

STREAMING VIDEO WITH BANDWIDTH ADAPTATION AND ERROR CONCEALMENT FOR LOW BIT RATE LIVE WIRELESS APPLICATIONS

Amir Asif, Uyen T. Nguyen, Guohua Xu, and Bin Song

Department of Computer Science and Engineering
York University, Toronto, ON, Canada M3J 1P3

ABSTRACT

We propose a real-time video transmission scheme, referred to as SNP/VQR, which is capable of providing spatial and temporal scalabilities. SNP/VQR avoids error propagation by combining bandwidth adaptability with error concealment. A real-time, software implementation of SNP/VQR is tested over a wireless network with simulated 5% to 20% of packet losses. SNP/VQR produces a relatively constant visual quality in our packet loss studies.

1. INTRODUCTION

With the emergence of broadband wireless channels, there is a renewed interest in real time video communications over wireless IP networks. Scalable video coding [1] is the preferred mode for sending video over wireless networks, where a subset of the compressed bit stream provides the base quality. Additional enhancement layers build on to the video quality in the spatial and temporal domains. Scalability allows the server to dynamically adjust to the bandwidth variations in the network by transmitting a reduced number of layers at times of congestion.

The paper presents a video compression scheme for transmission of streaming video over wireless networks. The scheme builds on the noncausal predictive codec with vector quantization and conditional replenishment (SNP/VQR) [2] in the following ways.

1. SNP/VQR uses a backward unilateral prediction approach based on modeling the video with a noncausal Gauss Markov random process (GMrp). Starting from the last frame ($k = N_k$), backward prediction encodes the video sequence in the reverse order ($N_k \geq k \geq 1$). In this paper, we derive an equivalent *forward* unilateral prediction model, which recursively transforms the video, in its natural order ($1 \leq k \leq N_k$), into a 3D uncorrelated error field. Forward prediction is, therefore, more suitable for live streaming video.
2. We propose using the forward SNP/VQR for flow and error control for low bit rate video communications. We design a bandwidth scalable version of SNP/VQR capable of adjusting its transmission rate according to the changing network conditions. This solves the flow control problem while error propagation is avoided by using an error concealment approach based on spatial and temporal median filtering.

A software implementation of the forward SNP/VQR is successfully tested for real time video cellular telephony over a wireless IP network with 5% to 20% of simulated packet losses. Our results show that the performance of the forward SNP/VQR is comparable

to the performance of MPEG4 under ideal transmission conditions. In our simulated packet loss studies, we observe that SNP/VQR produces a constant quality video over the entire transmission.

Section 2 reviews SNP/VQR and derives the forward unilateral prediction model. Section 3 describes the bandwidth adaptation and error concealment features of the forward SNP/VQR. Experimental results, including a comparison with MPEG4, are presented in section 4. Section 5 concludes the paper.

2. SNP/VQR CODEC

As shown in Fig. 1, the compression procedure of SNP/VQR has a predictive component followed by vector quantization (VQ). The predictive component has three stages: (1) Estimation of the vertical, horizontal, and temporal interactions $\{\beta_h, \beta_v, \beta_t\}$; (2) Transformation of the 3D noncausal model to an equivalent unilateral representation $\{L, F\}$; and (3) Generation of the uncorrelated error field $\vec{v}$ using a 3D noncausal Gauss Markov prediction model.

To achieve low bit rates, the error video $\vec{v}$ is compressed using the cascaded VQ shown in part II of Fig. 1. We apply conditional replenishment [3] at each stage of the cascaded VQ, where a VQ block is transmitted only if it is substantially different from the corresponding VQ block at the same location in the previous frame. The decoder reconstructs the video by inverting the steps of the encoder in the reverse order and is shown in part III of Fig. 1. We now develop the forward prediction model for SNP/VQR.

