

RESEARCH

Open Access

# Joint redundant motion vector and intra macroblock refreshment for video transmission

Jimin Xiao<sup>1,2</sup>, Tammam Tillo<sup>2\*</sup>, Chunyu Lin<sup>3</sup> and Yao Zhao<sup>3</sup>

## Abstract

This paper proposes a scheme for error-resilient transmission of videos which jointly uses intra macroblock refreshment and redundant motion vector. The selection of using intra refreshment or redundant motion vector is determined by the rate-distortion optimization procedure. The end-to-end distortion is used for the rate-distortion optimization, which can be easily calculated with the recursive optimal per-pixel estimate (ROPE) method. Simulation results show that the proposed method outperforms both the intra refreshment approach and redundant motion vector approach significantly, when the two approaches are deployed separately. Specifically, for the Foreman sequence, the average PSNR of the proposed approach can be 1.12 dB higher than that of the intra refreshment approach and 5 dB higher than that of the redundant motion vector approach.

**Keywords:** H.264/AVC, error resilience, end-to-end distortion, intra refreshment, redundant motion vector

## 1. Introduction

The H.264/AVC [1] video coding standard provides higher coding efficiency and stronger network adaptation capability in comparison to all the previously developed video coding standards. However, as previous video compression standards, it is still based on a hybrid coding method, which use transform coding with motion-compensated prediction (MCP). As a result, when transmitting the hybrid-coded video in packet loss environments, it suffers from error propagations and this leads to the well-known drifting phenomenon [2,3].

Due to the unreliable underlying networks, the development of error-resilient video coding techniques are a crucial requirement for video communications over lossy networks. Among all the error-resilient video coding techniques, two categories of robust coding approaches are promising. One category is based on intra macroblock refreshment, and another one is redundant coding.

The intra macroblock refreshment approach is standard compatible, and it is an useful tool to combat network packet losses. It can be employed to weaken the inter picture dependency due to inter prediction, and

eventually, cut-off the error propagations. The early intra macroblock refreshment algorithms are based on randomly inserting intra macroblocks [4] or periodically inserting intra contiguous macroblocks [5]. However, in both [4] and [5] the intra refresh frequency is determined in a heuristic way, and it is costly to code an entire picture by intra coding. So the trade-off between code efficiency and error resiliency need to be balanced. Zhang et al. first treated this problem as optimization of coding mode selection for each macroblock in [6], and proposed the well-known recursive optimal per-pixel estimate (ROPE) approach to determine intra macroblock. In [6] the expected end-to-end distortion for each pixel is calculated in a recursive way, then in the mode selection step, the expected end-to-end distortion is used in the rate-distortion optimization process. In [7], another flexible intra macroblock update algorithm was investigated to optimize the expected rate-distortion performance. In this approach, the end-to-end distortion is calculated by emulating the real channel behaviors, therefore, the computation complexity is tremendous. Among the methods to get the expected end-to-end distortion, [6] is pixel-based, another block-based approach [8] generates and recursively updates a block-level distortion map for each frame. The work in [6-8] are loss-aware end-to-end rate-distortion optimized intra macroblock refreshment algorithm, which are currently the

\* Correspondence: tammam.tillo@xjtlu.edu.cn

<sup>2</sup>Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, 111 Ren Ai Road, Suzhou, P.R. China  
Full list of author information is available at the end of the article

best known way for determining both the correct number and placement of intra macroblocks for error resilience.

Redundant coding is another effective tool for robust video communications over lossy networks. In [9], an optimal algorithm is presented to determine whether one picture needs redundant version. In [10], redundant slices are optimally allocated based on the slice position in the group of pictures (GOP), and the primary and redundant slices are then interleaved to generate two equal importance descriptions of the same data using the multiple description coding (MDC) paradigm. Whereas in [11], the two descriptions are generated by splitting the video pictures into two threads, and then redundant pictures are periodically inserted into the two threads. In both [9] and [11] redundant coding are optimized in frame level, namely all the macroblocks in one frame is encoded with the same redundant coding parameters. Whereas for [10], redundant information is allocated in slice level. In all the three approaches, redundant bitrate is allocated to both motion vectors and residual information. In [12] a new approach with only redundant motion vectors is proposed, as the redundant bitrate for motion vector is low, this approach improves the bandwidth utilization with limited primary picture quality degradation. In [13] a significant motion vector protection (SMVP) scheme for error-resilient transmission of videos is proposed. This scheme shows how to determine the significant motion vectors (SMVs) and how much rate should be dedicated to SMVs. The idea behind this scheme is to give more protection to SMVs.

Intra macroblock refreshment can stop errors in the previous frames, while redundant coding is a way of preventing and minimizing propagated errors in the future frames. Motivated by the two approaches, in this paper, we propose an innovative approach that jointly uses intra macroblock refreshment and redundant motion vector. For each macroblock, intra coding or redundant motion vector is chosen based on the rate-distortion optimization procedure. The loss-aware end-to-end expected distortion is used for this RD optimization, and the end-to-end distortion is calculated with the ROPE [6] method.

The rest of the paper is organized as follows. In Section 2, the ROPE method is presented as preliminary, as it is the method we adopt to calculate the end-to-end distortion, In Section 3, the proposed joint redundant motion vector and intra macroblock refreshment (JRVIR) approach is introduced. In Section 4, extensive simulation results are given, which validate our approach. Finally, some conclusions are drawn in Section 5.

## 2. Preliminary and the rope approach

In an ideal error-free environment, the rate-distortion optimized intra/inter mode decision is an efficient tool to determine the macroblock mode based on the cost function defined in [14], the cost function of all the macroblocks is defined as

$$J_{MB} = D_{MB} + \lambda_{mode} R_{MB} \quad (1)$$

where  $\lambda_{mode}$  is the Lagrange multiplier,  $D_{MB}$  and  $R_{MB}$  are the encoding distortion and the bitrate in different encoding modes, respectively. This optimization mode is tailored for error-free environments, and no packet loss is considered here.

However, when the compressed videos are transmitted over error-prone networks, traditional schemes cannot adaptively insert intra refresh macroblocks to efficiently stop the channel error propagations. The ROPE approach uses the end-to-end distortion in the RD optimization, which takes into account the channel packet losses. With the ROPE approach, intra macroblocks are optimally used to stop error propagations, and it is defined as follows.

