Poliphase Subsampling Multiple Description Coding Using Robust Edge-Preserving Interpolation

Stefania Colonnese[1], *Nicolás Herrero*[2], *Lorenzo Rossi*[1], *Gaetano Scarano*[1]

[1]INFOCOM Dep.t, Università "La Sapienza" di Roma
via Eudossiana 18, I-00184 Roma, Italy
{colonnese,rossi,scarano}@infocom.uniroma1.it

[2]Telefónica Research
Vía Augusta 177, Barcelona, Spain
nhm@tid.es

ABSTRACT

Multiple Description Coding (MDC) is aimed at achieving transmission diversity for error resilient purposes by transmitting different coded descriptions of the same data. In this paper we analyze a poliphase subsampling MDC technique implemented by pre-processing and post-processing stages of a H.264/AVC standard compliant encoder and decoder. The pre-processing step performs a suitable interleaving that builds a synthetic sequence where each frame carries different subsampled frames of the original sequence. The synthetic sequence is ordinarily encoded and transmitted. At the receiver side, the H.264 decoder performs error concealment exploiting the multiple data descriptions, and the de-interleaving stage generates a decoded version of the original sequence. A robust restoration algorithm based on edge-preserving interpolation is then applied. Simulations show that the analyzed scheme improves the decoded video sequence quality both from an objective and from a subjective point of view.

1. INTRODUCTION

Multiple description coding (MD Coding, or MDC) consists in providing different coded description of the same data, and to send them using different transport channels, achieving the transmission diversity needed for error resiliance. Surveys on principles in designing MD video coders and on different MD compression algorithms can be found in [1, 2]. MDC algorithms descriptions are generated by subsampling the data either in the spatial, temporal, or frequency domain. A possibly lost description can be estimated from the others by exploiting spatial or temporal adjacent video data samples correlation. In [3], two MDC algorithm based on polyphase down-sampling are investigated and their performances over unreliable networks assessed by numerical simulations. In [4] the authors analyze a mathematical framework for pre- and post-processing two descriptions of the original data, so as to implement the MDC paradigm by exploiting the native directional correlation characteristics of the image. Specifically, in the pre-processing stage the data splits into two subsets by means of a forward transform, that are separately encoded and transmitted. At the receiver side, data is recovered by an inverse transform making use only of the effectively available description. In [5, 6], MDC is achieved by originating four descriptions from the spatially downsampled polyphase components of the original frames; each description is independently H.264 coded, and concealment is applied at the decoder side in case of losses. In [7] a distributed video streaming framework using unbalanced MDC and unequal error protection for wavelet-based coders is proposed. In [8], a MDC technique based on the H.264/AVC slice group syntactic structure [9] is described. Recently, in [10] a novel MDC technique has been proposed, in the framework of H.264 coding. The coding algorithm exploits the H.264 redundant slices; at the receiver side, the received compressed bitstream must be pre-processed before being applied at the input of a standard compliant decoder.

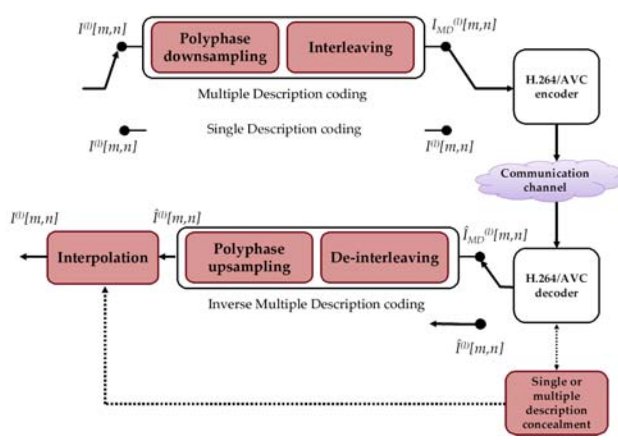

Figure 1: MDC scheme proposed.

