A Robust Resolution-Enhancement Scheme for Video Transmission over Mobile Ad-Hoc Networks

Liang Zhou, *Student Member*, *IEEE*, Baoyu Zheng, *Member*, *IEEE*, Anne Wei, Benoît Geller, Jingwu Cui

Abstract

Historically, Error-Resilient (ER) video transmission and Super-Resolution (SR) image reconstruction techniques have evolved separately. In this paper, we propose a coordinated application of ER and SR to enhance the resolution of image transmitted over mobile ad-hoc networks. In order to combat error propagation, a flexible multiple description coding method based on shifted 3-D SPIHT (3-D Set Partitioning In Hierarchical Trees) algorithm is presented to generate variable independent descriptions (substreams) according to the network condition. And then, a novel unequal error protection strategy based on the priority level is provided to assign a higher level of error protection to more important parts of bitstream. Moreover, a robust SR algorithm is proposed in the presence of different kinds of packet loss rate to enhance the image resolution. Experimental results indicate that the proposed robust resolution-enhancement scheme outperforms the competing methods from the aspects of PSNR (Peak-Signal-to-Noise Ratio) and visual quality under different packet loss rates.

Index Terms

error-resilient; super-resolution; mobile ad-hoc networks; resolution-enhancement

L.Zhou, B.Zheng and J.Cui are with the Electrical Engineering Department, Shanghai Jiao Tong University, Shanghai, China and the Institute of Signal Processing and Transmission, Nanjing University of Posts and Telecommunications, Nanjing, China; A. Wei is with the Laboratory of CEDRIC, Conservation National des Arts et Métiers, Paris, France; B.Geller is with the Laboratory of LEI, Ecole Nationale Supérieure des Techniques Avancées and with the Laboratory of SATIE, Ecole Normale Supérieure (E.N.S.) Cachan, Paris, France.

I. Introduction

In recent years, mobile ad-hoc networks (MANETs) has emerged as one of the high growth applications of wireless communication technology, and the interest in the transmission of video over MANETs has increased dramatically. However, this error-prone network is packet based where many potential reasons may result in packet loss which has a devastating effect on the visual quality of images at the receiver. Furthermore, in most electronic imaging applications, image with high resolution (HR) is desired and often required because HR image can offer more details that may be critical in various applications. Unfortunately, it is challenging to provide HR image transmitted over MANETs where no Quality of Service (QoS) is guaranteed at the network level. In order for HR, it would be essential to solve the following primary technical challenges:

• Trade-off between coding efficiency and error resilience. Most current standardized video codecs, including MPEG-2/4 and H.263/4 are designed to achieve high compression efficiency at the expense of error resilience. The coding efficiency in these codecs is achieved by using motion-compensated prediction to reduce the temporal and statistical redundancy between the video frames. This brings a severe problem, namely error propagation, where errors due to packet loss in a reference frame propagate to all of the dependent frames leading to visual artifacts that can be long lasting and annoying [1].

Therefore, Error resilience is needed to achieve robust video transmission. One strategy to achieve resilience is to insert redundant information systematically into compressed video signals so that the decoder can compensate transmission errors. The redundant information can be error correction codes [2], [3] or multiple descriptions [4], [5]. The former one combined with layered coding can provide good performance in prioritized networks while the latter is suitable for delivery over multiple channels to enhance reliability. However, error resilience is achieved at the expense of coding efficiency in both methods. Another way can be achieved with feedback mechanism to request retransmission or adjust encoding modes according to conditions. The methods proposed in [6], [7] rely on feedback from the decoder and are, therefore, application-limited.

• Enhance resolution under the scenario of packet loss. In order for HR image, one promising approach, which is called super resolution (SR), uses signal processing techniques to obtain a

HR image from observed multiple low resolution (LR) images [8]. In the past few decades, a variety of SR methods have been proposed for estimating the HR image from a set of LR images without taking into account the packet loss during the transmission. The critical requirement for traditional SR approach is that the observations contain different but related views of the scene [8], however, it can not be guaranteed under the framework of error-prone networks where the packet loss destroys the correlation between the related views of scene. Therefore, how to apply the SR approach to the packet loss scenario is still an open problem. Here, it is necessary to differentiate error concealment (EC) with SR. EC hides or recovers the errors by using correctly received image information without modifying source or channel coding schemes [9], which can only produce a visually acceptable (rather than exact) image from the available data, and can not enhance the physical resolution of the image. While SR extracts the exact detail information hidden among the different but related video frames to enhance the image resolution.

To meet these challenges, in this paper, we propose an entire scheme to get HR video transmitted over MANETs by integrating efficient ER strategy with robust SR algorithm, which not only provides relatively efficient compression and transport performance but also provides robust resolution-enhancement performance in the presence of various packet loss rates. An outline of the remainder of this paper is as follows. In section II, we describe the whole system and provide the related technique preliminaries used in this work; in section III, we present an adaptive error-resilient strategy to reduce the distortion due to the packet loss; and then, a robust SR algorithm to enhance the resolution of received image is proposed in section IV; section V provides the simulation results and compares the performance of the proposed scheme with other comparison systems. Finally, we give some concluding remarks.

II. PRELIMINARIES

In this section, we first overview system framework of the video transmission and processing, and then present some related technical preliminaries used in this work, such as shifted 3-D SPIHT algorithm and multiple description coding.

A. System Overview

The total architecture of video transmission and processing illustrated in Fig.1 is composed of three processes, such as image degradation, image transmission over error-prone networks and

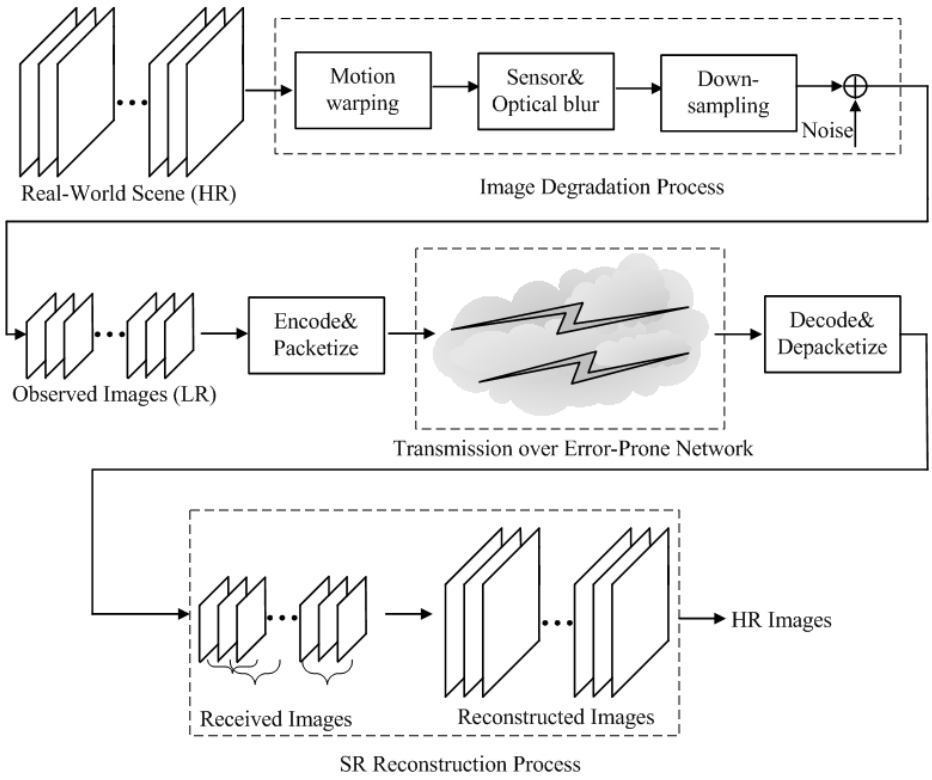


Fig. 1. The total architecture of video transmission and processing

the image SR reconstruction process.

