H.265/HEVC real scenarios. The errors in H.265/HEVC damage multiple consecutive coding tree units (CTUs), which are wide regions of a frame [14].
- The size of a corrupted region is relatively large in H.265/HEVC. Therefore, using only available neighboring MVs leads to low-quality concealed video. Furthermore, SEC techniques perform well for an adequately small damaged region which is not the case for H.265/HEVC.
- There have been several EC algorithms proposed in the literature that exploit side information. However, they are inappropriate in many real scenarios since they are not standard compliant, i.e., the codec requires some modifications. Therefore, these algorithms have limited applications.

This paper conducted several statistical experiments to estimate the probability distribution of the corrupted region's neighboring MVs. The experiments show that the mixture of Gaussians can accurately model the probability distribution of these MVs. According to these experiments, we propose a novel GMM clustering approach for H.265/HEVC EC.

First, the proposed method extracts the available neighboring MVs of the damaged block from the current and previous frames. Second, we develop a novel clustering algorithm for motion data and assign each MV to exactly one cluster. Third, the most probable cluster is selected based on the maximum likelihood (ML) criteria. Fourth, the new MVs are generated for the chosen cluster, and the selected MV is the one that can minimize the boundary-matching distortion between the replacing and neighboring blocks.

The paper is organized as follows. Section 2 provides an overview of the state-of-the-art related work EC techniques. Section 3 is devoted to a concise description of the problem and motivation for using GMM. Then, the problem formulation and the proposed method are presented in Sect. 4, and comprehensive experimental comparisons are demonstrated in Sect. 5. Finally, Sect. 6 concludes this paper with remarks.

## 2 Related work

### 2.1 Using neighboring correctly received MVs

Many TEC methods are proposed in the literature using spatially or temporally adjacent MVs. For example, Li et al. modeled neighboring MVs of the lost block using the plane fitting, which shows the changing tendency in small regions of the damaged frame [15]. However, fitting a plane to the MVs for the heavily corrupted video frame is not meaningful and requires further modifications. Moreover, several TEC techniques exploit filtering schemes to enhance the efficiency of the recovered MVs. [16–18] utilized Kalman filtering (KF) to extract and refine the MV candidates.

Despite improving the accuracy of the MVs, they cannot correctly track nonlinear and non-Gaussian motions in a frame. Hence, a novel particle filtering (PF) EC approach was introduced in [19] to mitigate this problem by denoising the MVs derived from the boundary matching algorithm (BMA). This method works well for non-stationary MVs. However, it is rather complex. Also, Lin et al. used the available nearest MVs to predict the missing MVs. They proposed a novel TEC method for estimating MVs' weighting using disparities among MVs of available neighboring blocks [20]. This method is inefficient for the large, lossy regions since neighboring MVs are unavailable for inner blocks.

In [21], the authors introduce a novel EC algorithm that aims to enhance the performance of a commonly used MV extrapolation algorithm by considering partition decision information from the previous frame. By incorporating this additional information, the proposed method achieves improved objective and subjective results compared to the existing approaches. Notably, their algorithm demonstrates a more substantial improvement in HEVC compared to H.264/AVC videos since the larger partition decision space is existed in HEVC that accommodates a wider range of moving object sizes within a frame. Also, the proposed motion-compensated EC method for HEVC presented in [22] aims to enhance the visual quality. The method utilizes the classification map of residual energy associated with each MV and employs a block-merging algorithm. By analyzing and classifying the reliability of MVs, the scheme identifies and merges blocks with unreliable MVs, ensuring the preservation of the structure of moving objects and edge information. Notably, this method demonstrates its effectiveness in EC by employing a novel approach that ensures the preservation of moving objects and edge information, leading to an overall improvement in video quality.

In the field of video EC, there have been ongoing efforts to develop effective approaches for reducing blocky effects along block boundaries. One recent contribution in this area

is the work by [23], where they propose a novel MV refinement approach for video EC. Their approach combines a traditional boundary-matching method with a downhill simplex approach to fine-tune MVs, resulting in reduced blocky effects along the prediction unit block boundary. Importantly, their method achieves this improvement while minimizing computational costs. Through extensive simulations, the authors demonstrate the robustness and effectiveness of their proposed approach. This work provides valuable insights into addressing block artifacts in video EC and contributes to the growing body of research in this field. In [24] the authors introduce a novel weighted boundary matching EC scheme for HEVC. This approach leverages the depth information of coding units (CUs) and the partitioning decisions of prediction units (PUs) in reference frames to address lost slices. By utilizing the surrounding largest CUs (LCUs), the summed CU-depth weight is calculated to determine the concealment order of each CU. Furthermore, the co-located partition decision from the reference frame is used to guide the concealment of PUs within each lost CU. The sequence of PUs to be concealed is sorted based on the texture randomness index weight, prioritizing the PU with the largest weight. Ultimately, the best estimated MV is selected for the concealed PU. The results demonstrate that the proposed method exhibits superior visual quality compared to state-of-the-art techniques.