In this paper we analyze an error resilient MDC scheme based on Polyphase SubSampling (PSS MDC). The analyzed MDC scheme is shown in Fig.1. The original video sequence is applied at the input of a spatial-temporal interleaving stage that generates a synthetic sequence, in which each frame conveys a fixed number of descriptions pertaining to different frames of the original sequence. The synthetic sequence is applied at the input of a standard video encoder; the transport channel diversity is achieved simply by mapping different encoded frames into different transport packets. Let us observe that, since MDs are managed only inside the interleaving stage, switching from a MDC scheme to a Single Description (SD) coding one can be performed dynamically at the encoder input at the expence of an interleaving delay, easily recoverable by suitable buffering at the decoder side. At the receiver side, the synthetic sequence is decoded and concealed using available MD and de-interleaving is applied to provide a coarse reconstruction of the original video sequence.

Once this coarse estimate of the video sequence has been provided, a fast restoration algorithm based on robust interpolation is applied. The interpolation algorithm exploits the local image directionality feature as well as the information on which descriptions have been correctly received and which have been concealed and it effectively improves the decoded video sequence quality both from an objective and from a subjective point of view. The herein analyzed scheme presents two interesting properties: the MDC technique employs a standard compliant video coding stage, so that it can be implemented at a slightly increased computational cost; besides, it appears to the transport layer as a SD coded data flow, and the increase of protocol overhead is limited, too.

The remainder of the paper is organized as follows: in Sect.2 the overall MDC coding scheme is outlined, while in Sect.3 the standard compliant encoding/decoding stages providing a coarse video sequence reconstruction are described. Sect.4 describes the final restoration stage performing a robust edge-preserving interpolation; the results of numerical simulations assessing the MDC algorithm performance are shown in Sect.5. The paper is concluded in Sect.6.

2. MD GENERATION

PSS MDC generates descriptions of each single frame by subsampling it with different initial phases in vertical and/or horizontal directions; each subsampled image provides a simplified frame description, and the original frame is recovered by suitably collecting different descriptions.

Benefits of PSS MDC are achieved when the different descriptions are transmitted in diversity. However, independent MD encoding and transmission not only results into heavier computational requirements due to encoding and decoding different descriptions, but also into increased protocol overhead and reduced bandwidth efficiency[1] Here, we describe how PSS MDC can be realized by an application layer interleaving scheme operating at the input of a standard compliant encoding module, so as to bound the computational requirement. Furthermore, the interleaving generates a synthetic sequence in which each frame contains different subsampled descriptions pertaining to different frames of the original video sequence; thus, diversity is straightforwardly achieved when each coded frame of the interleaved sequence is mapped into at least one transport-layer packet.

Let us denote the l-th frame of the original video sequence, of dimensions $M \times N$ by

$$I^{(l)}[m,n],\ m = 0,\cdots,M-1, n = 0,\cdots N-1$$

and let us suppose that a $K \times K$ downsampling factor is employed in the PSS stage, so that each frame is conveyed by $K \times K$ descriptions. The j-th poliphase subsampled description of the l-th frame, of size $M/K \times N/K$, is given by

$$\begin{aligned}&\Delta_j^{(l)}[m,n] \stackrel{\text{def}}{=} I^{(l)}[Km+m_j, Kn+n_j],\ j = 0,\cdots K^2-1\\ &m_j = j_{\text{mod}K}, n_j = \lfloor j/K \rfloor,\\ &m = 0,\cdots,M/K-1, n = 0,\cdots N/K-1.\end{aligned} \quad (1)$$

The K^2 descriptions can be juxtaposed into a single $M \times N$ frame, so as to associate at the original video sequence a new spatially interleaved sequence:

$$\begin{aligned}&I_{\text{SI}}^{(l)}[m+m_j*M/K, n+n_j*N/K] \stackrel{\text{def}}{=} \Delta_j^{(l)}[m,n],\\ &m = 0,\cdots,M/K-1, n = 0,\cdots N/K-1, j = 0,\cdots K^2-1,\end{aligned} \quad (2)$$