Generally, all of the video sequences we observe are LR images comparing to the real-world scenes which are viewed as the original HR images. That is because the degradation process affects the quality of images acquired by digital video camera which results from the lens' physical limits, such as motion warping, optical blur and additive noise. In addition, these LR images are usually down-sampled convenient for transmission or storage. Next, the observed LR images are encoded and packetized preparing for transmission over error-prone networks. In this paper, the method of encoding and packetizing is based on the shifted 3-D SPIHT algorithm to generate variable descriptions (substreams) at the sender, and different descriptions employ different error protection strategies according to its priority. As to the MANETs system, suppose that there are n mobile nodes in the system and n' $(1 \leq n' < n)$senders (sources) stream complementary substreams to a single receiver (destination) over different paths. In this system, sender-n'' streams substreams-n'' to the receiver over path-$n''(n'' = 1, 2..., n')$.At the receiver,

the received images can be reconstructed after depacketizing and decoding the received data.

B. Shifted 3-D SPIHT Algorithm

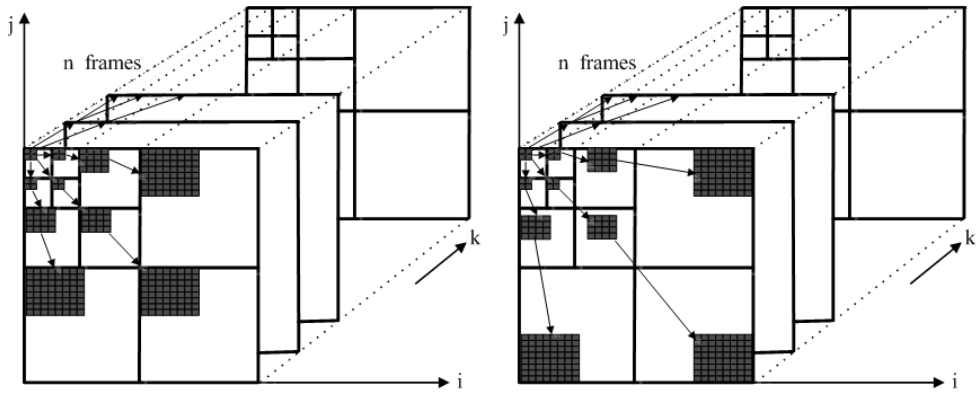
Wavelet zero-tree image-coding technique was first addressed by Shapiro [10], and further developed by Said and Pearlman [11], and have provided unprecedented high performance in image compression with low complexity. Later, extended 2-D zero-tree wavelet coding (EZW) by [11] had been introduced to three dimensions (3-D EZW) by Chen and Pearlman [12], and had shown promise of an effective and computationally simple video-coding system without motion compensation, obtaining excellent visual results. And then, Kim and Pearlman developed the 3-D SPIHT [13] coding algorithm based on the 3-D EZW mentioned in [12].

3-D SPIHT algorithm provides excellent rate-distortion performance along with a low encoding complexity, and spatiotemporal trees are defined as groups of wavelet transform coefficients organized into trees rooted in the lowest frequency subband and the spatially and temporally related coefficients in the higher frequency subbands, which is helpful to reduce the error propagation [13]. Fig.2(a) shows how coefficients in a 3-D transform are related according to their spatial and temporal domains. The parent-children linkage except at the highest and lowest pyramid levels (which do not have offspring) is:

$$\begin{aligned} O(i,j,k) = \quad & \{(2i,2j,2k),(2i,2j+1,2k),(2i,2j,2k+1), \\ & (2i+1,2j,2k),(2i+1,2j+1,2k),(2i+1,2j,2k+1), \\ & (2i,2j+1,2k+1),(2i+1,2j+1,2k+1)\} \end{aligned} \tag{1}$$

where $O(i,j,k)$ represents a 3-D of coordinates of all the offspring at node (i,j,k).

Although the SPIHT-coded bitstream has many advantages, it is also sensitive to data losses because of the dependence among wavelet coefficients in constructing a significance map. Motivated by the packetized 2-D SPIHT coding algorithm [14] partitions the wavelet coefficients into independent packets to combat error propagation by shifting 2-D wavelet tree structure, here we apply it to 3-D case by shifting the 3-D wavelet tree structure. The essential aim is the wavelet coefficients from different sub-bands are interleaved to form independent packets that can be decoded independently. The formation of the shifted wavelet trees is shown in Fig.2(b). It should be noted that although a lot of multi-interlaced partitioning and packetizing methods have been



(a) Traditional 3-D SPIHT (b) Shifted 3-D SPIHT

Fig. 2. Structure of the spatiotemporal relation of 3-D SPIHT

presented in recent years [23], they could enhance the encoding burden greatly. Therefore, we don't employ them here.

C. Multiple Description Coding

As to the way to protect data from packet losses induced by the error-prone channels, one of the most common ways is to add the redundant information at the bitstream so that the original video can be recovered in presence of the packet loss, and the code should be devised in such a way that the decoder is able to recover the lost information or conceal the irrecoverable errors.

One such popular approach is multiple description coding (MDC), which is similar in spirit to the multiple-substream approach. The fundamental principle of MDC is to generate multiple correlated descriptions of the source such that each description approximates the source with a certain level of fidelity [15]. MDC does not impose any dependency among its descriptions so that each extra successfully received description improves the quality further regardless of what has been received so far.

The benefits of using MDC in video streaming can be further amplified when MDC is combined with path diversity (PD) [16]. In this approach, each substream (or description) is explicitly transmitted over an independent path to receiver. PD exploits the fact that the probability of having all the paths simultaneously congested is relatively low. As a result, the use of PD in video streaming can achieve higher throughput and increase tolerance to packet loss.

III. Adaptive Error-Resilient Strategy

In this section, a novel error-resilient strategy is proposed based on partitioning the GOF (group of frames) into variable substreams with different priority levels adapting to the current network condition.

A. Unequal Error Protection

Progressive bitstreams provide a natural basis for unequal error protection (UEP), by which perceptually more important parts of the bitstream are assigned a greater level of error protection [17]. In this work, we propose a novel UEP based on the expected lifetime of the path to guarantee the important part (high priority) of substreams are received at the receiver. That is to say, important substream corresponds to the "good" path.