Let  $f_n^i$  denote the original value of pixel  $i$  in frame  $n$ , and let  $\hat{f}_n^i$  and  $\tilde{f}_n^i$  denote its encoder and decoder reconstruction, respectively. Because of possible packet loss in the channel,  $\tilde{f}_n^i$  can be modeled at the encoder side as a random variable. In the ROPE approach, the  $D_{MB}$  is redefined as the overall expected decoder distortion in one macroblock.

$$D_{MB} = \sum_{i \in MB} d_n^i \quad (2)$$

$$\begin{aligned} d_n^i &= E\{(f_n^i - \tilde{f}_n^i)^2\} \\ &= (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\} \end{aligned} \quad (3)$$

The overall expected mean-squared-error (MSE) distortion of a pixel is  $d_n^i$ , obviously, it is determined by the first and second moments of the decoder reconstruction. ROPE provides an optimal recursive algorithm to accurately calculate the two moments for each pixel in a frame.

Let us assume that packet loss events are independent for simplicity, and the packet loss rate (PLR)  $p$  is available at the encoder side. To make it more general, there is no limitation on the slice shape and size, so the motion vectors from neighboring macroblocks are not always available in the error concealment stage. Therefore, the decoder may not be able to use motion vector from neighboring macroblocks for concealment. Accordingly, we assume the decoder copies

reconstructed pixels from the previous frame for concealment. The prediction at the encoder only employs the previous reconstructed frame. The recursive formulas of ROPE are as follows.

- Pixel in the intra macroblock

$$E\{\tilde{f}_n^i\} = (1 - p)\hat{f}_n^i + pE\{\tilde{f}_{n-1}^i\} \quad (4)$$

$$E\{(\tilde{f}_n^i)^2\} = (1 - p)(\hat{e}_n^i)^2 + pE\{(\tilde{f}_{n-1}^i)^2\} \quad (5)$$

- Pixel in the inter macroblock

$$E\{\tilde{f}_n^i\} = (1 - p)(\hat{e}_n^i + E\{\tilde{f}_{n-1}^{i+mv}\}) + pE\{\tilde{f}_{n-1}^i\} \quad (6)$$

$$E\{(\tilde{f}_n^i)^2\} = (1 - p)((\hat{e}_n^i)^2 + 2\hat{e}_n^i E\{\tilde{f}_{n-1}^{i+mv}\} + E\{(\tilde{f}_{n-1}^{i+mv})^2\}) + pE\{(\tilde{f}_{n-1}^i)^2\} \quad (7)$$

where inter coded pixel  $i$  is predicted from pixel  $i+mv$  in the previous frame. The prediction residual  $e_n^i$  is quantized to  $\hat{e}_n^i$ .

It is important to notice that in order to make it simple, we apply ROPE in its simple setting, where the motion estimation is evaluated at pixel level accuracy, and we use constrained intra prediction, so there are no error propagations in the intra prediction. Recent advances in ROPE further expand its capability to accommodate sub-pixel prediction [15], bursty packet loss [16]. But they are not incorporated here so as to avoid diluting the focus. In the ROPE approach, the end-to-end distortion is only used in the mode selection stage. However, recently, in [17,18] end-to-end distortion is applied in the motion estimation and motion prediction stage, which is so-called loss-aware motion estimation and loss-aware motion prediction. With this extension the error-resilience capability of ROPE is improved further. The loss-aware motion estimation and loss-aware motion prediction is not used in our approach, because we extend ROPE in different direction, in fact, the gain can be accumulated if both the loss-aware motion estimation and loss-aware motion prediction are applied.

### 3. The proposed JRVIR approach

As both redundant motion vector and intra macroblock refreshment are powerful tools for error resilient video communications. In the proposed JRVIR approach, they are jointly applied to further protect the video stream. With the JRVIR approach, all the macroblocks of one frame are divided into three types, namely intra

macroblock, inter macroblock (including skip) without redundant motion vector and inter macroblock (including skip) with redundant motion vector. The redundant motion vectors are encapsulated in the redundant picture. Let us take macroblocks in Figure 1 as an example, let us suppose the last macroblock in the first row is a macroblock with redundant motion vector, accordingly, it is stored in the redundant picture. On the contrary, for intra refresh macroblock and inter macroblock without redundant motion vector, there will be no redundant information to be sent in the redundant picture. Therefore, for inter macroblock with redundant motion vector, if the macroblock in the primary picture is lost due to packet losses, the redundant extra motion vector can be used to recover the macroblock. It is important to note that, in the proposed JRVIR approach, as not all the macroblocks need to have redundant motion vector, a new flag is applied in each macroblock to indicate whether there is redundant motion vector. For these macroblocks with redundant motion vector, there will be no transformed coefficients to be encapsulated in the redundant macroblocks. Therefore, the proposed JRVIR would not be standard compatible, and some small modifications are required for both the encoder and decoder.