The spatially interleaved sequence $I_{\text{SI}}^{(l)}[m,n]$ exhibits more rapid luminance variations than the original sequence $I^{(l)}[m,n]$, thus presenting a higher coding cost. However, since in video coding intensive prediction techniques are used, the overall coding cost is strongly related to interframe correlation, and a suitable temporal interleaving increases the coding efficiency. Then, on the spatially interleaved sequence, a temporal interleaving is applied, aiming at

- assigning different descriptions of the j-th frame to different application layer packets
- preserving the inter-frame correlation properties typical of natural video sequences.

The spatio-temporal interleaved sequence is built as

$$I_{\text{MD}}^{(l)}[m+m_j*M/K, n+n_j*N/K] = \Delta_j^{(l+j)}[m,n] \quad (3)$$

for $m = 0,\cdots,M/K-1, n = 0,\cdots N/K-1, j = 0,\cdots K^2-1$ From (3), we recognize that the l-th frame of the interleaved sequence $I_{\text{MD}}^{(l)}[m,n]$ conveys K^2 subsampled descriptions, with different sampling phases, pertaining to K^2 different frames of the original sequence $I^{(l)}[m,n]$, namely $l, l+1,\cdots,l+K^2-1$ of the original sequence $I^{(l)}[m,n]$; conversely, the K^2 descriptions of each frame of $I^{(l)}[m,n]$ are conveyed by K^2 different frames of $I_{\text{MD}}^{(l)}[m,n]$. Thus, the interleaving introduces diversity when each frame of the interleaved sequence $I_{\text{MD}}^{(l)}[m,n]$ is sent using a different transport packet; this condition is a minimal requirement [**?**] that is expected to be satisfied by all video communication systems. Such interleaving also preserves the inter-frame correlation, in fact each description follows the correspondant description of the previous frame.

For instance, let us fix $K = 2$. Then, given the original sequence $I^{(l)}[m,n]$, the PSS followed by the interleaving stage generates the MD sequence $I_{\text{MD}}^{(l)}[m,n]$:

$$\begin{array}{ccccc} \cdots\left|\begin{matrix}\Delta_0^{(l-1)} & \Delta_1^{(l)}\\ \Delta_2^{(l+1)} & \Delta_3^{(l+2)}\end{matrix}\right\| & \begin{matrix}\Delta_0^{(l)} & \Delta_1^{(l+1)}\\ \Delta_2^{(l+2)} & \Delta_3^{(l+3)}\end{matrix}\Big\| & \begin{matrix}\Delta_0^{(l+1)} & \Delta_1^{(l+2)}\\ \Delta_2^{(l+3)} & \Delta_3^{(l+4)}\end{matrix}\Big\| & \begin{matrix}\Delta_0^{(l+2)} & \Delta_1^{(l+3)}\\ \Delta_2^{(l+4)} & \Delta_3^{(l+5)}\end{matrix}\Big|\cdots \\ \Uparrow & \Uparrow & \Uparrow & \Uparrow \\ \ldots\ \text{frame } l-1, & \text{frame } l, & \text{frame } l+1, & \text{frame } l+2, \ \ldots \end{array}$$

whose l-th frame is built by juxtapposing the $K^2 = 4$ descriptions $\Delta_0^{(l)}, \Delta_1^{(l+1)}, \Delta_2^{(l+2)}, \Delta_3^{(l+3)}$ pertaining to the frames $l, l+1, l+2, l+3$ of the original video sequence.