Here, we set the priority of the substreams according to the importance of the subband derived from the wavelet decomposition.In general, the LL_k provides more visual information than HH_k,HL_k and LH_k(where k denotes the decomposition level),so we set priority level of LL_k as 2^{2k},while the other three subbands are $2^{2(k-1)}$. To the priority of path, we use the ratio of remaining energy (E) to transmit power (P_T) to denote the expected lifetime of the node. So the higher the ratio is the more lifetimes of the node. In order to maximize the lifetime of the whole network system, to an arbitrary node i, the criterion of its next hop should satisfy that

$$R(i) = \max(E_i/P_{T_i}) \tag{2}$$

where $R(i)$ denotes the maximum expected lifetime for node i corresponding to the optimal energy transmission to next hop which locates on the desired path. Assuming that the discovered path j consists of nodes such as $i, i+1, ...i+n$ ($n \in N^+$) , R_j, the energy consumer function of path j, here we define

$$R_j = \min\{R(i), R(i+1), ...R(i+n)\} \tag{3}$$

This means that the minimum expected lifetime of each node determines the whole lifetime of the path. Because once one of the nodes in the path dies, the whole path dies as well. The more R_j the higher priority of the path, and correspond to higher level of encoded substream. Note that, in order to realize the proposed UEP, we modify the traditional DSR (Dynamic Source Routing) Protocol as follows: besides adding the node's ID to the request packet, each node also

adds the information of transmit power and remaining energy to the request packet, if the node receives packet, the information of the received power is also added to the packet. So when the sender node receives the request packet, the packet should consist of the route nodes' ID, the remaining energy of each route node and each node's transmit power and received power.

B. Flexible MDC

We focus on how to design the MDC according to the networking condition: one is how many descriptions are needed to guarantee the reconstructed video quality as well as keep the total bitstream as little as possible, while the other one is how to distribute these encoded data to the determinate substreams.

Obviously, the more substream the more data received at the destination, but it is infeasible for the practical wireless network. Here, we give an oversimplified method to compute the minimum needed substream number according to the packet loss rate (P_L) of the obtained channels.

With regard to the channel model, we use a two-state Markov model (i.e. Gilbert model) to simulate the bursty packet loss behavior [18]. The two states of this model are denoted as G (good) and B (bad). In state G, packets are received correctly and timely, whereas, in state B, packets are assumed to be lost. This model can be described by the transition probabilities p from state G to B and q from state B to G. Then the average P_L is given by

$$P_L = \frac{p}{p+q} \tag{4}$$

And the average length of burst errors L_B is given by

$$L_B = \frac{1}{p} \tag{5}$$

So the average channel P_L of path j is $p_j/p_j + q_j$. Suppose that all the transmit probability of potential paths are independent, the minimum number of substreams L is

$$\min_{L \in N^+} \prod_{j=1}^{L} (\frac{p_j}{p_j + q_j}) \leq Thr \tag{6}$$

where Thr is the tolerance P_L threshold which depends on the practical application requirements. Usually, its value varies from $10^{-1} \sim 10^{-4}$ [22], in this paper, we set $Thr = 10^{-2}$.

In the case of data distribution, it also contains two aspects: one is the decision of the wavelet decomposition level, and the other one is the data distribution among these determinate paths.

As to this point, we employ three basic principles:

• High priority level of the source data corresponding to high priority of the path, which is called equity principle.

• The whole transmission system only guarantees the most important source data, which has the highest priority level.

• As to other parts of the data, we use the best-effort strategy to transmit.

If a K-level dyadic wavelet decomposition is used, the number of wavelet coefficients in the k ($0 < k \leq K$) level spatial-frequency subband $C \in \{LL, LH, HL, HH\}$ is given by

$$C_k = (\frac{X}{2^k}) \times (\frac{Y}{2^k}) \tag{7}$$

where X and Y represent the frame width and height respectively. The level of the wavelet decomposition is decided by the expected lifetime of the highest priority path R_H. The essential requirement is that the GOF of the most important data should be guaranteed to transmit from source to destination over the highest priority path, so

$$\min_{K \geq 1} \lceil \frac{(LL_K/XY) \times r_b \times N_{GOF}}{F \times S} \rceil \leq R_H \tag{8}$$

where LL_K/XY denotes the ratio of the most important part in the total data streaming; N_{GOF} is the frame number of one GOF; r_b is the total source coding rate in bytes/s; F is the frame rate in frames/s; S is the packet size in bytes.

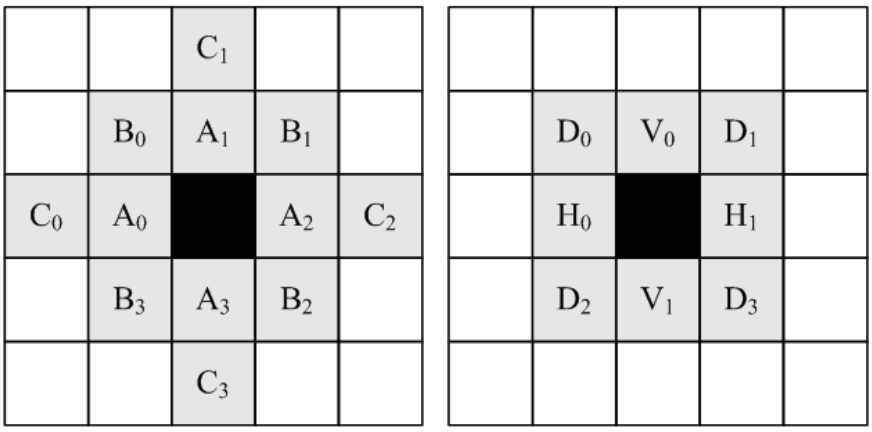
The fundamental distribution rule of data distribution is that the highest priority level data should be distributed at the each path while the others only distributed once at the path. Assuming $C' \in \{LH, HL, HH\}$, and expected life of path j is R_j, so each path should satisfy that

$$\lceil \frac{((LL_K + \sum_{k=K}^{1} C'_k)/XY) \times r_b \times N_{GOF}}{F \times S} \rceil \leq R_j \tag{9}$$

If there is the $C'_k (1 \leq k \leq K-1)$ which not transmitted by this path, it is transmitted by the other potential paths with lower priority level.

IV. Robust Super-Resolution Algorithm

Although adaptive error-resilient video transmission can reduce the transmission distortion, it can not exterminate the packet loss which has a devastating effect on visual quality. In this section, we propose a robust SR algorithm taking into consideration the various packet loss



(a) Mask used in low-frequency subband (b) Mask used in high-frequency subband

Fig. 3. Labeling of the weights used in calculation at the missing sample.

scenarios to enhance the resolution of received image. At first, we propose a simplified estimator to estimate the lost wavelet coefficients. And then, a series of convex sets which extract the exact detail information hidden among the adjacent images are constructed by taking advantage of the correlation of the wavelet coefficients.

A. Simplified Estimator

Motivated by the model of [19], here we propose a simplified estimator to estimate the lost coefficients. Since LL subband provide basic information (low frequency) for original image, missing samples in approximation subband LL are estimated prior to processing the high frequency subbands in LH, HL, and HH. Therefore, different strategies are employed to deal with the different kinds of packet loss.

