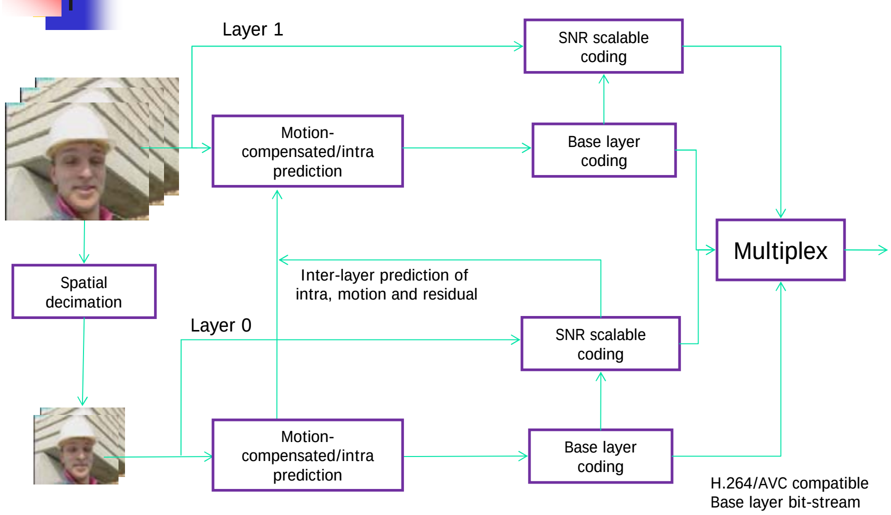
CASE IV

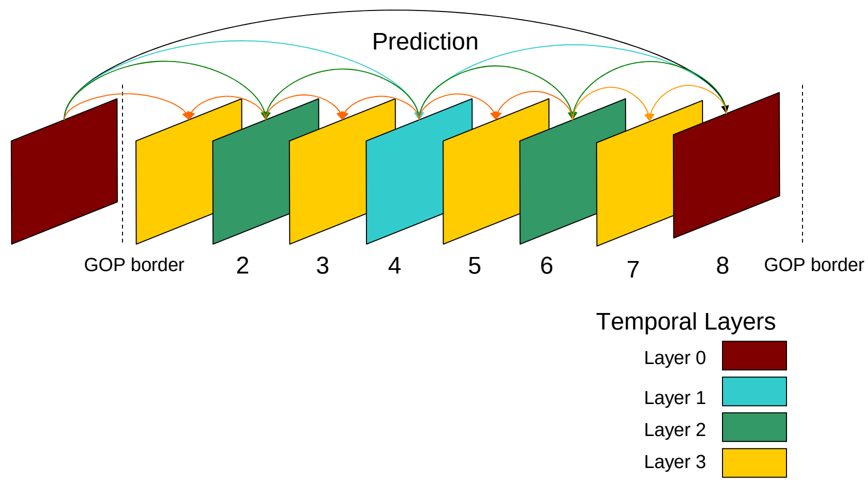
Optimal Rate Extraction and Channel Coding for Scalable Streaming over Packet Lossy Networks

SVC



Hierarchical prediction structure

Hierarchical prediction structure in a scalable video encoder

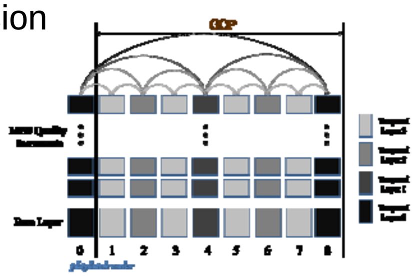


Problem Formulation

- Optimally extract source packets (BL or EL) from an SVC bit stream and allocate channel coding rate for efficient transmission

$$(\boldsymbol{\phi}^*, \boldsymbol{\psi}^*) = \min_{\boldsymbol{\phi} \in \Phi, \boldsymbol{\psi} \in \Psi} E\{D(\boldsymbol{\phi}, \boldsymbol{\psi})\}$$