3. MD ENCODING AND DECODING

The MD sequence $I_{\text{MD}}^{(l)}[m,n]$ is applied at the input of a standard video encoder, and transmission diversity of MD is implicitly assured when each coded frame is mapped into at least one independent transport packet. From now on, and without loss of generality, we will refer to a video communication scheme based on the most recent Joint Video Team coding standard H.264 [9]. In the H.264 framework, each

[1] In almost all the emerging video communication schemes, it is recommended a one-to-one correspondence between application layer packets and transport packets [**?**], resulting in protocol overhead in case of independent MD transmission.

frame can be coded in one or more Network Adaptation Layer Unit (NALU), and each NALU can be coded in one or more transport packets. In either case, coded video data pertaining to different frames are expected to be conveyed by different transport packets when IETF recommended packetization [11] is applied. Hence, from a technical point of view, the herein analyzed MDC scheme has the merit to bound the visibility of MDC to the application layer and to appear to the network as a unique media flow, so simplifying the protocol architecture required for the video communication.

At the receiver side, the video stream is decoded; in case of channel errors, data are lost and the decoder invokes error concealment procedures to provide the decoded video sequence $\hat{I}^{(l)}_{\mathrm{MD}}[m,n]$. In case of data loss, error concealment may use the same algorithms adopted in SD coding or exploit the MDC paradigm by searching in the decoder buffer for alternative descriptions of lost data. When the loss occurs on less than K^2 consecutive frames, the decoder can exploit the availability of MD pertaining to the same frame of $I^{(l)}[m,n]$ to recover the loss; otherwise it performs a generic concealment, for instance by exploiting descriptions belonging to adjacent frames.

Once the sequence $\hat{I}^{(l)}_{\mathrm{MD}}[m,n]$ has been generated, de-interleaving is applied to generate the decoded version $\hat{I}^{(l)}[m,n]$, which represents a coarse estimate of the trasmitted sequence $I^{(l)}[m,n]$.

4. VIDEO SEQUENCE RESTORATION BY MEANS OF ROBUST EDGE PRESERVING INTERPOLATION

After the decoding, error concealment and de-interleaving stages, the reconstructed video sequence $\hat{I}^{(l)}[m,n]$ is available. The luminance values $\hat{I}^{(l)}[m,n]$ may be error-free or affected by reconstruction errors, resulting from losses of coded data pertaining to the l-th frame or from propagation of errors occurred on preceding frames due to the employment of predictive coding. Moreover, the amount of error varies from a pixel to another. In fact, at the output of the decoding and de-interleaving stage, in $\hat{I}^{(l)}[m,n]$ we can distinguish

- error-free pixels,
- pixels that have been concealed using at least one correctly received description pertaining to the same frame
- pixels that have been concealed in absence of alternative descriptions pertaining to the same frame.

Hence, we recognize that the MDC concealment induces a fairly natural reliability hierarchy in the luminance values of the pixels of the sequence $\hat{I}^{(l)}[m,n]$. We formalize this hierarchy by introducing three classes of pixels, namely Class I (error-free), Class II (MD concealed), and Class III (SD concealed) and assigning a different reliability $r^{(l)}[m,n]$ to pixels belonging to different classes. Then, the decoding sequence quality can be improved by applying a restoration algorithm that takes into account not only the local image edges but also the pixel reliability. The restoration stage operates by replacing each concealed (Class II or Class III) pixel in $\hat{I}^{(l)}[m,n]$ with a suitably interpolated estimate.

The herein presented interpolation technique extends the classical edge-detection interpolation scheme Edge-based Line Average (ELA) into a Robust ELA (RELA) to exploits the reliability of the descriptions available for interpolation. Namely, for a given site (m,n), we define a 3×3 neighborhood $\eta(m,n)$ as illustrated in Fig.2Then, for each site (m,n), four pairs of pixels belonging to $\eta(m,n)$ are individuated as: $\{(m+\delta_v,n+\delta_h),(m-\delta_v,n-\delta_h),\delta_v,\delta_h\in\mathscr{S}_{\mathrm{ELA}}\}$ being $\mathscr{S}_{\mathrm{ELA}}$ defined as: $\mathscr{S}_{\mathrm{ELA}}\stackrel{\mathrm{def}}{=}\{(1,0),(0,1),(1,1),(1,-1)\}$. Each pair of pixels, indexed by δ_v,δ_h, identifies a candi-