In [25] the authors proposed a method to reduce the prediction mismatch at the decoder. This algorithm selects an optimal subset of the MVs and transmits them as side information. However, this technique requires increasing the bitrate and adjusting the standard codec.

## 2.2 Using available pixel information

These methods estimate the lost pixels using the temporally or spatially adjacent pixels. For example, in [26], the authors proposed a sequential recovery SEC algorithm to reconstruct the missing pixels successively using an adaptive linear predictor. The algorithm's efficiency decreases for large, corrupted regions. Therefore, it is not suitable for H.265/HEVC.

In [27], the authors proposed a shape preservation technique exploiting different EC strategies based on the position of the corrupted blocks. However, measuring the object's boundaries is challenging and limits its application. In [28], the authors presented a homography-based registration algorithm exploiting available pixels around the corrupted region and their matched ones in the reference frame. However, this procedure is not practical when there is a low correlation between pixels to determine the registrations. Therefore, it is often impossible for large corrupted areas to find good matching points.

## 2.3 Using the generative model approach

Several statistical EC methods have been recently introduced in the literature. These techniques employ deep neural networks (DNNs) to conceal corrupted blocks. We categorize them as generative EC algorithms which attempt to produce convincing data close to transmitted data. In [29], the authors presented a network with convolutional and Long short-term memory (LSTM) layers for optical flow prediction. This method effectively predicts optical flow when adequately trained, but the results are unsatisfactory if the available neighboring MVs are non-uniform.

In [30], a generative adversarial network (GAN) is applied to the corrupted blocks. A couple of neighboring frames and the proposed mask are fed into the completion network. Then, the corrupted pixels are recovered based on the current and previous frames. However, even a small loss requires training for the lost area before it occurs. Therefore, using GAN networks for EC is time-consuming since it needs a training phase for multiple lossy regions in each frame.

## 3 Problem statement

The main goal of modern video coding standards is to increase coding efficiency while preserving the quality of the transmitted videos. However, H.265/HEVC does not support error resiliency options such as FMO. Moreover, the compression ratio of H.265/HEVC is much higher than the previous codecs. Therefore, H.265/HEVC is more susceptible to channel errors and needs more robust loss concealment techniques [14].

Furthermore, EC methods attempt to exploit the neighboring information of the corrupted regions. However, for H.265/HEVC, the correctly received information is far from the damaged area since the loss area is often relatively large. Therefore, the traditional SEC algorithms are ineffective for H.265/HEVC since insufficient adjacent pixels are available.

This paper devised a novel generative approach to address these challenging issues. Supplementary Fig. 1 illustrates a simplified diagram of the proposed algorithm. First, a set of motion data is derived from the available neighboring MVs. Second, a GMM is trained in the set of MVs which means estimating the correct set of parameters that characterize the GMM. Third, the proper data points are generated and the missing MVs are extracted based on the minimum boundary distortion criteria.

## 3.1 Derivation of motion information for H.265/HEVC EC

H.265/HEVC contains coding tree units (CTUs) or, to be more precise, CUs, whose sizes vary from LCU to Smallest

CU (SCU). The LCU and SCU sizes are bounded to  $64 \times 64$  and  $8 \times 8$  pixels, respectively. A packet loss in H.265/HEVC stream affects several consecutive CUs since the packet loss results in the loss of all the information associated with the packet.

Furthermore, each CU consists of coding blocks (CBs) that are further divided into PUs and transform units (TUs). Moreover, the motion information is located in PUs. As a result, a lost CU causes the loss of corresponding PUs. To address this issue, we propose a CU partitioning method that is demonstrated in Supplementary Fig. 2. The damaged CU is split into the  $8 \times 8$  sub-blocks. Each sub-block  $B_{ij}$  is indicated by a row and a column, i.e.,  $i$  and  $j$ , respectively. The concealment process starts from the upper left sub-block  $B_{11}$  and finishes at the bottom right sub-block  $B_{88}$ , row by row. Additionally, each sub-block  $B_{ij}$  extracts the available information from its neighboring  $8 \times 8$  blocks.

In addition, the proposed partitioning algorithm exploits the available MVs to propagate the nearest motion information into the lost subblocks. We also use the co-located MVs of the previous frame. It allows the algorithm to accurately track the changes in the MVs.

Furthermore, consider  $F_t$  and  $F_{t-1}$  to denote current and previous frames, respectively. Also, assume the set  $S_{i,j}$  contains the MVs of the neighboring prediction blocks (PBs). Supplementary Fig. 2 illustrates the location of these blocks, i.e., AL, A, AR, and L, which are denoted by  $\overrightarrow{MV}_{i-1,j-1}$ ,  $\overrightarrow{MV}_{i-1,j}$ ,  $\overrightarrow{MV}_{i-1,j+1}$ ,  $\overrightarrow{MV}_{i,j-1}$ , respectively.