In the case of the LL subband packet loss, as the correlation of the wavelet coefficients is much less than the high-frequency subbands, it is difficult to use a common interpolation method to estimate the lost coefficients precisely. As a result, we propose a low-complexity solution which can replace the commonly used single interpolation mask by a set of masks. In this case, each of them is adapted to a specific direction of the predominant spatial correlation. Here we define the correlation from the aspects of horizontal and vertical direction respectively and use general 5×5 interpolation mask presented in Fig.3(a). The parameters $A_0 \sim A_3$, $B_0 \sim B_3$ and $C_0 \sim C_3$ are weights for the corresponding neighboring coefficients. The weights in each mask are chosen

according to the predominant correlation direction (horizontal or vertical) and according to the degree of correlation (strong or weak). In order to measure the predominant correlation direction and the degree of correlation we propose the following algorithm:

After the wavelet decomposition, the sender bi-linearly interpolates each scaling coefficients as if it is lost. This process is done twice: a first approximation is obtained by using a horizontal interpolation; a second approximation by using a vertical interpolation. The horizontal interpolation pass only calculates the mean value of the left and right neighbors and the vertical interpolation pass simply calculates the mean value of the upper and lower neighbors. In this way, two approximations of the subband LL_k are created: LL_{k_h} and LL_{k_v}. The sender then calculates the sum of the absolute differences (SAD) values for these two subbands compared to the original LL_k subband: SAD_v and SAD_h,

$$SAD = \sum_{i \in W} |I(i) - O(i)| \tag{10}$$

where I and O denote the interpolated and original coefficients respectively, and W is the comparison region, here is a 5×5 mask. We define a directional correlation measure as follows:

$$G_{h-v} = SAD_h - SAD_v \tag{11}$$

The value of G_{h-v} tells us how much one direction is better for interpolation than the other one. We define five classes: (a) strong horizontal correlation ($G_{h-v} > A$), (b) weak horizontal correlation ($A \geq G_{h-v} > B$), (c) isotropic ($B \geq G_{h-v} \geq -B$), (d) weak vertical correlation ($-B > G_{h-v} \geq -A$) and (e) strong vertical correlation ($-A > G_{h-v}$). The resulting values of A and B are $A = 15$ and $B = 5$.

• Mask (a): $A_0 = A_2 = 30$, $A_1 = A_3 = 20$; $B_0 = B_1 = B_2 = B_3 = -8$; $C_0 = C_3 = -5$, $C_1 = C_2 = -4$;

• Mask (b): $A_0 = A_2 = 35$, $A_1 = A_3 = 25$; $B_0 = B_1 = B_2 = B_3 = -7$; $C_0 = C_3 = -5$, $C_1 = C_2 = -4$;

• Mask (c): $A_0 = A_1 = A_2 = A_3 = 35$; $B_0 = B_1 = B_2 = B_3 = -10$; $C_0 = C_1 = C_2 = C_3 = -4$;

• Mask (d): $A_0 = A_2 = 25$, $A_1 = A_3 = 35$; $B_0 = B_1 = B_2 = B_3 = -7$; $C_0 = C_3 = -4$, $C_1 = C_2 = -5$;

• Mask (e): $A_0 = A_2 = 20$, $A_1 = A_3 = 30$; $B_0 = B_1 = B_2 = B_3 = -8$; $C_0 = C_3 = -4$, $C_1 = C_2 = -5$;

In the case of missing samples in LH, HL, and HH subbands, we label the weighting factors as shown in Fig.3(b). That is, weights connecting horizontal neighbors are labeled H_0 and H_1, those connecting vertical neighbors V_0 and V_1, diagonal neighbors D_0, D_1, D_2, and D_3. Note that the direction of low-pass filtering is the direction in which high correlation of samples may be expected. To exploit the correlation between the wavelet coefficients in high-frequency subbands, a linear estimated model is adopted to estimate the missing samples. So the weighting factors can be set that

- Mask (LH): In LH, $H_0 = H_1 = 6$, $V_0 = V_1 = D_0 = D_1 = D_2 = D_3 = 1$;
- Mask (HL): In HL, $V_0 = V_1 = 6$, $H_0 = H_1 = D_0 = D_1 = D_2 = D_3 = 1$;
- Mask (HH): In HH, $D_0 = D_1 = D_2 = D_3 = 4$, $H_0 = H_1 = V_0 = V_1 = 1$;

The value of G_{h-v} can be sent as three bits (which represent the five classes) in each packet. This causes nearly no additional transmission overhead and no computational overhead for the receiver. All calculations are done by the sender and for the scaling coefficients only. According to a number of experiments, for a wavelet decomposition with depth is 3, the scaling coefficients only 1.6% of the total number of the coefficients. It should be emphasized that these factors are obtained empirically on the basis of considerable experimentation.

B. Projection onto Convex Sets

In this subsection, a projection procedure is utilized to extract information hidden in a group of video frames to update the wavelet coefficients. Since these coefficients correspond to the high frequency information in the spatial domain, the exacted fine features from other frames augment the individual LR frame to a HR frame. The constructed convex set should satisfy the following two points: 1) enhance the resolution of the received images, 2) reduce the artifacts generated during the projection process. We account for them in the following and present the relevant notation descriptions in Table I.

Let $f'(n,t) \in L^2(R^2)$ denotes the original HR image[1], which can be expanded as a sum of approximation component in the LL band and three detail components in the LH, HL and HH

[1]the original value of $f'(n,t)$ can be achieved by expanding the LR image $f(n,t)$ using bi-linear method

TABLE I

NOTATION

Symbol	Description
$f(n,t)$	observed low-resolution frame
$f_c(n,t)$	the current constructed low-resolution frame
$f_r(n,t)$	the reference low-resolution frame
$f'(n,t)$	the original high-resolution frame
$\widetilde{f}'(n,t)$	the estimation of the original high-resolution frame
C_{inter}	the convex set using the inter-frame projection
C_{intra}	the convex set using the intra-frame projection

bands.

$$\begin{aligned} f'(n,t) \equiv & \underbrace{\sum_{j,k\in Z} \alpha_{j,k}\psi_{j,k}(n,t)}_{LL} + \underbrace{\sum_{j,k\in Z} \beta^h_{j,k}\varphi^h_{j,k}(n,t)}_{LH} \\ + & \underbrace{\sum_{j,k\in Z} \beta^v_{j,k}\varphi^v_{j,k}(n,t)}_{HL} + \underbrace{\sum_{j,k\in Z} \beta^d_{j,k}\varphi^d_{j,k}(n,t)}_{HH} \end{aligned} \tag{12}$$

where $\varphi^h_{j,k}(n,t)$, $\varphi^v_{j,k}(n,t)$ and $\varphi^d_{j,k}(n,t)$ are the translated wavelets at the next coarse scale level that capture detail information in the horizontal, vertical and diagonal directions respectively, and $\psi_{j,k}(n,t)$is the translated coarse scaling function. The approximation and detail wavelet coefficients are given by

$$\alpha_{j,k} = \iint f'(n,t)\psi_{j,k}(n,t)dndt \tag{13}$$

$$\beta^h_{j,k} = \iint f'(n,t)\varphi^h_{j,k}(n,t)dndt \tag{14}$$

$$\beta^v_{j,k} = \iint f'(n,t)\varphi^v_{j,k}(n,t)dndt \tag{15}$$

$$\beta^d_{j,k} = \iint f'(n,t)\varphi^d_{j,k}(n,t)dndt \tag{16}$$

Let $f_c(n,t)$ the current constructed LR frame and $f_r(n,t)$ be the reference LR image. The convex set can be defined from the above four aspects:

$$C_{inter} = \{\widetilde{f}'(n,t) | C^{\alpha}_{inter}, C^h_{inter}, C^v_{inter}, C^d_{inter}\} \tag{17}$$

where

$$C_{inter}^{\alpha} = \{\iint \widetilde{f}'\psi_{j,k}dndt = f_c\} \tag{18}$$

$$C_{inter}^{h} = \{\iint \widetilde{f}'\varphi_{j,k}^{h}dndt = \widetilde{\beta}_{j,k}^{h} = \iint f_r\varphi_{j,k}^{h}dndt\} \tag{19}$$

$$C_{inter}^{v} = \{\iint \widetilde{f}'\varphi_{j,k}^{v}dndt = \widetilde{\beta}_{j,k}^{v} = \iint f_r\varphi_{j,k}^{v}dndt\} \tag{20}$$

$$C_{inter}^{d} = \{\iint \widetilde{f}'\varphi_{j,k}^{d}dndt = \widetilde{\beta}_{j,k}^{d} = \iint f_r\varphi_{j,k}^{d}dndt\} \tag{21}$$