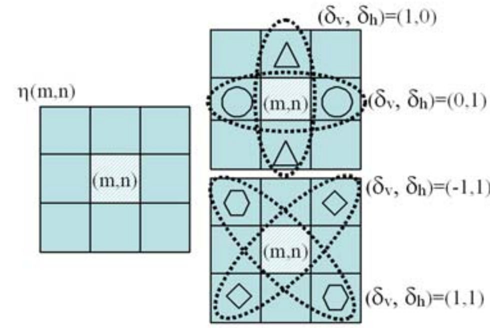

Figure 2: The considered neighborhood $\eta(m,n)$ and the associated four pixel pairs, indexed by $(\delta_v,\delta_h)\in\mathscr{S}_{\mathrm{ELA}}$.

date direction for interpolation; ELA searches for the direction of minimal luminance variation, *i.e.* for the pair (δ_v,δ_h) minimizing $\left|\hat{I}^{(l)}[m+\delta_v,n+\delta_h]-\hat{I}^{(l)}[m-\delta_v,n-\delta_h]\right|$ and estimates the luminance in $\hat{I}^{(l)}_{\mathrm{ELA}}[m,n]$ as the average between $\hat{I}^{(l)}[m+\delta_v,n+\delta_h]$ and $\hat{I}^{(l)}[m-\delta_v,n-\delta_h]$. Since it has been originally designed for fast upsampling of high quality images, the ELA interpolation algorithm does not take into account possible errors affecting the luminance of the pixels in $\eta(m,n)$; thus, it performs quite well on reconstructing missing pixels for error free descriptions, but it presents modest performances when the descriptions are affected by residual errors after concealment.

We design here a robust edge driven interpolation algorithm (RELA) taking into account the measurements reliability. Specifically, let us assume that a reliability measure $r^{(l)}[m,n]$ is associated to each pixel $[m,n]$ of the l-th frame after the concealment stage. This measure is used to operate a reduction of the set of directions which are candidate for interpolation, by limiting to the set of most reliable direction, *i.e.* the set $\mathscr{S}_{\mathrm{RELA}}$ as in (4), being θ a suitably defined threshold. Then, the optimal interpolation direction is determined as in (5), and estimates the luminance in (m,n) as in (6). From (6), we recognize that the robust interpolation attempts to restore the concealed pixels by directional smoothing, meanwhile using only the most reliable luminance values. Finally, we observe that the described interpolation strategy estimates the luminance value at the location (m,n) employing directional interpolation of pixels belonging to $\eta(m,n)$. The reformulation of this interpolation in terms of Bayesian interpolation of Markov Random Fields is currently under investigation.

5. NUMERICAL RESULTS

In this section we present a set of numerical simulation results assessing the performance of PSS MDC technique using the robust RELA interpolation. The experiments refer to the

$$\mathscr{S}_{\text{RELA}} = \left\{ (\delta_v, \delta_h) \in \mathscr{S}_{\text{ELA}}, (\delta_v, \delta_h) \text{ s.t.} \left| r^{(l)}[m+\delta_v, n+\delta_h] \right| + \left| r^{(l)}[m-\delta_v, n-\delta_h] \right| > \theta \right\} \tag{4}$$

$$(\delta_v^{(\text{RELA})}, \delta_h^{(\text{RELA})}) = \underset{(\delta_v,\delta_h)\in\mathscr{S}_{\text{RELA}}}{\arg\min} \left| \hat{I}^{(l)}[m+\delta_v, n+\delta_h] - \hat{I}^{(l)}[m-\delta_v, n-\delta_h] \right| \tag{5}$$

$$\hat{I}^{(l)}_{\text{RELA}}[m,n] = \frac{\hat{I}^{(l)}[m+\delta_v^{(\text{RELA})}, n+\delta_h^{(\text{RELA})}] + \hat{I}^{(l)}[m-\delta_v^{(\text{RELA})}, n-\delta_h^{(\text{RELA})}]}{2} \tag{6}$$

test sequences *Foreman* and *News*, CIF format, at 10 frames per second.. A number of $K = 2 \times 2$ descriptions has been selected. The RELA was realized by assigning $r^{(l)}[m,n] = 2$ to Class I pixels, $r^{(l)}[m,n] = 1$ to Class II pixels, $r^{(l)}[m,n] = 0$ to Class III pixels, and by setting $\theta = 1$.