Additionally, we add the MV of the previous frame's co-located PB and the average of MVs to the set  $S_{i,j}$ , which are denoted by  $\overrightarrow{MV}_{CO}$  and  $\vec{0}$ , respectively. Thus, the set  $S_{i,j}$  is represented by (1):

$$S_{i,j} = \left\{ \overrightarrow{MV}_{i-i_0,j-j_0} | (i_0, j_0) \in \left\{ (-1, -1), (-1, 0), (-1, +1), (0, -1) \right\} \cup \left\{ \vec{0}, \overrightarrow{MV}_{AV}, \overrightarrow{MV}_{CO} \right\} \right\} \quad (1)$$

In summary, three steps are required to create the set  $S_{i,j}$  for each sub-block  $B_{ij}$ . This procedure is outlined below:

- (1) Add MVs of adjacent PBs, i.e., AL, A, AR, and L, in  $F_t$  to the set  $S_{i,j}$ .
- (2) Add the average of the values in step 1 to the set  $S_{i,j}$ .
- (3) Find the co-located block in  $F_{t-1}$ , and add its corresponding MVs to the set  $S_{i,j}$ .

The set  $S_{i,j}$  contains motion information from the previous and current frames. This set is used for further analysis in Sect. 3.2.

### 3.2 Motivation for using Gaussian-mixture model

EC algorithms for H.265/HEVC suffer from the lack of neighboring available data around the corrupted regions of a frame. Therefore, a generative EC method is developed in this paper to address this issue. Additionally, as stated in Sect. 2.3, the training complexity of DNN for EC increases the computational costs dramatically. Also, it is inefficient for interactive applications since it cannot determine the lossy region a priori. Therefore, many consecutive frames are required to resolve this issue, which is not practical in real-time scenarios.

Consequently, one solution is to use a generative model that exploits the available neighboring temporal information of the damaged region. To develop such a generative model, we conducted several experiments to assess the distribution of the MVs for the damaged sub-blocks.

In this paper, we conducted several simulations to measure the probabilities of MVs, and the results are presented in Supplementary Tables 1 and 2. These experiments involved various types of video sequences, and we calculated the Root Mean Square Error (RMSE) for GMMs with different numbers of components, as shown in Supplementary Table 1. The video sequences are selected from publicly available standard video dataset from Xiph.org [31]: “Shields” (SH) and “Park run” (PR) (720p), “Four people” (FP), “Tractor” (TR) (1080p), “Rush field cuts” (RFC) (1080p) and “Crowd run” (CR) (1080p). To estimate the Probability Density Function (PDF) of the MVs, we utilized the set  $S_{i,j}$  for the corrupted subblock  $B_{ij}$ . Additionally, each LCU is denoted by  $LCU(r, c)$ , where  $r$  and  $c$  represent the row and column, respectively. The corrupted regions considered were  $LCU(10, 5)$  and  $LCU(15, 8)$  for the high-definition (HD) and full high-definition (FHD) video sequences, respectively. HD refers to a video resolution of  $1280 \times 720$  pixels and FHD specifically refers to a video resolution of  $1920 \times 1080$  pixels. The number of components in the GMMs is estimated using the Bayesian information criterion (BIC) approach, and the results are presented in Supplementary Table 2. Also, the mixing proportion and center position of each component are denoted as  $mp(x_0, y_0)$ , where  $mp$  represents the mixing proportion and  $(x_0, y_0)$  denotes the center coordinates of the corresponding component in Supplementary Table 2. These experiments demonstrate that the probability distribution of MVs can be effectively modeled by a GMM. Given the observations in the set  $S_{i,j}$  and the MVs' probabilities, the PDF is formulated by (2).

$$P_{i,j} = \sum_{n=1}^K L(C_n | S_{i,j}) P(C_n) \quad (2)$$

where  $P_{i,j}$  is the PDF value for the  $S_{i,j}$  and  $L(C_n|S_{i,j})$  is the likelihood of the  $n^{th}$  MV. Besides,  $P(C_n)$  is the probability of the  $n^{th}$  MV. The observed data in this experiment are assumed to be described by a GMM with  $K$  different normally distributed components. Accordingly, this paper proposes a GMM model for clustering the available neighboring MVs.

### 4 Proposed algorithm

In this paper, we propose a novel Gaussian mixture-based generative (GMG) clustering algorithm for H.265/HEVC EC. The proposed approach develops a probabilistic model, which enables the MV estimation for the corrupted blocks. Moreover, this method increases the number of MV candidates. Therefore, the proposed algorithm can accurately predict the missing MVs and enhance the EC performance for H.265/HEVC.