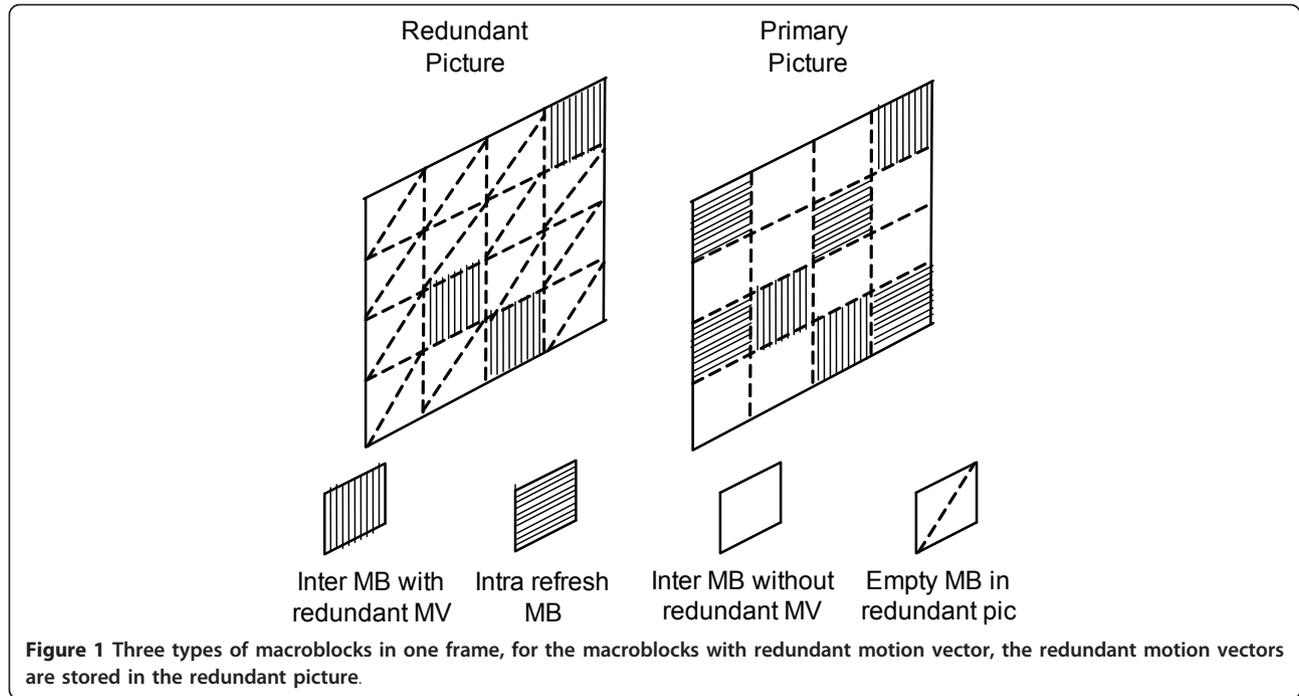
In general, intra coding is more expensive in terms of rate requirement with respect to redundant motion vector, therefore for the macroblocks with smooth texture and/or macroblocks with slow and translational movements, providing redundant motion vector would lead to better resource utilization, i.e., bitrate, with respect to the intra coding. Whether to encode one macroblock with intra mode, inter mode with redundant motion vector or without motion vector is determined by our JRVIR rate-distortion optimization process.

#### 3.1 The JRVIR rate-distortion optimization

As in other encoding approaches, in the JRVIR rate-distortion optimization process, the encoder selects the coding option  $O^*$  for the current encoding macroblock, so that the Lagrangian cost function is minimized.

$$O^* = \arg \min_{o \in \Gamma_{\text{JRVIR}}} (D_{\text{MB}}(o) + \lambda_{\text{mode}} R_{\text{MB}}(o)) \quad (8)$$

where  $D_{\text{MB}}(o)$  is the expected end-to-end distortion for mode  $o$ ,  $R_{\text{MB}}(o)$  is the rate for this mode and  $\lambda_{\text{mode}}$  is the Lagrangian multiplier.  $\Gamma_{\text{JRVIR}}$  is a set of encoding options which includes all encoding modes. For the original ROPE approach, the available encoding modes includes Intra mode, SKIP mode and Inter mode, so  $\Gamma_{\text{ROPE}} = \{\text{Intra, SKIP, Inter}16 \times 16, \text{Inter}16 \times 8, \text{Inter}8 \times 16, \text{Inter}8 \times 8\}$ . However, in our JRVIR approach, there are five new modes, they are SKIP, Inter $16 \times 16$ , Inter $16 \times 8$ , Inter $8 \times 16$  and Inter $8 \times 8$ , all with redundant



motion vector. For simplicity, let us use Skip\_dup, Inter\_dup16 × 16, Inter\_dup16 × 8, Inter\_dup8 × 16, Inter\_dup8 × 8 to denote the five new modes, with *dup* standing for duplicating motion vector. Therefore, for the JR VIR approach, the set of encoding options becomes  $\Gamma_{JR VIR} = \{\text{Intra, SKIP, Inter}_{16 \times 16}, \text{Inter}_{16 \times 8}, \text{Inter}_{8 \times 16}, \text{Inter}_{8 \times 8}, \text{Skip\_dup, Inter\_dup}_{16 \times 16}, \text{Inter\_dup}_{16 \times 8}, \text{Inter\_dup}_{8 \times 16}, \text{Inter\_dup}_{8 \times 8}\}$ .

### 3.2 The JR VIR end-to-end distortion

When calculating the expected end-to-end distortion, we can still use Equations 4 and 5 for intra macroblock, and Equations 6 and 7 for inter macroblock without redundant motion vector. For inter macroblock with redundant motion vector, first and second moments of the decoder reconstruction are as follows.

$$E\{\tilde{f}_n^i\} = (1 - p)(\hat{e}_n^i + E\{\tilde{f}_{n-1}^{i+mv}\}) + p(1 - p)\{\tilde{f}_{n-1}^{i+mv}\} + p^2 E\{\tilde{f}_{n-1}^i\} \quad (9)$$

$$E\{(\tilde{f}_n^i)^2\} = (1 - p)((\hat{e}_n^i)^2 + 2\hat{e}_n^i E\{\tilde{f}_{n-1}^{i+mv}\} + E\{(\tilde{f}_{n-1}^{i+mv})^2\}) + p(1 - p)E\{(\tilde{f}_{n-1}^{i+mv})^2\} + p^2 E\{(\tilde{f}_{n-1}^i)^2\} \quad (10)$$

For those inter macroblocks with redundant motion vector, the probability of receiving the primary information is  $1 - p$ . The probability of receiving the redundant motion vector while losing the primary information is  $p(1 - p)$ , and the probability of both the primary information and the redundant motion vector get lost is  $p^2$ . With all those probabilities, we can easily get Equations

9 and 10 for inter macroblock with redundant motion vector.

### 3.3 The JR VIR rate

In the RD optimization procedure, the rate of the redundant motion vector should be taken into account. For those redundant motion vectors, encoding them without exploiting the correlation among them can cost a significant number of bits. Motion vectors for neighboring macroblocks are often highly correlated and so each motion vector is predicted from vectors of nearby, and previously coded macroblocks. Therefore, the motion vector encoding procedure in H.264 standard [1], which includes motion vector prediction, is adopted to encode the redundant motion vector to reduce bits. However, it is worth noticing that, in our JR VIR approach we do not provide redundant motion vector for all the inter macroblocks. For example, in Figure 1 only three macroblocks have redundant motion vector, when encoding the redundant motion vector for the macroblock in row 3, there are no motion vectors to predict from, because its up and left macroblocks do not have redundant motion vectors. As a result, the performance is compromised. In future, more sophisticated prediction method will be investigated, for example, the redundant motion vector would be predicted from the concealment emulated motion vector.