The interleaved sequences are encoded using the reference JM H.264 coder version 11.0 [12], one NALU per frame. The GOP structure is given by a primary SP frame followed by 9 P frames; in each frame, 40 macroblocks are INTRA encoded for Random Intra Refresh purposes. The H.264 Video Coding Layer encodes one slice per frame, and the Network Adaptation Layer followed by the RTP packetizer using the so called simple packetization method maps each slice into an RTP packet.

The first set of numerical simulations refers to the encoding of 100 frames of the sequence News, at a bit-rate of 600 kbps. We analyze here in detail a run characterized by PLP=13%. MDC using RELA reduces the visually relevant artifacts that are observed on the decoded video sequence in presence of transmission errors. Fig.3 shows selected details of a few snapshots captured within the sequence decoded using ELA and RELA; the visual quality improvement achieved by adopting MDC with RELA is clearly appreciated. The quality of the video sequences decoded in different conditions has been also evaluated in terms of Peak to Signal Noise Ratio (PSNR), defined as: $\text{PSNR} \overset{\text{def}}{=} 255^2/\text{MSE}$, proving that the RELA stage significantly improves the overall MDC performance, resulting into a PSNR gain of 1.5 dB over MDC without interpolation and 1.1 dB over ELA.

The second set of numerical simulations refer to the encoding of 100 frames of the sequence Foreman, at a bit-rate of 750 kbps. A transport channel characterized by a packet loss probability (PLP) equal to 10% has been simulated over 100 Montecarlo runs. The MDC scheme using RELA, ELA and MDC without interpolation have been compared, by evaluating the decoded sequence PSNR values observed on each of the 100 frames, and by characterizing statistically the PSNR values observed using the different schemes. Fig. 4 reports the PSNR histograms of the three schemes, while Table 1 reports selected parameters characterizing the PSNR distributions.

6. CONCLUSION

In this paper, we have analyzed a PSS MDC scheme. The underlying assumption of general MDC schemes is that each description is a tight approximation of the others; in the particular case of PSS MDC, this is the consequence of the correlation between the neighboring pixels of a natural image. In case of losses, the availability of multiple descriptions is exploited to perform a more accurate error concealment. The concealment induces a fairly natural pixel hierarchy that can be exploited by a post-processing stage. Here, we analyze a fast restoration stage that makes use not only of the local directionality information but also of the reliability of the decoded pixes. The scheme effectively improves the decoded video sequence quality on lossy channels both from an objective and from a subjective point of view. The interesting performance of the interpolation stage are related to the markovian nature of natural images; the relation between the interpolation algorithm and markovian image interpolation [13] is currently under investigation. The herein presented MDC scheme is realized in the form of pre-processing and post-processing stage in H.264/AVC standard compliant encoder and decoder pair; besides, transmission diversity is straightforwardly obtained by mapping one application packet in at least one transport packet, while maintaining a single transport data flow.

REFERENCES

[1] V. Goyal, Multiple description coding: compression meets the network, *IEEE Sig. Proc. Mag.*, vol. 18, pp. 74-93, Sept. 2001.

[2] Y. Wang, A.R. Reibman, and S. Lin, Multiple description coding for video delivery, *Proc. of the IEEE*, vol. 93, no. 1, pp. 57-70, Jan. 2005.

[3] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, Multiple description video coding for scalable and robust transmission Over IP, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp.321-334, March 2005.