#### 4.1 Problem formulation for MV recovery

In the present manuscript, we aim to find the MVs of the corrupted PBs successively. This procedure improves the loss concealment accuracy because the MV candidates are generated based on the successfully received PBs. We define the MV of a corrupted sub-block as a 2D vector  $\overrightarrow{MV}$ , which is denoted in (3).

$$\overrightarrow{MV} = (MV_x, MV_y) \tag{3}$$

where  $MV_x$  and  $MV_y$  are  $x$  and  $y$  components of the corresponding MV, respectively. Also, assume that  $\overrightarrow{MV}$  is a 2D Gaussian random variable. Then, the set  $S_{i,j}$  consists of multivariate Gaussians with probability  $p(X_n|\theta)$ , which is formulated in (4).

$$p(X_n|\theta) = \sum_{k=1}^K w_k \times g(X_n|\bar{\mu}_k, \sigma_k) \tag{4}$$

where  $X_n$  is the  $n^{th}$  MV in the set  $S_{i,j}$ , i.e.,  $X_n \in S_{i,j}$ . Each Gaussian distribution is defined in  $\mathfrak{R}^l$  space and  $K$  is the number of components in a Gaussian mixture model.

Also,  $\theta = \{w_1, \mu_1, \sigma_1, w_2, \mu_2, \sigma_2, w_3, \mu_3, \sigma_3, \dots, w_K, \mu_K, \sigma_K\}$  where  $\mu_k \in \mathfrak{R}^l$  and  $\sigma_k \in \Sigma^l$  are the mean and covariance of  $X_n$ , respectively. It should be noted that a Gaussian distribution is completely determined by  $\sigma_k$  and  $\mu_k$ . Also,  $\sigma_k$  specifies the width of each Gaussian component. Additionally,  $w_k \in [0, 1]$  is a mixing probability that describes the effectiveness of each Gaussian function. Also, the sum of  $w_k$  must equal one, i.e.,  $\sum_{k=1}^K w_k = 1$ .

Therefore, the MVs' distribution, i.e.,  $(X_n|\theta)$ , is defined as the Gaussian distribution, i.e.,  $g(X_n|\theta)$ , which is completely determined by (5).

$$g(X_n|\theta) = \frac{1}{(2\pi)^{\frac{l}{2}}(\sigma_k)^{\frac{l}{2}}} \times \exp\left(-\frac{1}{2}(X_n - \bar{\mu}_k)^T \sigma_k^{-1}(X_n - \bar{\mu}_k)\right) \tag{5}$$

The goal of this paper is to find the optimal parameter  $\theta$  by maximizing the MVs' likelihood. In other words, the challenge is to determine which set of model parameters leads to the missing MVs' recovery.

Mathematically, it can be formulated as the minimization of a log-likelihood function, which produces the parameters  $\theta$ :

$$\min_{\bar{\mu}_1, \dots, \bar{\mu}_K, \bar{\sigma}_1, \dots, \bar{\sigma}_K} - \sum_{j=1}^N \log\left(\sum_{k=1}^K w_k g(X_n|\bar{\mu}_k, \sigma_k)\right) \tag{6}$$

where the mean  $\{\bar{\mu}_k\}_{k=1}^K$  and covariance  $\{\sigma_k\}_{k=1}^K$  should be estimated. Besides, the prior probability can be calculated based on the allocation of each Gaussian component. Moreover, the number of Gaussian components can be found by using the BIC procedure.

#### 4.2 Training the GMM with Expectation–Maximization

Training a GMM means determining the correct set of parameters  $\theta$ . There are two iterative techniques to solve the Eq. (6) the gradient descent (GD) algorithm and the Expectation–Maximization (EM) method. The former is an optimization algorithm for finding a local minimum of an equation, while the latter is a method to find maximum a posteriori (MAP) estimates of the parameters of the statistical models. EM is an appropriate algorithm since it is faster and more accurate than GD. Also, EM always converges to a  $K$ -component Gaussian mixture model, but GD often converges to a subset of the model [32]. EM method provides two successive steps for better estimating the model parameters, which do the following:

- (1) Initialization: choose initial guesses for the locations and shapes of the Gaussian components.
- (2) Iteration: repeat the two following steps until the stopping condition is met.
  - (a) Expectation step (E-step): the method calculates the function  $g(\cdot)$  using the current guess  $\theta$ . It evaluates the expectation value for the latent variables in the posterior distribution. In other words, this step finds the weights for the data points belonging to each cluster.

- (b) Maximization step (M-step): given the expectations in the previous step, the procedure updates the parameters, locations, and shapes by utilizing the weights of all the data points.