Forward Unilateral Prediction: Extending Woods' minimum mean square error (MMSE) approach [4] to 3D models, a first order video is represented by the bilateral autoregressive linear model

$$x_{ijk} - \beta_v(x_{i+1jk} + x_{i-1jk}) - \beta_h(x_{ij+1k} + x_{ij-1k}) - \beta_t(x_{ijk+1} + x_{ijk-1}) = e(i, j, k), \quad (1)$$

where x_{ijk} is the pixel intensity at spatial location (i, j) in frame k of the $(N_I \times N_J \times N_K)$ input video and e_{ijk} is the correlated input noise. Using a row-major order to arrange pixels x_{ijk} in frame k into vector $X^{(k)}$ and then stacking the resulting vectors, gives

$$\vec{X} = [X^{(1)T} \ X^{(2)T} \ \dots \ X^{(N_K)T}]^T. \quad (2)$$

By rearranging pixels x_{ijk} in vector $\vec{X}$ and the correlated input noise e_{ijk} in $\vec{e}$, Eq. (1) is expressed as $\mathcal{A}\vec{X} = \vec{e}$. For a first order GMrp, (1), the covariance matrix Σ_x of the video field is $\sigma^2 \mathcal{A}^{-1}$, where σ^2 is the variance of the pixel intensities in the video. The structure of the potential matrix $\mathcal{A}$ is described in Theorem 1.

Theorem 1: The expression $\mathcal{A}\vec{X} = \vec{e}$ represents a 3D, first-order, noncausal GMrp with zero Dirichlet boundary conditions (b.c.) iff

$$\mathcal{A} = I_{N_K} \otimes A_1 + H_{N_K}^1 \otimes A_2 \quad (3)$$

This work was supported in part by the Natural Science and Engineering Research Council (NSERC), Canada under Grant No. 228415-03.

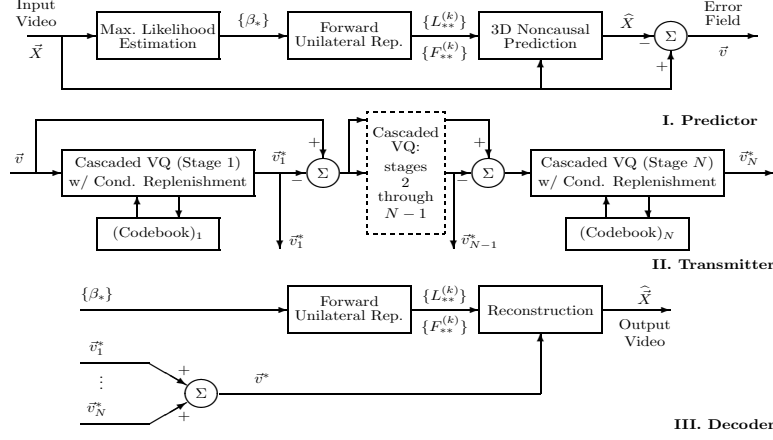


Fig. 1. Block Diagram of the forward SNP/VQR.

with $A_1 = I_{N_I} \otimes B + H_{N_I}^1 \otimes C$ and $A_2 = I_{N_I} \otimes D$ (4)

where $\otimes$ denote the Kronecker product. The constituent blocks are

$$B = -\beta_h H_{N_J}^1 + I_{N_J}, \quad C = -\beta_v I_{N_J}, \quad \text{and} \quad D = -\beta_t I_{N_J}. \quad (5)$$

In (3)-(5), symbols I_{N_K} and I_{N_I} denote identity matrices, while $H_{N_K}^1$ and $H_{N_I}^1$ are Toeplitz matrices that have zeros everywhere except for the first upper and lower diagonals, which are composed of all ones. The subscript denotes the order of the matrix. ■

Theorem 1 defines the prediction model ($\mathcal{A}\vec{X} = \vec{e}$) for a 3D non-causal field in terms of parameters $\{\beta_v, \beta_h, \beta_t, \sigma^2\}$. The structure of the potential matrix $\mathcal{A}$ includes pixels from both the past frame ($k-1$) and the future frame ($k+1$) for prediction of pixels in frame k . Such a representation precludes recursive computations. Theorem 2 derives an equivalent forward unilateral model.