[4] T. Tillo, and G. Olmo, Data-dependent pre- and post-processing multiple description coding of images, *IEEE Trans. on Image Processing*, vol. 16, no. 5, pp. 1269-1280, May 2007.

[5] R. Bernardini, M. Durigon, R. Rinaldo, L. Celetto, A. Vitali, Polyphase spatial subsampling multiple description coding of video streams with H264, *Proc. ICIP 2004*, Singapore, October 24-27. 2004, pp. 3213-3216.

[6] A. Vitali, Multiple description coding: a new technology for video streaming over the Internet, *EBU Technical Review*, Bergamo, October 2007.

[7] J. Kim, R.M. Mersereau, and Y. Altunbasak, Distributed video streaming using multiple description coding and unequal error protection, *IEEE Trans. on Image Processing*, vol. 4, no. 7, pp. 849-861, July 2005.

[8] C. Su, J.J. Yao, and H.H. Chen, H.264/AVC-Based multiple description coding scheme, *Proc. ICIP 2007*, San Antonio, September 16-19. 2007, pp. 265-268.

[9] T. Wiegand *et all.*, "Overview of the H.264 video coding standard", *IEEE Trans. on Circ. and Sys. for Video Tech.*, vol.13, no.7, pp.560-576, July 2003.

[10] T. Tillo, M. Grangetto, and G. Olmo, Redundant slice optimal allocation for H.264 multiple description coding, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp.59-70, Jan. 2008.

[11] S. Wenger, H.264/AVC over IP, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, pp. 645-656, July 2003.

[12] H.264/AVC reference software, `http://iphome.hhi.de/suehring/tml/`

[13] M. Li, and T.Q. Nguyen, Markov random field Model-Based edge-directed image interpolation, *IEEE Trans. on Image Processing*, vol. 17, no. 7, pp. 1121-1128, July 2008.

Figure 3: Details for frame 46 of the sequences decoded using R-ELA and ELA interpolation, News sequence, CIF, 600 kbps.

PSNR	MDC using RELA	MDC using ELA	MDC without interpolation
Mean	30.59	29.41	28.14
Standard Deviation	3.42	4.03	4.56
Median	31.03	29.41	27.63

Table 1: PSNR of the MD encoded sequences, decoded using R-ELA and ELA interpolation, and without interpolation: Foreman sequence, CIF, 750 kbps, 100 encoded frames, 50 Montecarlo runs.

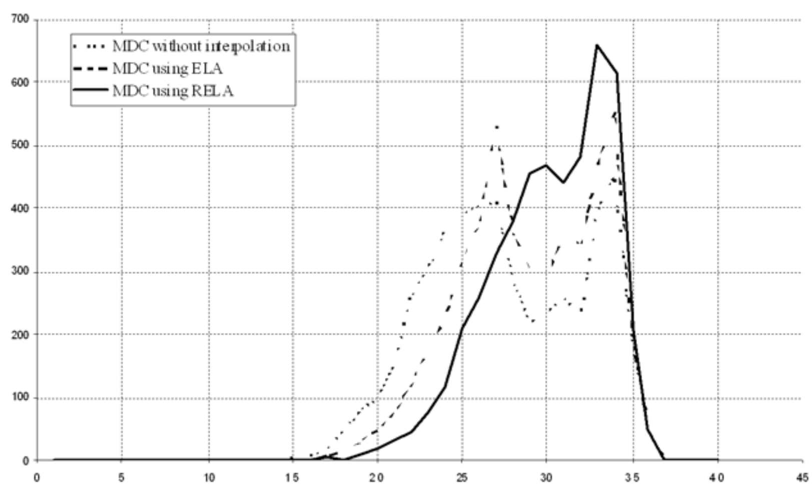

Figure 4: PSNR Histograms for the MD encoded sequences, decoded using R-ELA and ELA interpolation, and without interpolation: Foreman sequence, CIF, 750 kbps, 100 encoded frames, 50 Montecarlo runs.
.