To determine the required rate to encode a macroblock using the JR VIR algorithm, let us assume that encoding the macroblock itself and its redundant

motion vector would use  $R_{mb}$  and  $R_{mv}$  bits, respectively. For encoding mode  $o \in \{\text{Intra, Skip, Inter}16 \times 16, \text{Inter}16 \times 8, \text{Inter}8 \times 16, \text{Inter}8 \times 8\}$ ,  $R_{MB}(o)$  in (8) equals to  $R_{mb}$

$$R_{MB}(o) = R_{mb} \quad (11)$$

Whereas, for encoding mode  $o \in \{\text{Skip\_dup, Inter\_dup}16 \times 16, \text{Inter\_dup}16 \times 8, \text{Inter\_dup}8 \times 16, \text{Inter\_dup}8 \times 8\}$ , the value of  $R_{MB}(o)$  is

$$R_{MB}(o) = R_{mb} + R_{mv} \quad (12)$$

### 3.4 Lagrange multiplier selection

The Lagrange multiplier  $\lambda_{mode}$  in (8) controls the rate-distortion trade-off. For the error-prone environment, extensive experimental evidence suggests that there is no significant performance difference between using the Lagrange multiplier tailored to the error-free or the error-prone environment. This argument has also been confirmed in [7]. So  $\lambda_{mode}$  is set as the one tailored to error-free environment.

$$\lambda_{mode} = 0.85 \times 2^{(QP-12)/3} \quad (13)$$

where QP is the quantization parameter.

### 3.5 The pseudo code of JRVIR algorithm

The whole mode selection process of the proposed JRVIR approach is described in Algorithm 1. It is important to note that, in the proposed JRVIR approach, the end-to-end distortion is used in the rate-distortion optimization process, and five new encoding modes are adopted. Upon the optimal encoding mode is selected, the first and second moments for all the pixels in current macroblock are recorded based on the selected encoding mode, and those values will be recursively used in the rate-distortion optimization process of next frame. At the decoder side, if the primary slice is available, the redundant motion vector will be discarded, whereas when the primary slice is lost while the redundant motion vector is available, the motion vector will be used to conceal the lost region by copying into the lost macroblock the region indicated by the redundant motion vector. In general, with the correct motion vector, the concealed pixels will be much more accurate than those generated by Temporal Replacement (TR), which copies the pixels from the same positions in the previous frame.

## 4. Simulation result

Our simulation setting builds on the JM14.0 H.264 codec [19], with constrained intra prediction and CABAC entropy coding used. Rate control mechanism in the JM codec is used with one common quantization

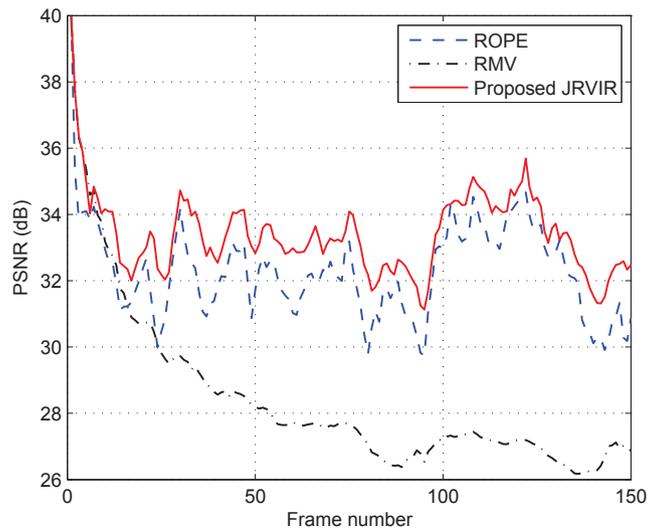
scale to all the macroblocks of one row. Pixel level accuracy motion estimation and prediction is used. Each slice contains one row of macroblocks (22 macroblocks for the CIF video sequences) for both primary and redundant frames, and one slice per network packet is adopted, therefore the term packet and slice are used interchangeably. The IPPP...GOP structure is used, and it is assumed that the I-frame is transmitted over secure channel. A random packet-loss generator is used to drop the packets according to the required packet loss rate, except the burst packet-loss is specified explicitly. The luminance PSNR (Y-PSNR) is averaged over 200 trials to get statistical meaningful results. To evaluate the proposed JRVIR approach, we use conventional ROPE [6] and redundant motion vector (RMV) [12] as benchmark.

First, the frame by frame average PSNR are presented in Figure 2, for the three approaches, namely the JRVIR approach, ROPE and RMV. Both CIF sequence Foreman and Silent are encoded at 1 Mbps bitrate. The packet loss rate is 10%. From the figures, it is observed that for all the frames, the JRVIR frame quality is always better than that of RMV and ROPE. For the Foreman sequence, for some frames the PSNR of JRVIR can be up to 2.5 dB higher than that of ROPE, and up to 8.5 dB higher than that of RMV. At the beginning of the sequence, the PSNR of JRVIR and RMV are quite similar, but with the increase of frame number the quality gap between the two approaches increases dramatically. This phenomena indicates that when the GOP length is small, the RMV approach can protect the video stream effectively, but when the GOP length is relatively large, the RMV approach can not work properly. For Foreman, the average PSNR for JRVIR approach is 33.39 dB, it is higher than that of ROPE and RMV, which are 32.27 dB and 28.39 dB, respectively. For Silent, the average PSNR for JRVIR approach is 37.56 dB, while for ROPE and RMV it is 36.83 dB and 31.27 dB. For Foreman, the gap between JRVIR and ROPE is larger than that of Silent, this is because the movement in Foreman is more translational than that in Silent, and this leads to more inter macroblocks with redundant motion vectors being used in the Foreman case. Interestingly, with the JRVIR approach 8.18% of macroblocks in P-frame are intra coded macroblocks, while 35.02% are inter macroblocks with redundant motion vector. In the ROPE approach 18.40% macroblocks are intra macroblocks, which is more than that of JRVIR by nearly 10%.