Theorem 2: The following are equivalent representations for a 3D, first order, noncausal GMrp with zero Dirichlet b.c.

$$1. \text{ Bilateral Representation: } \mathcal{A}\vec{X} = \vec{e} \quad (6)$$

$$2. \text{ Unilateral Representation: } L^{(1)}X^{(1)} = v^{(1)} \quad (7)$$

$$\text{and } F^{(k)}X^{(k-1)} + L^{(k)}X^{(k)} = v^{(k)}, \quad (2 \leq k \leq N_K)$$

where $v^{(k)}$ represents the row major ordered pixels in frame k of the whitened error field $v(i, j, k)$, obtained from the transformation, $\vec{v} = \mathcal{L}^{-T}\vec{e}$. Blocks $\{L^{(k)}, F^{(k)}\}$ are the constituent blocks on the main and first lower block diagonals of the block bidiagonal matrix $\mathcal{L}$, derived by the Cholesky factorization $\mathcal{L}\mathcal{L}^T = \mathcal{A}$. All other block diagonals in $\mathcal{L}$ are zero blocks. ■

Since the covariance matrix $\Sigma_x = \sigma^2 \mathcal{A}^{-1}$, it is straightforward to show that the covariance $\Sigma_v = \sigma^2 I$. In other words, the noise vector $\vec{v}$ is white and we have completely uncorrelated the 3D video field with the unilateral transformation, (7). The Cholesky blocks $\{L_k, F_k\}$ are obtained directly from $\{A_1, A_2\}$ by expanding $\mathcal{L}^T \mathcal{L} = \mathcal{A}$ in terms of the constituent blocks as

$$L^{(k)} = \text{chol}(A_1 - \bar{\delta}_{kN_k} F^{(k+1)T} F^{(k+1)}), \quad (8)$$

with $F^{(k)} = (L^{(k)})^{-T} A_2$ for $(N_K \geq k \geq 1)$. The symbol $\bar{\delta}_{kN_k}$ is 0 for $k = N_K$. For all other values of k , $\bar{\delta}_{kN_k}$ is 1. We note that L_k is an upper triangular matrix, while F_k is lower triangular.

2.1. Structure of Cholesky Blocks

To illustrate the structures of the Cholesky blocks $\{L^{(k)}, F^{(k)}\}$, we run a simulation of (8) with $\beta_v = 0.156631$, $\beta_h = 0.166309$, and $\beta_t = 0.167446$ defined on a 3D ($24 \times 24 \times 24$) lattice. These values of field interactions are obtained from a real video sequence. For comparison, we also compute their steady state values $\{L^\infty, F^\infty\}$ by iterating (8) until convergence is achieved. Based on the computed values, we make the following observations.

Property 1: Fig. 2(a) plots the norms $\|L^{(k)} - L^\infty\|$ and $\|F^{(k)} - F^\infty\|$ for $(N_K \leq k \leq 1)$. The plot highlights the rapid geometric convergence of the sequences $L^{(k)}$ and $F^{(k)}$ with respect to k .

Property 2: Fig. 2(b) plots the norm $\|F_{\ell_1+1\ell_1+1}^\infty - F_{\ell_1\ell_1}^\infty\|$ of the constituent blocks in F^∞ as a solid line marked with symbol ‘*’. The norm $\|L_{\ell_1+1\ell_1+1}^\infty - L_{\ell_1\ell_1}^\infty\|$ in L^∞ is shown with symbol ‘o’. We observe that the constituent blocks in F^∞ and L^∞ themselves converge along the respective block diagonals ℓ_1 . Other nonzero subblocks outside the main block diagonal in both F^∞ and L^∞ also converge in a similar manner.