$\widetilde{f}'(n,t)$ denotes the estimated value of the $f'(n,t)$. So the projection of $\widetilde{f}'(n,t)$ to C_{inter} (denoted by P_{inter}) is defined as

$$\begin{aligned} f'(n,t) &\equiv P_{inter}[\widetilde{f}'(n,t)] \\ &\equiv \sum_{j,k\in Z} f_c\psi_{j,k}(n,t) + \sum_{j,k\in Z} \widetilde{\beta}_{j,k}^{h}\varphi_{j,k}^{h}(n,t) \\ &+ \sum_{j,k\in Z} \widetilde{\beta}_{j,k}^{v}\varphi_{j,k}^{v}(n,t) + \sum_{j,k\in Z} \widetilde{\beta}_{j,k}^{d}\varphi_{j,k}^{d}(n,t) \end{aligned} \tag{22}$$

Next, two steps of wavelet transform operation are implemented on the image $f'(n,t)$ which is produced after the inter-frame projection. Firstly, we don't employ the down-sampling, and get the three high frequency bands such as LH_1, HL_1 and HH_1. Secondly, we employ the down-sampling, and get the low frequency band LL and other three high frequency bands LH_2, HL_2 and HH_2. In addition, wavelet transform operation is implemented on the LL band and non-sampling, so we can get three bands of LH_3, HL_3 and HH_3. At last, the LS (Least Square) estimation of prediction is utilized to get the three bands of $\widetilde{LH}$, $\widetilde{HL}$ and $\widetilde{HH}$. The convex can be constructed as follow:

$$C_{intra} = \{\widetilde{f''}(n,t)|C_{intra}^{LH}, C_{intra}^{HL}, C_{intra}^{HH}\} \tag{23}$$

where

$$C_{intra}^{LH} = \{[r_{LH} = \widetilde{LH} - \sum_{j,k\in Z} \widetilde{\beta}_{j,k}^{h}\varphi_{j,k}^{h}] \in [-\delta,\delta]\} \tag{24}$$

$$C_{intra}^{HL} = \{[r_{HL} = \widetilde{HL} - \sum_{j,k\in Z} \widetilde{\beta}_{j,k}^{v}\varphi_{j,k}^{v}] \in [-\delta,\delta]\} \tag{25}$$

$$C_{intra}^{HH} = \{[r_{HH} = \widetilde{HH} - \sum_{j,k\in Z} \widetilde{\beta}_{j,k}^{d}\varphi_{j,k}^{d}] \in [-\delta,\delta]\} \tag{26}$$

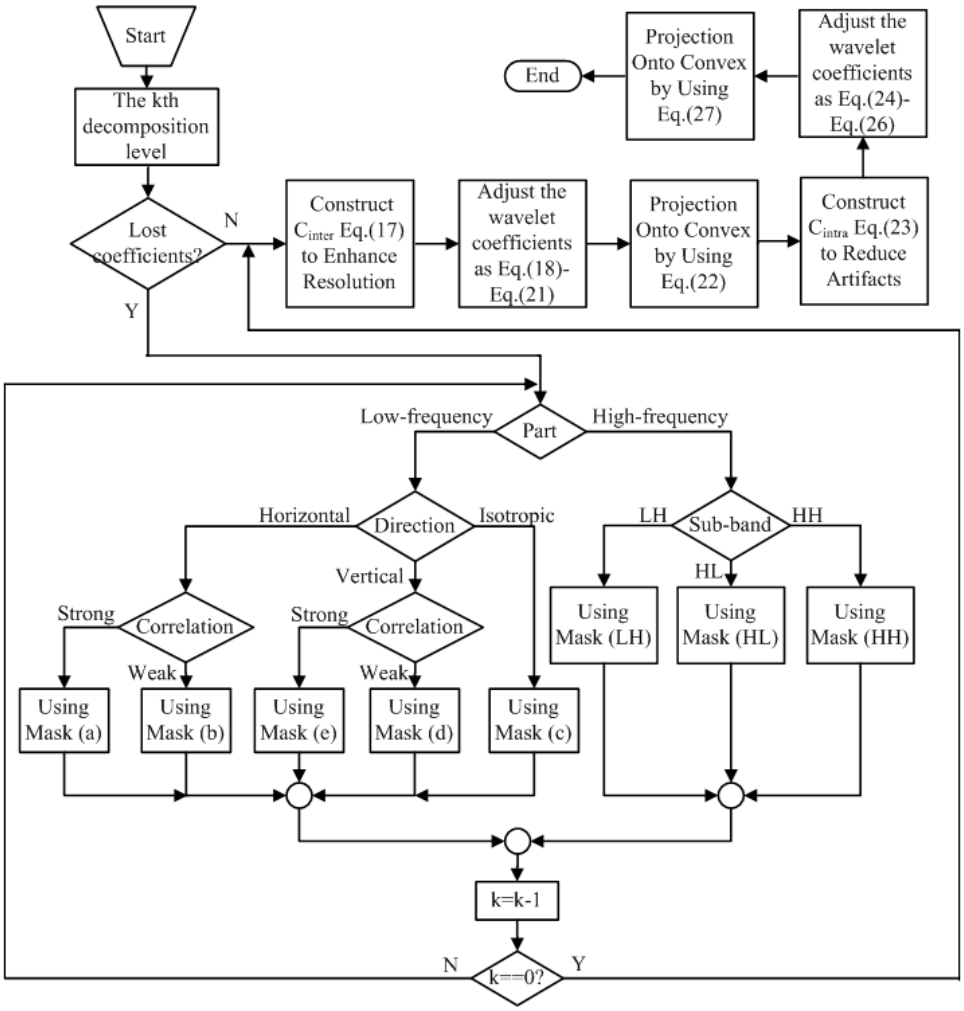


Fig. 4. Flow chart of the proposed robust SR method

where r_{LH}, r_{HL} and r_{HH} denote the residue between the estimated HR image through degradation and the original LR image in LH, HL and HH subband, respectively. $[-\delta, \delta]$ shows the boundary of the noise magnitude. After projecting by P_{intra}, we get (here we only take the LH part for

example, the other two bands are similar to this)

$$(LH)' = P_{intra}[\widetilde{LH}] = \widetilde{LH} + \begin{cases} \widetilde{LH} - \sum\limits_{j,k \in Z} \widetilde{\beta}^h_{j,k} \varphi^h_{j,k}, & r_{LH} > \delta \\ \widetilde{LH} + \sum\limits_{j,k \in Z} \widetilde{\beta}^h_{j,k} \varphi^h_{j,k}, & r_{LH} \leq -\delta \\ 0, & else \end{cases} \tag{27}$$

From the $(LH)'$, $(HL)'$, $(HH)'$ and the known LL, the HR image can be get easily. The flow chart of the whole SR algorithm is illustrated in Fig.4.