Furthermore, we identify the GMM parameters that can generate data points similar to what we observed. Mathematically, the EM method applies by the following procedure:

- (1) Initialization: initialize mean  $\mu_0$ , variance  $\sigma_0$ , and mixing probability  $w_0$ . Then, calculate the log-likelihood starting value.
- (2) E-step: evaluate posterior probability using the Bayes rule and assess the expectation  $\gamma_{n,k}$ :

$$\gamma_{n,k} = \frac{w_k \times N(\mathbf{X}_n | \bar{\mu}_k, \sigma_k)}{\sum_{k=1}^K w_k \times N(\mathbf{X}_n | \bar{\mu}_k, \sigma_k)} \quad (7)$$

- (3) M-step: re-estimate the parameters applying (7):

$$\bar{\mu}_k = \frac{\sum_{n=1}^N \gamma_{n,k} \mathbf{X}_n}{\sum_{n=1}^N \gamma_{n,k}} \quad (8)$$

$$\sigma = \frac{\sum_{n=1}^N \gamma_{n,k} (\mathbf{X}_n - \bar{\mu}_k)(\mathbf{X}_n - \bar{\mu}_k)^T}{\sum_{n=1}^N \gamma_{n,k}} \quad (9)$$

$$w_k = \frac{N_k}{N} = \frac{1}{N} \sum_{n=1}^N \gamma_{n,k} \quad (10)$$

$$\ln p(\mathbf{X} | \bar{\mu}, \sigma, w) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K w_k g(\mathbf{X}_n | \bar{\mu}_k, \sigma_k) \right\} \quad (11)$$

- (4) Calculate the log-likelihood:

Lastly, if there is no convergence in the fourth step, repeat the second step's procedure. Furthermore, the EM procedure is intrinsically a density estimation algorithm. It describes the distribution of some data by assuming a generative probabilistic model. In this paper, we exploit the EM technique to find the distribution of the MV candidates for MV clustering, which is explained in Sect. 4.3 in detail.

### 4.3 The Gaussian mixture-based MV clustering algorithm for H.265/HEVC EC

The proposed method is a Bayesian unsupervised clustering technique. In this scenario, each Gaussian component is a cluster that maximizes the likelihood of the data points. The proposed method generates new MVs based on the distribution parameters derived using the available neighboring motion information.

As it is illustrated in Supplementary Tables 1 and 2, MVs can be assumed to be generated from a mixture of  $K$  Gaussians. Moreover, the Bayesian information criterion (BIC) is used to solve the overfitting issue. It is a method for model selection and suppresses the possibility of too many components. Using this technique, GMG can determine the best-fitting model [33].

Supplementary Fig. 3 demonstrates the GMG method in summary. In this figure,  $i$  and  $j$  are the number of erroneous CUs and their corresponding subblocks, respectively.  $\mathbf{X}_n$  is drawn from a GMM with  $K$  Gaussian components given in (5) with density (4). Hence, one can estimate the  $j$ th sample generated by the  $k$ th Gaussian component. After parameters recalculation in (8), (9), and (10), the expected log-likelihood can be achieved by (11). Furthermore, the initial guess estimates the posterior parameters for the whole distribution. Therefore, we propose the initial guess from the set  $S_{i,j}$  that is more precise than a random guess. Moreover, the termination criterion is determined by a predefined cut-off point based on the mixing coefficients.

First, the erroneous slice is detected in a received frame, and the corrupted CUs are extracted. Second, the algorithm splits each erroneous CU into  $8 \times 8$  sub-blocks and builds the set  $S_{i,j}$  for each of them. Third, it calculates the expectation value  $\gamma_{n,k}$ , and the parameters are iteratively evaluated until only a small change occurs in the mixing coefficients:

$$\|w_l - w_{l+1}\| < \varepsilon \quad (12)$$

where  $w_l$  is the mixing coefficient for the  $l$ th iteration and  $\varepsilon$  is the threshold value for terminating the iteration. Fourth, the algorithm chooses the appropriate cluster  $C_{\text{chosen}}$  according to its  $N_k$  value, which determines the sufficient number of points assigned to the  $k$ th cluster. Finally, the latent variables, i.e., the missing data, are generated by using the selected cluster. Then, the new candidate MVs are generated based on the  $C_{\text{chosen}}$ , and the one which minimizes the boundary distortion is exploited for loss concealment. Algorithm 1 describes the proposed GMG procedure in detail.

## 5 Experimental results

This paper conducts a comprehensive assessment of the performance and complexity of the proposed algorithms compared with other state-of-the-art EC schemes such as Lin [21], Kim [23], Xu [24], and Chang [22] algorithms. The video encoding process uses the main profile with an input bit depth of 8. The CUs are limited to a maximum width and height of 64 pixels, and the maximum partition depth is set to 4. We do not use the decoding refresh type, and we choose a group of pictures (GOP) size of 4. Motion estimation is performed using the test zone (TZ) search method with a search

range of 64. To explain the relationship between lost packets, slices, and CUs, we divide the video frames into slices, and each slice is further divided into CUs. If a packet is lost during transmission, it can include one or more slices, and each slice can include one or more CUs. Therefore, the loss of a packet can result in the loss of one or more CUs, which affects the video quality in the corresponding region [34].