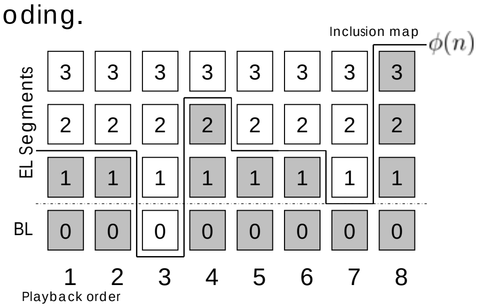
$$\text{s.t.} \quad R(\boldsymbol{\phi}, \boldsymbol{\psi}) \leq R_T,$$



$\phi(n)$	Inclusion map: specifies number of NAL units included for each packet
$\psi(n,q)$	Channel coding map: specifies channel rate of NAL unit
$R(\boldsymbol{\phi}, \boldsymbol{\psi}))$	Total rate of the video stream

Problem Formulation

- The minimization solution, ϕ^*, determines the inclusion map for the group of pictures
- The optimization requires evaluation of the distortion for any ϕ. Since ME/MCP in SVC is conducted using the highest available quality, missing quality increments cause *drift* to propagate to lower temporal levels. Thus, in principle any ϕ requires a decoding.



Distortion Estimation

- Total distortion of a frame originates from two sources:
  - D^d , distortion due to drift, inherited from parent frames
  - D^e , error due to quality increment truncation for the current frame
- Total distortion of frame *n*, $D^t_n(q)$, is estimated as

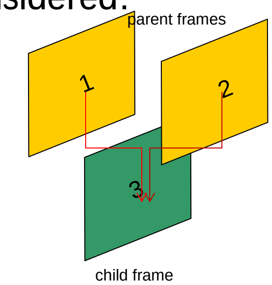
$$
\begin{aligned}
D^t_n(q) &= ||\boldsymbol{f}_n - \boldsymbol{f}^d_n + \boldsymbol{e}_n(q)||^2 \\
&= D^d_n + D^e_n(q) + 2(\boldsymbol{f}_n - \boldsymbol{f}^d_n)^T \boldsymbol{e}_n(q) \\
&\approx D^d_n + D^e_n(q) + 2\kappa\sqrt{D^d_n}\sqrt{D^e_n(q)} \\
&\leq D^d_n + D^e_n(q) + 2\sqrt{D^d_n}\sqrt{D^e_n(q)}
\end{aligned}
$$

$$0 \leq \kappa \leq 1$$

symbol	description
$\boldsymbol{f}^d_n$	Reconstructed image when drift exists
$\boldsymbol{e}_n(q)$	error vector due to inclusion of only q increments
$\boldsymbol{f}_n$	fully reconstructed image

Drift Distortion Estimation

- For any frame *n* two cases are considered:
  - Base layer decodable
    - Find coefficients α_i and β_{ij} from training data and use during optimization



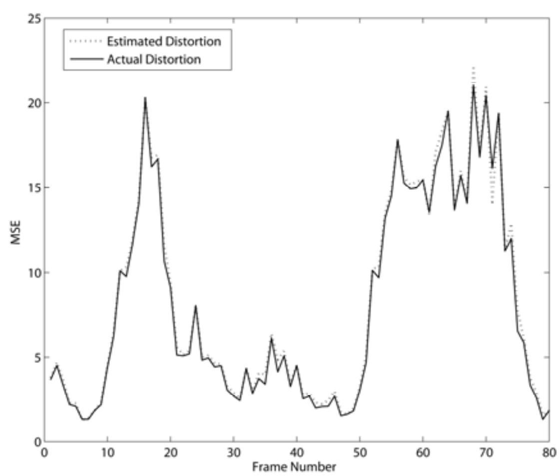
$$D_n^d \approx \sum_{i \in \Lambda_n} \alpha_i D_i^t + \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n} \beta_{ij} D_i^t D_j^t$$