Property 3: Figs. 2(c) and 2(d) plot the norms of the blocks $\|L_{\ell_1\ell_2}^\infty\|$ and $\|F_{\ell_1\ell_2}^\infty\|$ along block row ℓ_1 . We note that these blocks converge to 0 along ℓ_1 on either side of the main diagonal of F^∞ .

Based on Properties 1-3, we approximate each Cholesky block $L^{(k)}$ in $\mathcal{L}$ by an M_1 block banded, lower triangular matrix. Similarly, each Cholesky block $F^{(k)}$ in $\mathcal{L}$ are approximated by an M_2 block banded, upper triangular matrix, such that

$$F_{\ell_1\ell_2}^{(k)} = 0 \vee (\ell_2 - \ell_1) \leq M_1 \quad \text{and} \quad L_{\ell_1\ell_2}^{(k)} = 0 \vee (\ell_1 - \ell_1) \leq M_2.$$

2.2. Practical Implementation

To derive the simplified implementation of the unilateral representation, we expand (7) with $L^{(k)}$ and $F^{(k)}$ approximated with M_1 -upper triangular and M_2 lower triangular block banded matrices, respectively. For $M_1 = M_2 = 3$, the resulting expressions are

$$\vee (1 \leq k \leq N_K) \text{ and } (1 \leq \ell_1 \leq N_I),$$

$$\bar{\delta}_{kN_k} \cdot \sum_{\tau=\ell_1}^{\min(\ell_1+3, N_I)} F_{\ell_1\tau}^{(k)} X_\tau^{(k-1)} + \sum_{\tau=\max(1, \ell_1-3)}^{\ell_1} L_{\ell_1\tau}^{(k)} X_\tau^{(k)} = v_{\ell_1}^{(k)}, \quad (9)$$

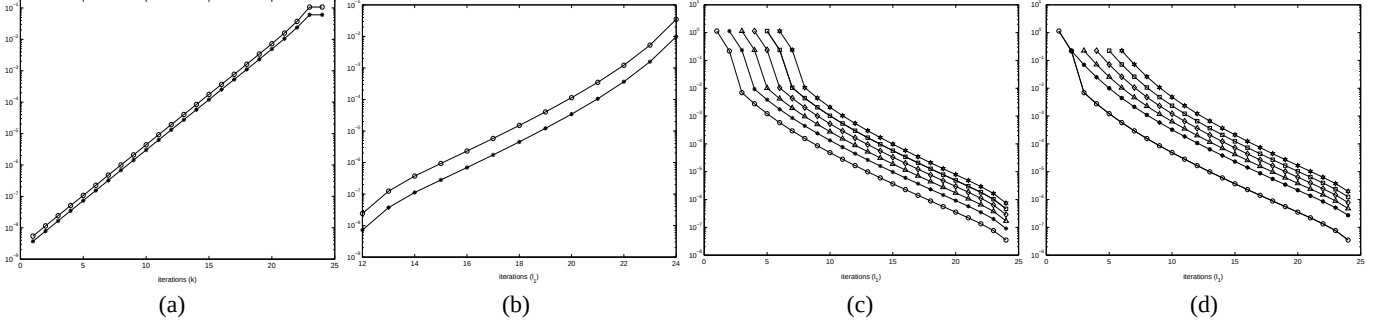


Fig. 2. Illustration of convergence in the Cholesky blocks for $\{\beta_v = 0.156631, \beta_h = 0.166309, \beta_t = 0.167446\}$. (a) Plots of $\|L^{(k)} - L^\infty\|$ and $\|F^{(k)} - F^\infty\|$ versus frame index k . (b) Plots of $\|L_{\ell_1+1\ell_1+1}^\infty - L_{\ell_1\ell_1}^\infty\|$ and $\|F_{\ell_1+1\ell_1+1}^\infty - F_{\ell_1\ell_1}^\infty\|$ versus block row ℓ_1 . (c) Plots of $\|L_{\ell_1\ell_2}^\infty\|$ for the last five rows ℓ_1 and $(1 \leq \ell_2 \leq \ell_1)$ in L^∞ . (d) Plots of $\|F_{\ell_1\ell_2}^\infty\|$ for the first five rows ℓ_1 and $(\ell_1 \leq \ell_2 \leq N_I)$ in F^∞ .