We utilized slice mode in our experiments to divide the video frames into smaller slices, each containing an integer number of CTUs and a maximum size of 1500 bytes per slice. This was done to align with the MTU constraint, which is common in video transmission over networks.

The coding structure is “*IPPP...I*” with only one intra-frame (I) after 50 inter-frames (P). Objective quality estimation typically uses PSNR as an evaluation metric adopted for the 200 frames in the experiments. The PSNR is determined between two video sequences by (13).

$$\text{PSNR} = 10 \log \frac{M^2}{\text{MSE}} \quad (13)$$

where  $M$  is the most considerable value of the signal, 255 in our experiments, and  $\text{MSE}$  is the mean square error difference between two video frames.

On the other hand, we exploit the structural similarity index measurement (SSIM) and MS-SSIM (multi-scale SSIM), which captures essential.

Also, the concealed video sequences were tested using VMAF (video multi-method assessment fusion) which incorporates spatial and temporal features by extracting them from the video frames and then aggregating them to produce a single feature value per frame. These features capture information related to spatial details and temporal variations within the video [14]. The simulations were conducted on a personal computer (PC) equipped with an Intel Core i5-6400 processor. The CPU operates with a frequency of 2.7 GHz and features four cores. The PC was also equipped with 16 GB of RAM. These hardware specifications provided the necessary computational resources to perform the simulations efficiently. The quantization parameter (QP) is set to 22, 27, 32, and 37. Additionally, the number of randomly generated lost packets to the total number of packets, i.e., packet loss rate (PLR), is 10 and 20%.

We have utilized H.265/HEVC reference software HM 18.0 for these experiments. We evaluated the proposed video encoding method on a publicly available video dataset from Xiph.org [31], which includes a diverse set of video sequences covering a wide range of content types. The proposed method was tested on five different video sequences: “Shields” and “Park run” (720p) HD signal format with 50 frames per second, “Four people” (720p) HD signal format with 60 frames per second, “Tractor” (1080p) FHD signal format with 25 frames per second, and “Crowd run” (1080p) FHD signal format with 50 frames per second, respectively. Also, we used 4:2:0 color subsampling throughout the simulations.

Supplementary Tables 3, 4, 5, 6, 7, 8, 9 and 10 exhibit the test results of the PSNR, SSIM, MS-SSIM, VMAF, and decoding time. With PLRs of 10, and 20%, the average PSNR of the proposed method is 5.67 dB greater than the Chang [22] algorithm. Furthermore, it achieves an average total gain of 0.1135 in SSIM than the other compared methods. According to these tables, the proposed method is more resilient than the other methods in reliability against error propagation. Compared to the other techniques, the better performance of the proposed approach is consistent with our expectations.

In other words, the clustering hypothesis for MV recovery exceeds the motion-compensated method, which is used in the Chang design. Due to the high compression ratio of H.265/HEVC, error propagation is severe in successive video frames. It causes rapid degradation for most EC techniques in H.265/HEVC since even minor error results in high degradation in the following GOP frames. The PSNR gain is prominently consistent in the proposed scheme, and the visual quality is better than the compared algorithms.

Supplementary Figs. 4 and 5 depict the visual comparison demonstrating the 121th frame of the video sequence “Park run” and the 55th frame of the “Tractor” sequence, respectively. These video sequences were tested at the PLR of 10 and 20%, with a corresponding QP of 27 and 37 for each experiment. In Supplementary Figs. 4 and 5b-e, the frames are concealed using the Lin [21], Kim [23], Xu [24], and Chang [22] algorithms. The comparison clearly shows that the proposed method significantly enhances the visual quality and preserves shapes and patterns better than the other algorithms. Furthermore, the proposed method accurately predicts the missing motion information, while the compared techniques fail to maintain the moving structures. These failures manifest as blockiness and deformed curves near the objects’ boundaries.

**Algorithm 1** Algorithm of GMG approach**Input:** Erroneous slice.**Output:** The MV candidates for the erroneous CUs.

```

1: for each missing CU do
2:    $j = 0$ .
3:   for each sub-block do
4:     Build the set  $S = \{X_1, X_2, \dots, X_N\}$ .
5:     Initialize parameters:  $\mu_0, w_0, \sigma_0$ .
6:     Calculate posterior probability  $\gamma_{n,k}$ .
7:     Recalculate the parameters:  $\mu_l, w_l, \sigma_l$ .
8:     Measure log-likelihood  $\ln p(X|\mu_l, w_l, \sigma_l)$ .
9:     if condition in equation (12) is true then
10:      Choose the cluster  $C_{chosen}$  with maximum
 $N_k$ .
11:      Generate new MVs using the
 $C_{chosen}$ 
12:      Select the MV candidate that minimizes
boundary differences.
13:      Recover the corrupted block using the
selected MV.
14:       $j \leftarrow j + 1$ .
15:      if  $j$  is less than or equal to  $L$  then
16:        go to 4
17:      else
18:        continue with the next CU.
19:      end if
20:    else
21:       $l \leftarrow l + 1$ .
22:      go to 6
23:    end if
24:  end for
25: end for

```

Note that the concealment errors propagate through consecutive slices spatially and temporally, which leads to severe quality degradation. Furthermore, when an erroneous slice is detected, the moving region of a video frame does not provide precise PU/TU information at the slice level. Consequently, applying only the neighboring MVs can lead to consecutive slice errors and blocking artifacts, especially when MVs' disparity increases. The proposed clustering model initially exploits the available neighboring MVs and then generates new MV candidates for loss concealment.