V. Simulation Results and Discussions

In this section, we conduct simulation experiments to study the performance of the proposed robust resolution-enhancement scheme in a distributed video streaming framework. First of all, we describe the simulation environment. Secondly, we present the main simulation results where we show the objective and subjective results of the performance of the proposed system under different scenarios. Finally, we conclude this section by summarizing the conclusions to be drawn based on the selected simulation results described.

In order to provide a meaningful comparison between our proposed approach and other alternative approaches, we consider use of a recent unbalanced MDC with UEP mentioned in [15] as a comparison system. This paper presents a distributed video streaming framework using unbalanced MDC and UEP under the video streaming framework that two senders simultaneously stream complementary descriptions to a single receiver over different paths. To minimize the overall distortion and exploit the benefits of multi-path transport when the characteristics of each path are different, this paper also propose an unbalanced MDC method for wavelet-based coders combined with a TCP-friendly rate allocation algorithm. In addition, three-level wavelet decomposition is applied to a group of 16 frames and the 3-D wavelet coefficients are divided into two unequal-sized groups according to the priority level. Each substream is independently protected using the Forward Error Correction (FEC) based UEP algorithm proposed in [20]. Because this method can not enhance the resolution of image, we utilize the proposed SR method to give a fair comparison. For simplicity, we note this method as unbalanced method. In addition, a fixed wavelet decomposition $K = 2$ (the remaining is the same as our proposed

approach) is compared with the flexible MDC; and traditional bilinear method (the remaining is the same as our proposed approach) is employed to compare with our SR algorithm. Similarly, the above two methods are noted as fixed method and bilinear method, respectively.

A. Simulation Environment

For these experiments, the two standard video sequences, Foreman and Weather forecast, are encoded with shifted 3-D SPIHT algorithm. These video sequences are 352×288 pixels per frame; frame rate $F = 30 frames/s$; down-sampling parameter $q = 2$; the blur is Gaussian blur where the support of the blurring kernel is 29; frame number of one GOF $N_{GOF} = 16$. At the receiver, the desired HR frame is reconstructed from several LR 176×144 degraded frames transmitted by MANETs. In order for objective comparison, PSNR at the receiver relative to the original HR video sequence is used and its definition is

$$PSNR(dB) = 10 \log_{10}(\frac{255^2}{MSE}) \tag{28}$$

where MSE is the mean-square error between the original the reconstructed luminance frame. To the network part, 30 nodes that move randomly at maximum speed 2m/s are dispersed in the area 200×200 square meters, and the routing algorithm is based on MRDSR (Multiple Route Dynamic Source Routing) [21]. In addition, the packet size $S = 512$bytes and initial energy of nodes is 1 Joule with uniform distribution. It should be noted that all the simulation results in this section have been obtained using 30 runs in order to obtain statistically meaningful average values.

B. Selected Simulation Results and Discussions

In Fig.5, we illustrate a plot of PSNR versus the packet loss rate P_L with burst length $L_B = 4$ for the Foreman sequence at $r_b = 96$Kbps, which is the range of low-bit services. The proposed method can be seen to achieve a much higher performance in terms of end-to-end PSNR compared to the representative fixed method which apply fixed wavelet decomposition level regardless of the condition of the network and unbalanced method which compulsively divided stream into two unbalanced substreams. For the above two methods, there is a considerable performance disadvantage due to the improper use of the decomposition level and data distribution. Obviously, for little loss transmission ($P_L \leq 5\%$) the fixed method achieves

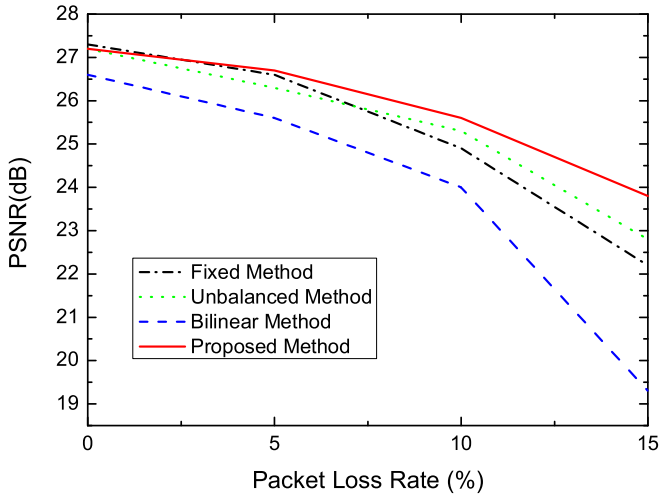


Fig. 5. Performance achieved by proposed method for the Foreman sequence, at the $r_b = 96$Kbps and $L_B = 4$.

almost the same performance compared to the proposed scheme, and proposed method only gets marginal performance gain compared to unbalanced method when ($8\% \leq P_L \leq 10\%$). However, as P_L increases, the performance gap is increased dramatically since when the packet loss rate is high, more packets are unsuccessfully delivered obviously, non-adaptive scheme, including fixed and unbalanced method, can not adapt transmission strategy to this case while proposed method employ more decomposition to guarantee the most important data successfully transmitted. For example, when $P_L = 5\%$, the 2-level decomposition is adopted by the proposed method, so the gap between the proposed method with the fixed method is only about 0.1 dB, and when $P_L = 10\%$, the decomposition is adaptive to 3-level, the gap between the proposed method with the unbalanced method is about 0.3 dB, while the gap become 1.0dB when the $P_L = 15\%$ and 4-level decomposition is employed.

It should also be noted from Fig.5 that the performance achieved by the proposed method is also super to the traditional bilinear method to get higher resolution. Though proposed SR method can only achieve little gain compared to the traditional bilinear method in the case of lossless transmission ($P_L = 0$), the performance gap also increases sharply as the P_L increases. In particular, as can be seen from Fig.5, the performance gap between proposed SR methods

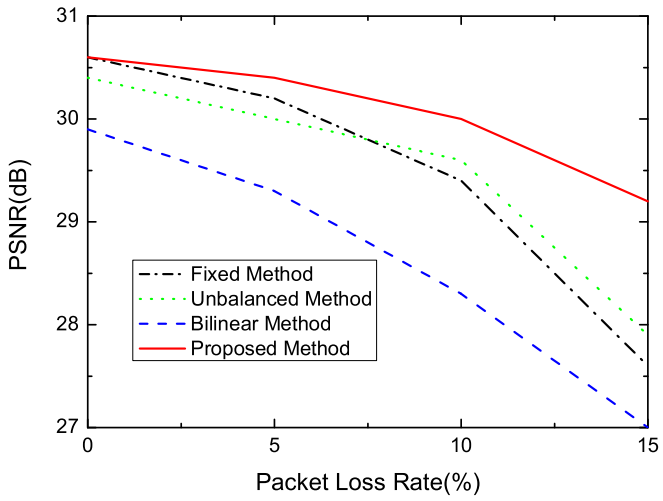


Fig. 6. Performance achieved by proposed method for the Foreman sequence, at the $r_b = 256$Kbps and $L_B = 4$.

with bilinear comparison method is more than 1.5dB when $P_L \geq 10\%$. The reason is that, when $P_L \leq 5\%$, few packets are lost and correlation of the adjacent packets is strong, bilinear as a simply interpolation method can also provide a relatively accurate estimation, while when $P_L > 5\%$ (especially $P_L \geq 10\%$), the packet correlation is so weak that bilinear can't performs well in this case. To our proposed SR method, adjacent and current frames information is utilized to enhance the resolution of the received image, therefore, it can achieve a more satisfying performance whether packet loss rate high or not.