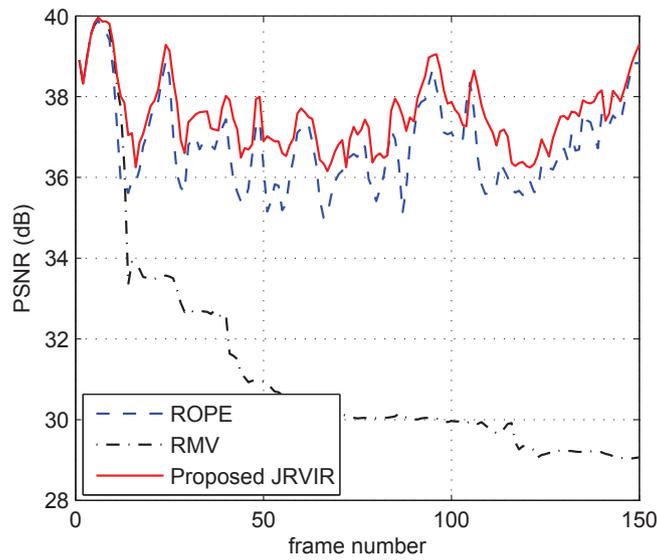
**Algorithm 1** The whole algorithm of mode selection in JRVIR

```

RD_cost ← ∞
best_mode ← ∞
for for each mode  $o \in \Gamma_{JRVIR}$  do
    if  $o \in \{\text{INTRA}\}$  then
    
```



(a) Foreman



(b) Silent

**Figure 2** Frame by frame comparison with bitrate 1 Mbps, packet loss rate 10%. a) CIF Forman b) CIF Silent.

calculate  $D_{MB}$  using Equations 2-5

calculate  $R_{MB}$  using Equation 11

**else**

**if**  $o \in \{\text{Skip}, \text{Inter}16 \times 16, \text{Inter}16 \times 8, \text{Inter}8 \times 16, \text{Inter}8 \times 8\}$  **then**

calculate  $D_{MB}$  using Equations 2, 3, 6 and 7

calculate  $R_{MB}$  using Equation 11

**else**

**if**  $o \in \{\text{Skip\_dup}, \text{Inter\_dup}16 \times 16, \text{Inter\_dup}16 \times 8, \text{Inter\_dup}8 \times 16, \text{Inter\_dup}8 \times 8\}$

**then**

calculate  $D_{MB}$  using Equations 2, 3, 9 and 10

calculate  $R_{MB}$  using Equation 12

**end if**

**end if**

calculate  $J_{MB}$  using Equations 1 and 13

**if**  $J_{MB} < \text{RD\_cost}$  **then**

$\text{RD\_cost} = J_{MB}$

$\text{best\_mode} = o$

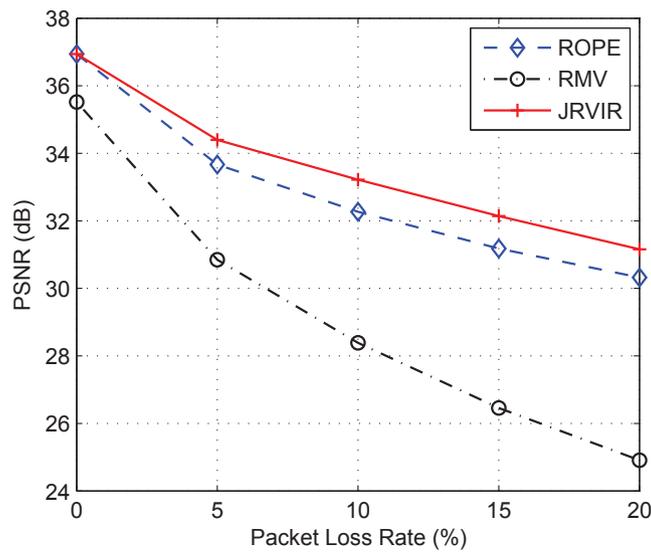
record the value of  $E\{\tilde{f}_n^i\}$  and  $E\{\tilde{f}_n^i\}^2$

**end if**

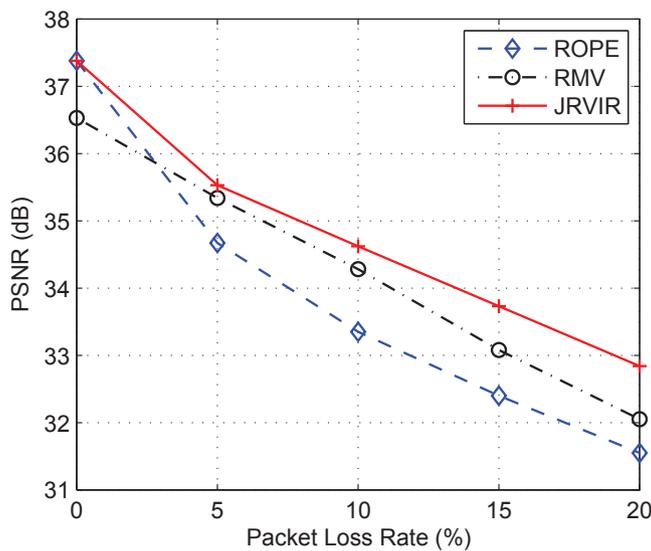
**end for**

In order to further evaluate the error-resilient performance of the JR VIR approach, we compare the video quality for different packet loss rate in Figure 3, with GOP length 150 and 15 in Figure 3.a, b, respectively. CIF Foreman sequence is used, and the target bitrate is 1 Mbps. From the figures, we can see that for different packet loss rates (0-20%) and GOP length the JR VIR approach can provide the best video quality among the three approaches. In Figure 4 video quality versus the

bitrate is presented for the three approaches. CIF Foreman sequence is used, the packet loss rate is 10%, and the GOP length is 150 and 15. In the 200 Kbps to 1 Mbps bitrate range, the proposed JR VIR approach outperforms the other two approaches, and the performance gap between JR VIR and the other two approaches increases with the bitrate. In both Figures 3 and 4, it is interesting to note that, with long GOP length ROPE can provide better video quality than RMV, while for short GOP length, RMV outperforms