  - Base layer missing/undecodable
    - For each concealment option "*i*" find μ_i and ν_i from training data and use during the optimization

$$D_{n,i}^{con} \approx \mu_i + \nu_i D_i^t,$$

Estimated Distortion

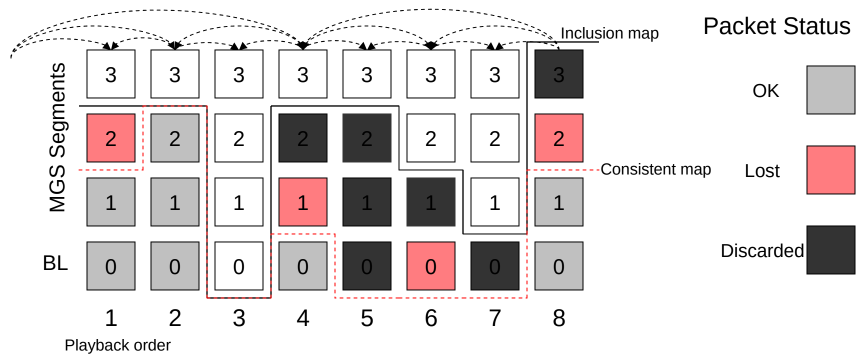
- Estimated vs. actual distortion for a random inclusion path (foreman sequence)



$\kappa = 0.05$

Expected Distortion

- Since a packet is not decodable unless all lower level packets (in both temporal and MGS directions) are received intact, any <u>consistent</u> received map that falls below the inclusion map is probable and has to be taken into account for expected distortion

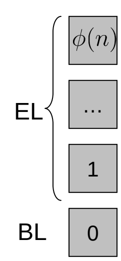


Expected Distortion

- Expected distortion is calculated in two steps. First, the conditional expected distortion given the base layer received is evaluated

$D_n(q)$ The n^{th} frame decoded using q enhancement segments

p_n^q Loss probability of the q^{th} segment of the n^{th} frame EL (BL: q=0)



- The conditional ED can be expressed as

$$E\{\tilde{D}_n|BL\} = \sum_{q=1}^{\phi(n)} p_n^q D_n(q-1) \prod_{i=0}^{q-1}(1-p_n^i) + D_n(\phi(n)) \prod_{i=0}^{\phi(n)}(1-p_n^i)$$

Expected Distortion

- In case that a base layer is lost, it is concealed using the nearest available (decodable) frame. Using this concealment strategy, the expected frame value is calculated according to

$$E\{\tilde{D}_n\} = \sum_{i \in \Delta_n} p_i^0 \, D_{n,k}^{con} \prod_{\substack{j \in \Delta_n \\ j \prec i}} (1 - p_j^0) + E\{\tilde{D}_n | BL\} \prod_{j \in \Delta_n} (1 - p_j^0)$$

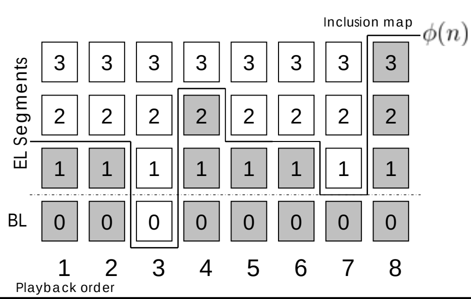
Where *"k"* indicates the nearest temporal ancestor frame

$E\{\tilde{D}_n \| BL\}$	The n^{th} frame conditional expected value
p_n^0	Loss probability of the of the base layer

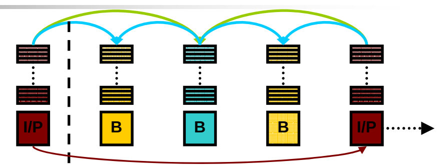
Source Optimization

- The base layer of the key picture is given the highest priority and is the first packet to be added. Then, packets are added one at a time based on their global distortion gradient, i.e., at time step i the next quality increment of frame n_i^* is included to the queue:

$$n_i^* = \arg\max_n \left| \frac{\partial D(\boldsymbol{\phi}) / \partial \phi(n)}{\partial R_s(\boldsymbol{\phi}) / \partial \phi(n)} \right|$$