In conclusion, our proposed technique produces better quality results than the state-of-the-art generative technique presented in [30]. The performance of the proposed algorithm in [30] decreases with slice loss rate (SLR) growth due to the accumulation of errors between P-frames. However, in our proposed algorithm, the GMM is used to cluster the MVs of each CU separately, which reduces the error propagation to adjacent frames.

**5.1 Computational complexity analysis**

We calculate and compare the required decoding time for various EC techniques in Supplementary Tables 3, 4, 5, 6, 7, 8, 9 and 10. According to these tables, our proposed method performs well in various EC experiments. The other EC algorithms often consider a fixed number of MVs adjacent to the missing blocks, which limits their accuracy. For example, the EC method in [28] only exploits the available neighboring MVs. However, the proposed algorithm's computational complexity is context-adaptive and depends on the degree of MVs' uniformity around the corrupted region. In other words, the smaller the MVs' variations, the lower number of computations required.

**6 Conclusion**

This paper devised a novel generative EC algorithm for H.265/HEVC. We have done several experiments to find the probability distribution of the spatially and temporally neighboring motion information surrounding the damaged regions. As a result, it is observed that a GMM describes the PDF of the MVs correctly, leading to a novel MV classification approach. Moreover, the parameters of the GMM are calculated locally for each erroneous CU. Thus, the new MVs are generated based on the proposed model. Next, the erroneous blocks are concealed with the MVs, which minimizes the boundary distortion. The main attraction of our proposed algorithm is that it can generate new MVs for dealing with relatively large lossy regions of a video frame which is the case for the new video codecs. Thus, it is an important area for future study. According to the experimental results, the proposed method preserves the structure of moving objects in the concealed regions of the recovered frames. Furthermore, we calculated the decoding time for the proposed method and compared it with other state-of-the-art techniques. The results confirm that the proposed technique outperforms other EC schemes while balancing video quality and computational complexity.

**Supplementary Information** The online version of this article (<https://doi.org/10.1007/s11760-023-02739-0>) contains supplementary material, which is available to authorized users.

**Acknowledgements** This work was supported by Babol Noshirvani University of Technology through Grant program No. BNUT/389059/1401.

**Author's Contribution** AR contributed to conceptualization, methodology, software, writing—original draft. AA contributed to supervision, validation, resources, writing—review & editing. SMHA contributed to methodology, investigation, validation, writing—review & editing.

**Data Availability** All data generated or analyzed during this study are included in this published article, and its supplementary information files are available in the Zenodo repository, 110.5281/zenodo.8001714.