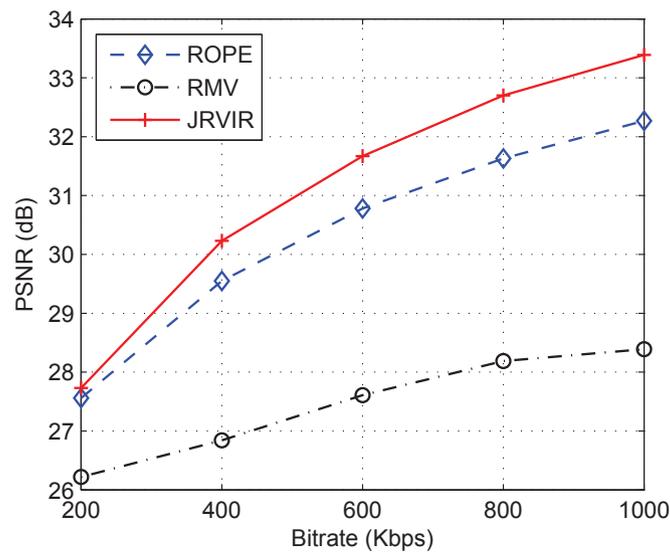


(a) GOP length 150

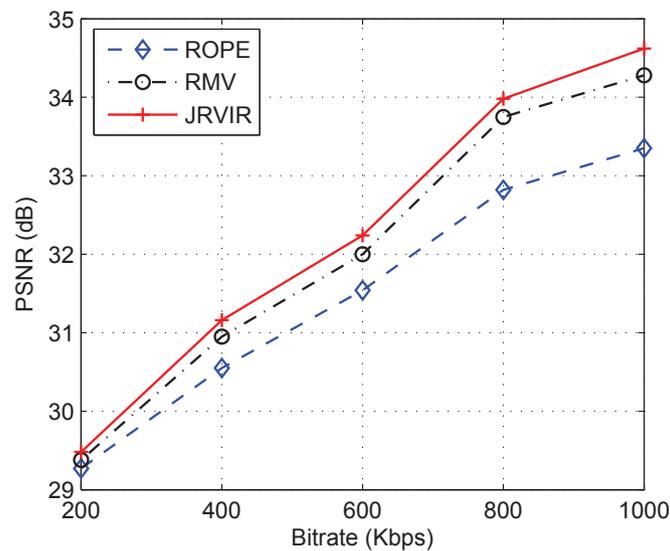


(b) GOP length 15

**Figure 3** PSNR comparison under different packet loss rates, CIF Foreman sequence is used, target bitrate is 1 Mbps, a) GOP length 150 b) GOP length 15.



(a) GOP length 150



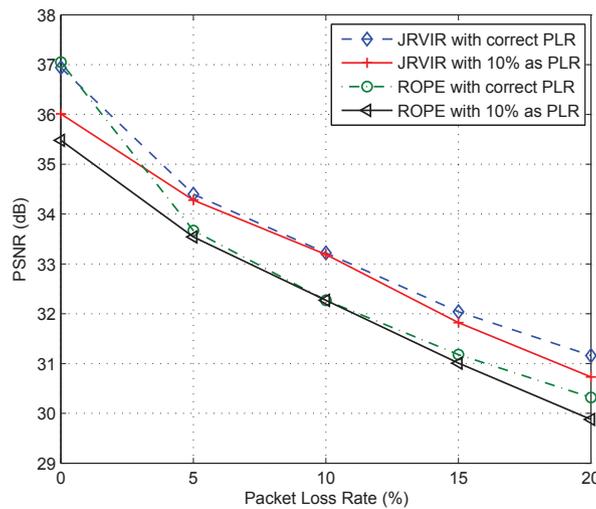
(b) GOP length 15

**Figure 4** PSNR comparison under different bitrate, packet loss rate is 10%, CIF Foreman sequence is used, a) GOP length 150 b) GOP length 15.

ROPE. This is because, in the ROPE approach, intra coding macroblocks are optimally inserted, the PSNR inside one GOP is more stable than RMV, while for the RMV approach, intra coding is not used, and consequently the PSNR inside one GOP drops incessantly.

In all the previous experiments, the channel packet loss rate is assumed to be available at the encoder, this can be implemented with the real time control protocol (RTCP) [20]. However, in practical situation, feedback packet loss rate information may be delayed from the

decoder. Therefore, the packet loss rate used by the encoder in its RD optimization process may not be exactly identical to the actual packet loss rate. To further evaluate the performances of the proposed JRVIR approach when the estimated packet loss rate does not match the actual one, we use 10% packet loss rate in the RD optimization process, whereas, the actual packet loss rate is varied from 0% to 20%. Figure 5 shows that the penalty because of the mismatch of the packet loss is not significant, especially when the packet



**Figure 5** JRVIR performance when packet loss rate (PLR) used in the encoder mismatches the actual PLR. The PLR used for RD optimization is 10%, and CIF Foreman sequence is used, GOP length 150.

loss rate is in the range from 5% to 15%, the gap is less than 0.25 dB. Comparing with the ROPE approach, the proposed JRVIR approach can provide better video quality when the estimated packet loss rate is not matched with the real situation.

In Table 1, simulation results for video sequences with varies degree of movement and bitrate are presented, and the percentage of intra macroblocks and macroblocks with redundant motion vector is given. For all the video sequences, GOP length is 150. We can observe that, in different test environments, the

proposed JRVIR approach always outperforms both ROPE and RMV approach. It is interesting to notice the JRVIR approach uses less intra macroblocks than ROPE, so as to allocate bitrate for redundant motion vector. Note that for the ROPE approach, the higher the packet loss rate, the more macroblocks are encoded with intra mode, whereas for the JRVIR approach, the total number of intra macroblocks and macroblocks with redundant motion vector increases. Table 1 shows that for the Foreman and Stefan sequences, nearly 20% of all the macroblocks are encoded with redundant motion

**Table 1** Video quality of joint redundant motion vector and intra macroblock refreshment (JRVIR), redundant motion vector (RMV) and recursive optimal per-pixel estimate (ROPE) for different bitrate and packet loss rate