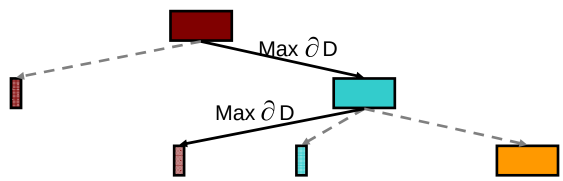


Packet Ordering Methods



- Basic Extractor
  - Base first

- Optimized Extractor
  - Maximum D gradient



Joint Source-Channel Optimization

- Similarly to source optimization, inclusion function $\phi(n)$ initially only includes the base layer of the key pictures with an initial channel coding rate less than 1, then
  - At each step find

$$\delta ED^* = \max_n \max_{q<\phi(n)} \left|\frac{\partial ED(\boldsymbol{\phi},\boldsymbol{\psi})/\partial\psi(n,q)}{\partial R_t(\boldsymbol{\phi},\boldsymbol{\psi})/\partial\psi(n,q)}\right|$$

$$\delta ED^\star = \max_n \max_{\psi(n,q)\in\Psi} \left|\frac{\partial^2 ED(\boldsymbol{\phi},\boldsymbol{\psi})/\partial\phi(n)\partial\psi(n,q)}{\partial^2 R_t(\boldsymbol{\phi},\boldsymbol{\psi})/\partial\phi(n)\partial\psi(n,q)}\right|, \quad q=\phi(n)$$

$\delta ED^* > \delta ED^\star$: channel protection rate of the already included Packet $\pi(n^*,q^*)$ is incremented to the next level by padding additional parity bits

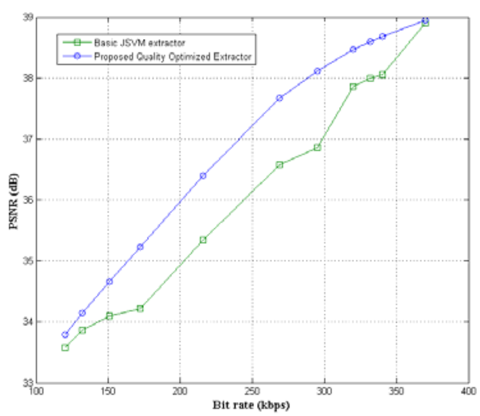
$\delta ED^* < \delta ED^\star$: the source packet $\pi(n^\star,q^\star)$ is included in the transmission queue with a channel coding rate of $\pi(n^\star,q^\star)$

Experimental Results

- Sequences are encoded using the scalable extension of the H.264/AVC (JSVM) reference software
- The GOP size is set to 8 and the base layer is encoded with quantization parameter *QP* = 36. Moreover, one enhancement layer (*QP* = 25) is considered
- The EL is then divided into 5 MGS layers
- For channel coding RS codes with symbol length of 5 are considered
- Bernoulli channel model

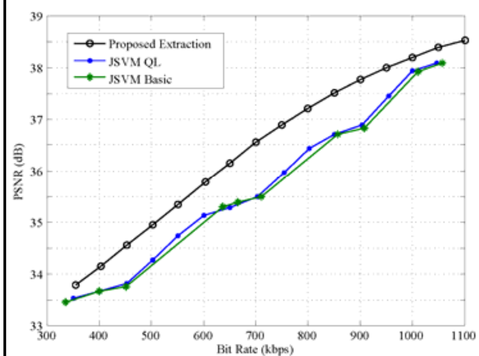
Source Optimized Extraction

- Foreman QCIF

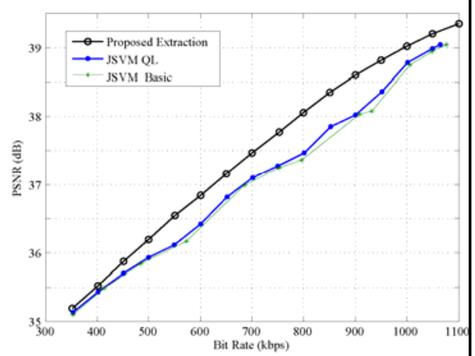


Source Optimized Extraction

City CIF



Foreman CIF



Source and Channel Encoding

Average decoded PSNR for FOREMAN QCIF :

(green) Joint Source Extraction and Channel Coding

(blue) Proposed Optimized Extraction with fixed channel coding (average rate of UEP system used)

(red) Basic extraction with a fixed channel coding rate (average rate of UEP system used)

Packet loss rate before channel coding is 0.10.