where the Cholesky blocks $\{F_{\ell_1\ell_2}^{(k)}, L_{\ell_1\ell_2}^{(k)}\}$ are obtained directly from blocks $\{B, C, D\}$ by expanding (8) and exploiting the block banded structure of the matrices.

The practical implementation of the forward SNP/VQR uses (9) to generate the whitened error image v_{ijk} , which is compressed using a 3-stage cascaded vector quantizer. Additional compression is achieved by applying conditional replenishment at each stage of the cascaded VQ, where a vector quantized block is transmitted only if it is substantially different from the corresponding vector quantized block at the same location in the previous frame. The compressed bit stream is then multiplexed with audio and other data types, if required, segmented into fixed length packets, and transmitted over the communication network.

3. FLOW AND ERROR CONTROL

For live video communications, reduction of transmission rates is only one of the required steps. A second equally important task is handling transmission errors and packet losses in the network. Current error control mechanisms include some form of retransmissions and/or forward error correction (FEC). Retransmission based approaches fail in real time applications, particularly when the round-trip propagation delay is large. FEC schemes add additional bits to the compressed data and are not particularly useful for low bit rate communications. In addition, FEC schemes are also deemed ineffective when packet losses are bursty. Since we are primarily interested in *live, low bit rate* multimedia applications, retransmission and FEC based approaches are not viable options for us. Instead, the forward SNP/VQR achieves error resilience by combining bandwidth adaptability with error concealment.

Bandwidth Adaptation: The practical implementation (9) of the forward SNP/VQR can achieve scalability in both the spatial and temporal domains. The *spatial scalability* is a direct consequence of cascaded VQ used to compress the prediction error. In the experimental setup, we use a 3-stage, 6-bit vector quantizer with a ‘321’-bit distribution between the three stages. For the best spatial quality video, the outputs of all three stages are transmitted to the receiver. The intermediate quality uses the outputs from stages 1 and 2, while the lowest quality uses only the output from stage 1.

The *frame rate scalability* is achieved by representing the input video sequence $X^{(k)}$, $(1 \leq k \leq N_K)$, in three layers. The base layer consists of every fifth frame, $X^{(4k-3)}$, and is transmitted to all receivers. The first enhancement layer encodes interme-

diates frames, $X^{(4k-1)}$, while the second enhancement layer encodes the remaining frames $X^{(2k)}$. To fully exploit the temporal redundancy, prediction in the first enhancement layer uses frames reconstructed from the base layer as well as the original frames $X^{(4k-1)}$. Similarly, prediction in the second enhancement layer uses frames reconstructed from the base and first enhancement layers in addition to the original frames $X^{(2k)}$.

By combining the spatial and frame rate scalabilities, the forward SNP/VQR offers three quality of services: Gold, Silver, and Bronze, at bit rates between 10kbps to 500kbps. The *bronze service* uses the base layer of the temporal feed compressed with only the first stage of the cascaded VQ. The *silver service* couples the base and first enhancement layers. Each layer is compressed with the first two stages of the cascaded VQ. Finally, the *gold service* uses all three temporal layers and a 3-stage cascaded VQ.

Using the above scalability feature, the forward SNP/VQR allows decoding at multiple rates from the same bit stream. Bandwidth adaptation is achieved by using different stages of the cascaded VQ and selecting some or all of the enhancement layers. This solves the flow control problem, while error propagation is avoided by concealing the effects of transmission errors.