## Declarations

**Conflict of interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Tommasi, F., De Luca, V., Melle, C.: Packet losses and objective video quality metrics in H.264 video streaming. *J. Vis. Commun. Image Represent.* **27**, 7–27 (2015)
- International Standard ISO/IEC Recommendation ITU-T H.265, (2016)
- Hou, J., Chau, L.P., Magnenat-Thalmann, N., He, Y.: Low-latency compression of mocap data using learned spatial decorrelation transform. *Computer Aided Geom.* **43**, 211–225 (2016)
- Ababneh, J., Almomani, O.: Survey of error correction mechanisms for video streaming over the internet. *Int. J. Adv. Comput. Sci. Appl.* **5**(3), 050322 (2014)
- Lee, J.S., Ebrahimi, T.: Perceptual video compression: a survey. *IEEE J. Sel. Top. Sign. Process.* **6**(6), 684–697 (2012)
- Chellappa, R., Theodoridis, S.: Academic press library in signal processing. *Image and Video Processing and Analysis and Computer Vision*. Academic Press, UK (2017)
- Wu, H.R., Rao, K.R.: Digital video image quality and perceptual coding. CRC Press, USA (2017)
- Ko, M.G., Hong, J.E., Suh, J.W.: Error detection scheme for the H.264/AVC using the RD optimized motion vector constraints. *IEEE Trans. Consumer Electron.* **58**(3), 955–962 (2012)
- Bull, D.R.: Communicating pictures: a course in image and video coding. Academic Press, UK (2014)
- Usman, M., He, X., Xu, M. and Lam, K.M.: Survey of error concealment techniques: Research directions and open issues, in *Picture Coding Symposium (PCS)*, pp. 233–238, (2015)
- Cui, Z., Gan, Z., Zhan, X., Zhu, X.: Error concealment techniques for video transmission over error-prone channels: a survey. *J. Comput. Inf. Syst.* **8**(21), 8807–8818 (2012)
- Kazemi, M., Ghanbari, M., Shirmohammadi, S.: A review of temporal video error concealment techniques and their suitability for HEVC and VVC. *Multimed. Tools Appl.* **80**(8), 12685–12730 (2021)
- Korhonen, J.: Study of the subjective visibility of packet loss artifacts in decoded video sequences. *IEEE Trans. Broadcast.* **64**(2), 354–366 (2018)
- Kazemi, M., Ghanbari, M., Shirmohammadi, S.: The performance of quality metrics in assessing error-concealed video quality. *IEEE Trans. Image Process.* **29**, 5937–5952 (2020)
- Li, Y., Chen, R.: Motion vector recovery for video error concealment based on the plane fitting. *Multimed. Tools Appl.* **76**(13), 14993–15006 (2017)
- Mazinan, A.H., Amir-Latifi, A.: Applying mean shift, motion information and Kalman filtering approaches to object tracking. *ISA Trans* **51**(3), 485–497 (2012)
- Cui, S., Cui, H., Tang, K.: Error concealment via Kalman filter for heavily corrupted videos in H.264/AVC. *Signal Process Image Commun* **28**(5), 430–440 (2013)
- Lie, W.N., Gao, Z.W.: Video error concealment by integrating greedy suboptimization and Kalman filtering techniques. *IEEE Trans. Circ. Syst. Video Technol.* **16**(8), 982–992 (2006)
- Radmehr, A., Ghasemi, A.: Error concealment via particle filter by Gaussian mixture modeling of motion vectors for H.264/AVC. *Signal Image Video Process* **10**(2), 311–318 (2016)
- Lin, T.L., Wei, X., Wei, X., Su, T.H., Chiang, Y.L.: Novel pixel recovery method based on motion vector disparity and compensation difference. *IEEE Access* **6**, 44362–44375 (2018)
- Lin, T.L., Yang, N.C., Syu, R.H., Liao, C.C. and Tsai, W.L.: Error concealment algorithm for HEVC coded video using block partition decisions, in *2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)* pp. 1–5, (2013)
- Chang, Y.L., Reznik, Y.A., Chen, Z. and Cosman, P.C.: Motion compensated error concealment for hevc based on block-merging and residual energy, in *20th International Packet Video Workshop* pp. 1–6, (2013)
- Kim, D., Kwon, Y., Choi, K.: Motion-vector refinement for video error concealment using downhill simplex approach. *ETRI J.* **40**(2), 266–274 (2018)
- Xu, J., Jiang, W., Yan, C., Peng, Q. and Wu, X.: A novel weighted boundary matching error concealment schema for HEVC, in *25th IEEE International Conference on Image Processing (ICIP)* pp. 3294–3298, (2018)
- Carreira, J.F.M., Assunção, P.A., de Faria, S.M.M., Ekmekcioglu, E., Kondo, A.: A two-stage approach for robust HEVC coding and streaming. *IEEE Trans. Circ. Syst. Video Technol.* **28**(8), 1960–1973 (2017)
- Liu, J., Zhai, G., Yang, X., Yang, B., Chen, L.: Spatial error concealment with an adaptive linear predictor. *IEEE Trans. Circuits Syst. Video Technol.* **25**(3), 353–366 (2015)
- Hsia, S.C., Hsiao, C.H.: Fast-efficient shape error concealment technique based on block classification. *IET Image Process* **10**(10), 693–700 (2016)
- Chung, B., Yim, C.: Bi-sequential video error concealment method using adaptive homography-based registration. *IEEE Trans. Circuits Syst. Video Technol.* **30**(6), 1535–1549 (2019)
- Sankisa, A., Punjabi A., Katsaggelos A.K.: Video error concealment using deep neural networks, in *25th IEEE International Conference on Image Processing (ICIP)*, pp. 380–384, (2018)
- Xiang, C., Xu, J., Yan, C., Peng, Q., Wu, X.: Generative adversarial networks based error concealment for low resolution video, in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1827–1831, (2019)
- C. Montgomery, Xiph.org video test media (Derf’s collection), the Xiph open source community. <https://media.xiph.org/video/derf/>. Accessed 7 Mar 2023
- Yu, L., Yang, T., Chan, A.B.: Density-preserving hierarchical EM algorithm: simplifying Gaussian mixture models for approximate inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(6), 1323–1337 (2019)
- Ding, J., Tarokh, V., Yang, Y.: Model selection techniques: an overview. *IEEE Signal Process Mag.* **35**(6), 16–34 (2018)
- Bossen, F.: Common test conditions and software reference configurations, document JCTVC-L1100. JCT-VC, San Jose (2012)
- Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. *J. Comput Commun.* **7**(3), 8–18 (2019)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.