In Fig.6, we repeat the results for Foreman sequence with $r_b = 256$Kbps and $L_B = 4$. It can be seen that the performance gap between the fixed or unbalanced method with the proposed method is much larger than Fig.5, which is the case for low-bit rate services, while Fig.6 is for high-rate services. This follows the fact that, when the available bit rate is high, the reconstructed video quality benefits considerably from the use of flexible MDC substreams and SR reconstruction technique which uses adaptive strategy to adapt current network condition and correlation information hidden in adjacent frames to enhance the received image resolution. Still, it can be seen that when $P_L = 10\%$, the proposed method can achieve a performance gain of approximately 0.4dB compared to the unbalanced method, and 1.7dB compared to the bilinear

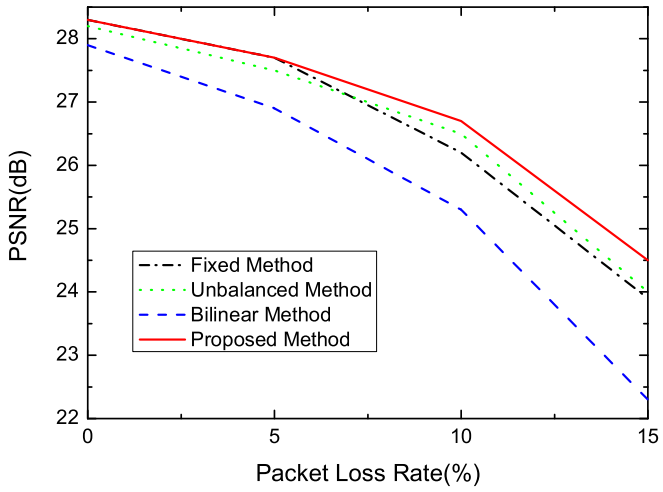


Fig. 7. Performance achieved by proposed method for the Weather forecast sequence, at the $r_b = 96$Kbps and $L_B = 4$.

interpolation method, and when $P_L = 15\%$, the gap increases to 1.3dB and 2.2dB, respectively.

As discussed previously, the Foreman sequence is a high-motion sequence, while the Weather forecast sequence is a low-motion sequence. In order to provide a more comprehensive evaluation of the proposed adaptive scheme, we repeat the results for Weather forecast sequence in Fig.7 and Fig.8, for $r_b = 96$Kbps and $r_b = 256$Kbps, respectively, and similar behaviors as in Fig.5 and Fig.6 are observed. It should be note that the performance gap between the proposed adaptive scheme and the compared methods are less obvious than that in Fig.5 and Fig.6 for the high-motion Foreman sequence, because the correlation between the adjacent frames increases as the low-motion sequence implies. However, the proposed adaptive scheme also has the advantage of other alternative methods no matter what the P_L is.

Some subjective results to demonstrate the relative performance achieved by the proposed method are illustrated in Fig.9 corresponding to the objective results in Fig.8 for $P_L = 15\%$. Clearly, in Fig.9, the subjective results are consistent with the objective results in Fig.8. In particular, the proposed method can maintain more visual information comparing to the other alternative methods.

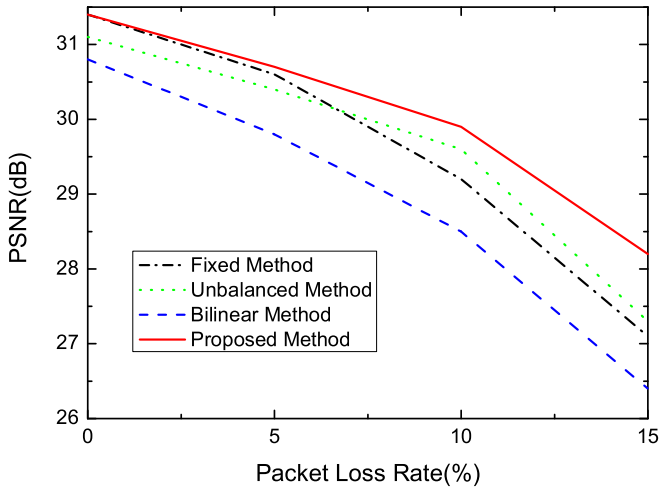


Fig. 8. Performance achieved by proposed method for the Weather forecast sequence, at the $r_b = 256$Kbps and $L_B = 4$.

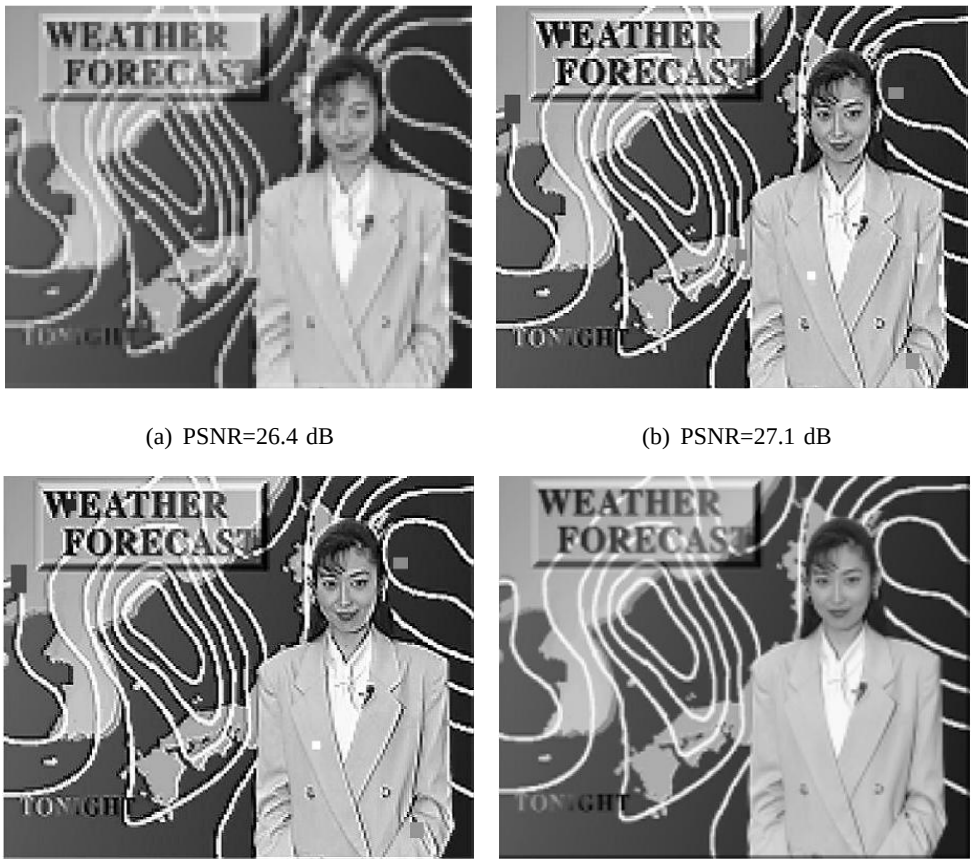
C. Observations

Based on the selected objective and subjective simulation results described above, there are several main observations:

• The adaptive error-resilient strategy has played an important role in the whole video transmission system. Comparing to the fixed method or the unbalanced method, the performance of the adaptive strategy is greater than those of them, especially when the packet loss rate is high.