Error Concealment: To conceal the distortion introduced by VQ blocks damaged due to transmission errors or lost due to packet losses, the forward SNP/VQR uses spatial and temporal interpolation. A lost packet in SNP/VQR implies that certain blocks of the error image $\vec{v}$ are not available for reconstruction of the video. Unavailable pixels in these damaged blocks are interpolated from correctly received pixels in adjacent blocks from the same and previous frames. To facilitate interpolation, the forward SNP/VQR interleaves the VQ blocks during packetization so that loss of one packet does not lead to loss of contiguous blocks.

4. EXPERIMENTS

We test the performance of the forward SNP/VQR when subjected to random packet losses during transmission. In our simulations, we use the monochrome video sequences *Carphone* with a QCIF resolution of (144×176) pixels per frame and a display rate of 30 frames/s. The test sequences are encoded with the forward SNP/VQR and transmitted over a channel with 5% to 20% of packet losses. We present results of a simulation with a data rate of 82.56 kbps and 20% of random packet loss. Data from the correctly received packets is used to reconstruct the video.

Fig. 3 plots the peak signal to noise ratio (PSNR) at 82.56 kbps for the following four schemes: 1. SNP/VQR without packet loss; 2. MPEG4 without packet loss; 3. SNP/VQR with packet loss and no compensation for error; and 4. SNP/VQR with packet loss and error concealment. Schemes 1 and 2 assume no transmission errors or loss of packets and allows us to compare the performance of the forward SNP/VQR with MPEG4 under ideal transmission conditions. Schemes 3 and 4 assume a 20% packet loss during transmission of the vector quantized field $\vec{v}$. In scheme 3, the luminance of any pixel in $\vec{v}$, whose information is lost due to packet loss, is set to zero. Scheme 4 replaces the values of such pixels with the median of the corresponding undamaged pixels in the three spatially neighboring blocks in the same frame and one block at the same location from the previous frame. To keep the decoder simple for real time applications, we avoid more complex approaches such as the projection onto convex sets (POCS) interpolation.

Fig. 3 illustrates that the performance of the forward SNP/VQR is comparable to the performance of MPEG4 at 82.56 kbps under ideal transmission conditions. Further comparison of the two codecs is completed in Table 1, which lists the mean PSNR (averaged over all frames) for video sequences compressed with the forward SNP/VQR and MPEG4 at bit rates below 125 kbps. Results from Table 1 support our earlier conclusion. Under conditions of 20% simulated packet loss, we observe from Fig. 3 that scheme 4 provides a fairly uniform quality of reconstructed video. Even a simple error concealment approach like median filtering used in the forward SNP/VQR eliminates error propagation. To provide subjective evaluation of the reconstructed video, a representative frame (frame 27) is extracted from the sequences compressed with schemes 1–4. Fig. 4 shows the original and four reconstructed frames. We observe that scheme 4 (Fig. 4(e)) conceals most of the distortion introduced by the loss of packets (Fig. 4(d)). All features of the original image corresponding to the lost packets are distinctly visible in the video reconstructed with scheme 4.

5. SUMMARY

In this paper, we successfully tested a real-time, bandwidth-scalable video transmission scheme for wireless applications. The proposed scheme allows spatial and frame control scalability at different rates from the same bit stream. This solves the flow control problem while error propagation is avoided by using a relatively simple time and spatial median filtering approach.

6. REFERENCES

- [1] W. Tan and A. Zakhor, "Video Multicast using Layered FEC and Scalable Compression," *IEEE Trans. Circuits, Syst. Video Technol.*, vol. 11(3), Mar. 2001, pp. 373-386.
- [2] M. Kouras and A. Asif, "Noncausal Predictive Video Codec Offering Hierarchical QoS," in *Proc. ICASSP*, vol. 3, 2004, pp. 737-80.
- [3] M. Goldberg and H. Sun, "Image Sequence Coding using Vector Quantization," *IEEE Trans. Communications*, vol. COM-34, 1986, pp. 792-800.
- [4] T. J. Woods, "Two Dimensional Discrete Markovian Fields," *IEEE Trans. Inform. Thry.*, vol. IT-18, Mar. 1972, pp. 232-40.
- [5] S. M. Schweizer and J. M. F. Moura, "Hyperspectral Imagery: Clutter Adaptation in Anomaly Detection," *IEEE Trans. Inform. Thry.*, vol. 46(5), Aug. 2000, pp. 1855-71.