Sequence	Rate (Kbps)	Method	Packet loss rate (PLR)			
			5%	10%	15%	20%
News	256	RMV	30.89dB	29.70dB	28.71dB	27.70dB
		ROPE	32.27dB (2.68%)	31.42dB (3.56%)	30.78dB (4.15%)	30.20dB (4.62%)
		JRVIR	<b>32.29dB</b> (1.54, 2.52)%	<b>31.70dB</b> (1.82, 3.90)%	<b>31.18dB</b> (2.11, 4.41)%	<b>30.73dB</b> (2.46, 4.94)%
Silent	384	RMV	30.75dB	29.00dB	28.22dB	27.52dB
		ROPE	33.56dB (4.70%)	32.75dB (5.96%)	32.10dB (6.76%)	31.60dB (7.29%)
		JRVIR	<b>33.76dB</b> (2.93, 3.49)%	<b>33.05dB</b> (3.43, 5.08)%	<b>32.45dB</b> (3.95, 5.76)%	<b>31.96dB</b> (4.54, 6.04)%
Foreman	512	RMV	29.32dB	27.42dB	25.89dB	24.56dB
		ROPE	31.48dB (6.78%)	30.29dB (9.40%)	29.42dB (11.32%)	28.66dB (13.06%)
		JRVIR	<b>32.03dB</b> (3.84, 18.94)%	<b>31.16dB</b> (4.84, 25.40)%	<b>30.32dB</b> (5.54, 27.52)%	<b>29.56dB</b> (6.90, 27.97)%
Highway	1024	RMV	34.73dB	32.54dB	30.07dB	29.48dB
		ROPE	37.71dB (11.68%)	36.64dB (15.09%)	35.76dB (16.52%)	35.01dB (18.10%)
		JRVIR	<b>38.06dB</b> (7.10, 9.87)%	<b>37.20dB</b> (8.80, 12.34)%	<b>36.50dB</b> (10.25, 13.85)%	<b>35.74dB</b> (11.70, 13.67)%
Stefan	2048	RMV	25.02dB	21.81dB	19.69dB	18.22dB
		ROPE	28.31dB (15.27%)	26.63dB (19.53%)	25.49dB (22.13%)	24.60dB (23.65%)
		JRVIR	<b>29.54dB</b> (6.38, 19.85)%	<b>27.50dB</b> (9.59, 18.81)%	<b>26.09dB</b> (12.86, 16.82)%	<b>24.99dB</b> (15.24, 14.71)%

For ROPE the percentage of intra macroblock is provided in brackets, while for JRVIR the first number in brackets is the percentage of intra macroblock, the second is the percentage of macroblock with redundant motion vector. The bold number is the highest PSNR among the three approaches.

vectors. Accordingly, the gaps between the JR VIR and ROPE approaches for these two video sequences are relatively larger than other sequences.

The actual network loss behavior have addressed by many papers, and it is agreed that Internet packet loss often exhibits finite temporal dependency, which means if current packet is lost, then the next packet is also likely to be lost. This leads to burst packet losses, with average burst length two for the Internet [21]. Therefore, besides *i.i.d.* random packet loss model, we also use burst loss model for simulation, and as indicated in [21], we set the average burst length as two. In practical burst loss environments, the transmission order of the primary and redundant packets would affect the performance. In our simulations, all the redundant packets of one frame are transmitted after the last primary packet of this frame, therefore there is no interleaving delay. In Figure 6, the PSNR versus bitrate curves in burst loss environments are plotted. The results are similar with that in the *i.i.d.* case, and the proposed JR VIR approach can provide best video quality among the three approaches. This makes us conclude that, the error resilient performance of proposed JR VIR approach is robust on different error distribution models.

In Table 2, we compare the encoding time of JR VIR with JM 14.0. In order to have fair comparison, we use the same configuration file for the two approaches. It is interesting to see that, the time costs for the two approaches are quite similar. In all cases, JR VIR costs less than 5% extra encoding time, this makes the JR VIR approach suitable for the real-time hand-device applications, where the battery capacity is usually the

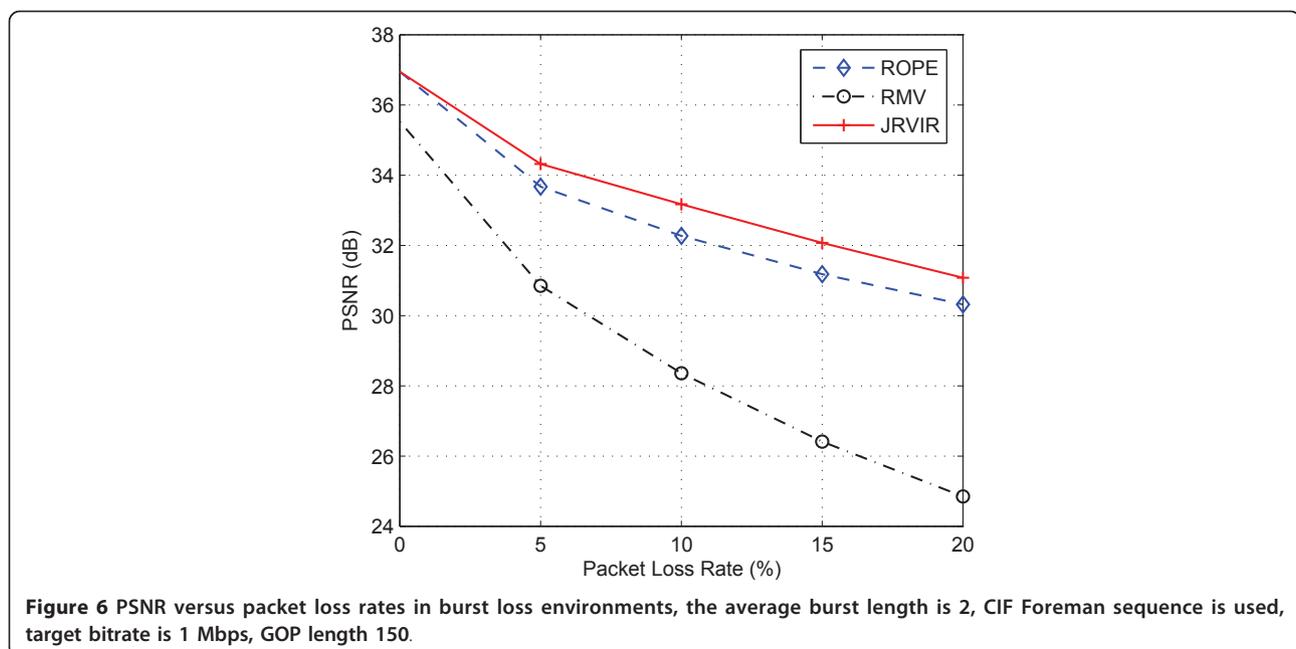
**Table 2 The time spent on encoding 30 frames for various video sequences and bitrates, for joint redundant motion vector and intra macroblock refreshment (JR VIR) and JM software, we assume the packet loss rate is 10%**

Sequence	Bitrate (Kbps)	JR VIR (s)	JM 14.0 (s)
News	256	41.19	40.79
Silent	384	40.51	39.35
Foreman	512	42.86	40.97
Highway	1024	42.61	41.63
Stefan	2048	42.25	40.87

bottleneck. This is because in the H.264/AVC encoding process, the motion estimation step is the main time-consuming task, so in comparison with this step, the end-to-end distortion calculation and new mode selection task costs much less time.