• The proposed SR algorithm actually can enhance the resolution of the received image, and much detail information is maintained by taking advantage of the projection onto convex sets. Comparing to the traditional bilinear method, the proposed SR algorithm at least has the 0.5dB PSNR gain, furthermore, the advantage gets more evident as the packet loss rate increases.

• The proposed method has great robust. No matter the video sequence is high-motion or low-motion, the packet loss rate is high or low, the proposed method can perform well all the time.



(a) PSNR=26.4 dB (b) PSNR=27.1 dB

(c) PSNR=27.3 dB (d) PSNR=28.2 dB

Fig. 9. Subjective results achieved by proposed method and other comparison schemes, for the Weather forecast at $r_b = 256$Kbps, $L_B = 4$ and $P_L = 15\%$. (a) bilinear method; (b) fixed method; (c) unbalanced method; (d) proposed method.

VI. Conclusions and Future Work

In this paper, we propose a robust resolution-enhancement scheme for video stream transmission over mobile ad-hoc networks. As detailed in the paper, the scheme can adaptively respond to the dynamic network condition by adjusting the coding fashion at the encoder and error protection strategy during the transmission process. Furthermore, the SR algorithm used in the proposed scheme is so robust that it performs well in presence of different kinds of packet loss rates. Experiment results demonstrate that the proposed scheme outperforms the competing methods under the basis of the same simulation condition. We should note that our proposed method is hard to provide a real-time processing in wireless video system, therefore, our future work is to reduce its complexity to adapt to the real-time wireless video transmission.

ACKNOWLEDGMENTS

This work is supported by the International Project PRA-SI (financed by France and China government) under Grant No. SI04-03, the Key Project of Nature Science Foundation of Jiangsu (China) under Grant BK2007729 and the Climbing Plan of Nanjing University of Posts and Telecommunications under Grant NY207061. Moreover, we also thank the anonymous reviewers for insightful comments and suggestions.

REFERENCES

[1] W.Y. Kung, C.S.Kim, and C.C.Jay Kuo, "Spatial and temporal error concealment techniques for video transmission over noisy channels", IEEE Trans. on Circuits and Systems for Video Technology, vol.16, no.7, pp. 789-802, July 2006.

[2] W.J. Hwang, J.F. Chen, Y.C. Huang and T.Y. Tsai, "Layered video coding based on displaced frame difference prediction and multi-resolution block matching", IEEE Trans. on Communications, vol.52, no.9, pp.1504-1513, Sept. 2004.

[3] R. Mathew and J. F. Arnold, "Efficient layered video coding using data partitioning," Signal Process: Image Communication, vol. 14, pp. 761-782, 1999.

[4] C.S. Kim and S.U. Lee, "Multiple description coding of motion fields for robust video transmission," IEEE Trans. on Circuits System Video Technology, vol. 11, no. 9, pp. 999-1010, Sep. 2001.

[5] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," IEEE Trans. on Circuit System Video Technology, vol. 12, no. 2, pp. 438-452, Jun. 2002.

[6] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," IEEE Journal of Selected Areas Communication, vol. 18, no. 6, pp. 966-976, Jun. 2000.

[7] E. Steinbach, N. Farber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environment," IEEE Trans. on Circuit System Video Technology., vol. 7, no. 6, pp. 872-881, Dec.1997.

[8] S.C. Park, M.K. P, and M.G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," IEEE Signal Processing Magazine, vol.20, no.3, pp.21-36, May 2003.

[9] W.N. Lie, and Z.W. Gao, "Video Error Concealment by Integrating Greedy suboptimization and Kalman Filtering Techniques," IEEE Trans. on Circuits and Systems Video Technology, vol.16, no.8, pp.982-992, Aug. 2006.

[10] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficient," IEEE Trans. Signal Processing, vol. 41, pp. 3445-3462, Dec.1993.

[11] A. Said and W. A. Pearlman, "A New, fast and efficient image codec based on set partitioning in hierarchical trees," IEEE Trans. on Circuits System. Video Technology, vol. 6, pp. 243-250, June 1996.

[12] Y. Chen and W. A. Pearlman, "Three-dimensional subband coding of video using the zero-tree method," SPIE Visual Communication Image Processing, pp. 1302-1309, Mar. 1996.

[13] B.J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees," in Proc. Data Compression Conf., pp. 251-260, Mar. 1997.

[14] Y.Sriraja, T.Karp, "A packetized SPIHT algorithm with over complete wavelet coefficients for increased robustness", Eurasip journal on applied signal processing, special issue on frames and over complete representations in signal processing, communications, and information theory, vol. 13, pp. 112-120, 2005.

[15] J. Kim, R. M. Mersereau, and Y.Altunbasak. “Distributed Video Streaming Using Multiple Description Coding and Unequal Error Protection”, IEEE Trans. Image Processing, vol.14, no.7, pp.849-861, July 2005.

[16] N. Gogate, D. M. Chung, S. S. Panwar, and Y.Wang, “Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding,” IEEE Trans. on Circuits System Video Technology, vol. 12, no. 6, pp. 777-792, Sep. 2002.

[17] A. Mohr, E. Riskin, and R. Ladner, “Unequal loss protection: Graceful degradation of image quality over packet erasure channel through forward error correction,” IEEE Journal Selected Areas Communication, vol. 18, no. 6, pp. 818-828, Jun. 2000.

[18] U. Horn, K. Stuhlmuller, M. Link, and B. Girod, “Robust internet video transmission based on scalable coding and unequal error protection,” Signal Processing: Image Communication, vol. 15, pp. 77-94, Sept. 1999.

[19] I.V. Bajic, “Adaptive MAP Error Concealment for Dispersively Packetized Wavelet-Coded Images,” IEEE Trans. on Image Processing, vol.15, no.5, pp.1226-1235, May 2006.

[20] J. Kim, R. M. Mersereau, and Y. Altunbasak, “Error-resilient image and video transmission over the internet using unequal error protection,” IEEE Trans. on Image Process., vol. 12, no. 2, pp. 121-131, Feb. 2003.

[21] A. Nasipuri, R. Castaneda, and S. Das, “Performance of multipath routing for on-demand protocols in mobile ad-hoc networks”, ACM/Kluwer mobile networks and applications, vol.6 pp.339-349, 2001.

[22] M.Kang and M.Alouini, “Transmission of Multiple Description Codes Over Wireless Channels Using Channel Balancing”, IEEE Trans. Wireless Communications, vol. 4, no. 5, pp.2070-2075, Sep. 2005.

[23] I. V. Bajic and J. W. Woods, “Domain-based multiple description coding of images and video”, IEEE Trans. Image Processing, vol. 12, no. 10, pp. 1211-1225, Oct. 2003.