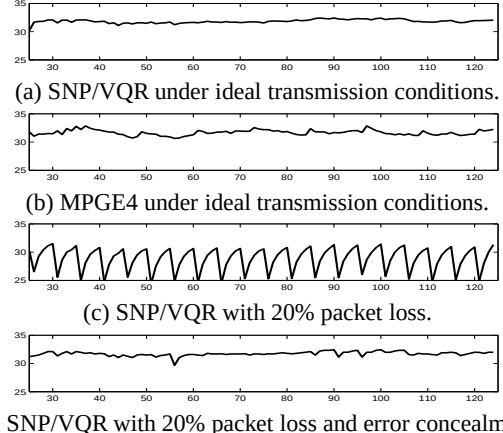


Fig. 3. Variability of PSNR at 20% simulated packet loss.

Bit rate (Kbps)	Mean PSNR	
	SNP/VQR	MPEG4
106.13	32.84 dB	32.15 dB
82.56	31.64 dB	31.20 dB
52.19	30.56 dB	29.74 dB
38.40	29.47 dB	29.05 dB
26.52	29.08 dB	28.61 dB
18.73	28.68 dB	28.46 dB

Table 1: Comparison of mean PSNR between SNP/VQR and MPEG4.

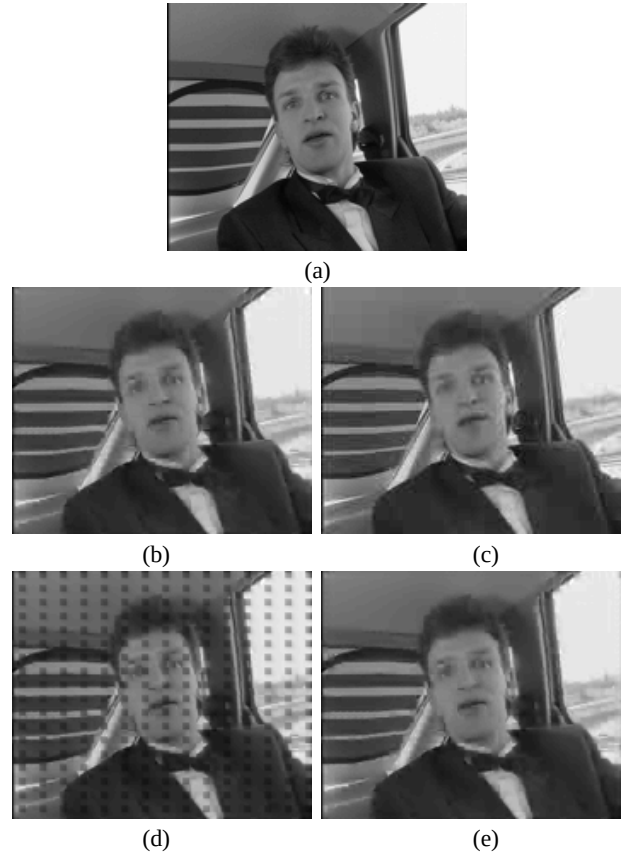


Fig. 4. Frame no. 27: (a) Original; Extracted from video sequences compressed to 82.56 kbps using: (b) SNP/VQR (scheme 1); (c) MPEG4 (scheme 2); (d) SNP/VQR with 20% packet loss (scheme 3); and (e) SNP/VQR with 20% packet loss but restored (scheme 4).