### 5. Conclusions

In this paper, a novel joint redundant motion vector and intra macroblock refreshment approach has been proposed to combat packet loss. Besides the traditional skip, inter and intra mode, we add a set of new modes, which are inter coding modes with redundant motion vector. Given the packet loss rate and the channel bitrate, the reconstructed distortion at the decoder side and the total bitrate for each mode are estimated at the encoder during the mode selection process. Based on the estimated end-to-end RD cost, the optimal encoding mode is selected. Equipped with the two tools, namely intra macroblock refreshment and redundant motion vector, extensive experimental results show that the



proposed approach outperforms other error-resilient approaches. Our future work would be investigating more sophisticated prediction method to compress the redundant motion vector. In addition, in this paper, we use pixel accuracy motion estimation and motion prediction, extending the current approach to quarter-pixel accuracy motion estimation and motion prediction would also be promising.

#### Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 60972085, No. 60903066), the Sino-Singapore JRP (No. 2010DFA11010) and National Science Foundation of China for Distinguished Young Scholars (No. 61025013).

#### Author details

<sup>1</sup>Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, L69 3GJ, UK <sup>2</sup>Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, 111 Ren Ai Road, Suzhou, P. R. China <sup>3</sup>Beijing Key Laboratory of Advanced Information Science and Network Technology, Institute of Information Science, Beijing Jiaotong University, Beijing, 100044, P.R. China

#### Competing interests

The authors declare that they have no competing interests.

Received: 7 January 2011 Accepted: 30 September 2011

Published: 30 September 2011

#### References

1. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 ISO/IEC 14496-10 AVC), document JVT-G050.doc, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. 2003.
2. Wenger S: **H.264/AVC over IP**. *IEEE Trans Circ Syst Video Technol* 2003, **13**(7):645-656.
3. Stockhammer T, Hannuksela MM, Wiegand T: **H.264/AVC in wireless environments**. *IEEE Trans IEEE Trans Circ Syst Video Technol* 2003, **13**(7):657-673.
4. Côté G, Kossentini F: **Optimal intra coding of blocks for robust video communication over the internet**. *Signal Process Image Commun* 1999, **15**(1-2):25-34.
5. Zhu QF, Kerofsky L: **Joint source coding, transport processing and error concealment for H.323-based packet video**. In *Proc SPIE VCIP'99. Volume 3653*. San Jose, CA; 1999:52-62.
6. Zhang R, Regunathan SL, Rose K: **Video coding with optimal inter/intra-mode switching for packet loss resilience**. *IEEE J Selected Areas Commun* 2000, **18**(6):966-976.
7. Stockhammer T, Kontopodis D, Wiegand T: **Rate-distortion optimization for JVT/H.26L coding in packet loss environment**. *Paper presented at the Packet Video Workshop 2002* Pittsburgh, PA, April; 2002.
8. Zhang Y, Gao W, Lu Y, Huang Q, Zhao D: **Joint source-channel rate-distortion optimization for H.264 video coding over error-prone networks**. *IEEE Trans Multimedia* 2007, **9**(3):445-454.
9. Zhu CB, Wang YK, Hannuksela MM, Li HQ: **Error resilient video coding using redundant pictures**. *IEEE Trans Circ Syst Video Technol* 2009, **19**(1):3-14.
10. Tillo T, Grangetto M, Olmo M: **Redundant slice optimal allocation for H.264 multiple description coding**. *IEEE Trans Circ Syst Video Technol* 2008, **18**(1):59-70.
11. Radulovic I, Frossard P, Wang YK, Hannuksela M, Hallapuro A: **Multiple description video coding with H.264/AVC redundant pictures**. *IEEE Trans Circ Syst Video Technol* 2010, **20**(1):144-148.
12. Dissanayake MB, Hewage CTER, Worrall ST, Fernando WAC, Kondoz AM: **Redundant motion vectors for improved error resilience in H.264/AVC coded video**. *Proc IEEE ICME* 2008, 25-28.
13. Chen JR, Lu CS, Fan KC: **A significant motion vector protection-based error-resilient scheme in H.264**. *Proc IEEE 6th Workshop on Multimedia Signal Processing* 2004, 287-290.
14. Sullivan GJ, Wiegand T: **Rate-distortion optimization for video compression**. *IEEE Signal Process Mag* 1998, **15**(6):74-90.
15. Yang H, Rose K: **Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC**. *IEEE Trans Circ Syst Video Technol* 2007, **17**(7):845-856.
16. Heng BA, Apostolopoulos JG, Lim JS: **End-to-end rate-distortion optimized mode selection for multiple description video coding**. *Proc IEEE ICASSP* 2005, 5:905-908.
17. Wan S, Izquierdo E: **Rate-distortion optimized motion-compensated prediction for packet loss resilient video coding**. *IEEE Trans Image Process* 2007, **16**(5):1327-1338.
18. Yang H, Rose K: **Optimizing motion compensated prediction for error resilient video coding**. *IEEE Trans Circ Syst Video Technol* 2010, **19**(1):108-118.
19. Available online at. [http://iphome.hhi.de/suehring/tml/download].
20. Schulzrinne H, Casner S, Frederick R, Jacobson V: **RTP: A transport protocol for real-time applications**. *Internet Engineering Task Force - RFC 1889* 1996.
21. Loguinov D, Radha H: **End-to-end internet video traffic dynamics: Statistical study and analysis**. *Proc of IEEE INFOCOM '02* 2002, 723-732.

doi:10.1186/1687-5281-2011-12

Cite this article as: Xiao et al.: Joint redundant motion vector and intra macroblock refreshment for video transmission. *EURASIP Journal on Image and Video Processing* 2011 **2011**:12.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)