# Quality-Controlled Motion-Compensated Interpolation

Mina Makar, Derek Pang, Yao-Chung Lin and Bernd Girod
Information Systems Laboratory, Department of Electrical Engineering, Stanford University
{mamakar, dcypang, yclin79, bgirod}@stanford.edu

***Abstract*—Low-complexity video encoding is important for mobile and power-sensitive applications. One way to reduce encoding complexity is to drop frames at the encoder and perform motion-compensated frame interpolation at the decoder. We propose a method to estimate the quality of the interpolated frames at the decoder by transmitting a small amount of error control information in lieu of an omitted frame. Typically, this information is obtained from a projection of the frame on a suitable low-dimensional basis. The projection coefficients can be compressed by conventional techniques or, more efficiently, by Slepian-Wolf coding. Based on the quality estimate, the decoder can recognize and suppress occasional frames for which motion-compensated interpolation does not yield satisfactory picture quality. Experimental results demonstrate that our approach eliminates most interpolation artifacts and achieves much better visual quality at a negligible increase in bit-rate.**

## I. Introduction

Video content delivery over mobile networks is expected to undergo a huge increase in the coming years as higher bandwidth will become available and lower latency will be achieved with the introduction of 4G/LTE networks [1]. This increase in demand requires the development of low-complexity video coding techniques that better suit mobile and power-sensitive applications.

One way to reduce encoding complexity is to drop frames at the encoder and perform motion-compensated (MC) frame interpolation at the decoder. MC interpolation often results in spatial artifacts especially in video frames with high motion. In the case of low-quality interpolated frames, it is usually better for the decoder to suppress these frames and lower the frame rate of the video instead of displaying visible artifacts.

In this paper, we propose a method to estimate the quality of the interpolated frames at the decoder by transmitting a small amount of error control information in lieu of an omitted frame. In the case when the decoder detects a high-quality frame, it displays the current interpolated frame, and thus keeps the frame rate of the original video. On the other hand, in the case when the decoder detects a low-quality frame, it displays a copy of the nearest high-quality frame which causes a temporary drop in the output frame rate.

Prior efforts [2]–[6] explored the idea of transmitting a projection of the original frame for performing image authentication [2], [3], estimating PSNR [4], [5] and evaluating the effectiveness of error concealment [6]. Based on the observation that a similar projection can be generated at the decoder from a compressed and/or error-prone video frame, the authors use distributed source coding (DSC) [7] techniques to efficiently compress the projection data. Our work extends this DSC-based approach to the problem of quality-controlled motion-compensated interpolation.

The remainder of the paper is organized as follows. Section II discusses the proposed technique in details and explains the reasons behind our specific choice of the projection data. In Section III, we discuss the efficient compression of the projection data using DSC. Finally, in Section IV, we present experimental results showing the performance of the proposed low-quality frame detection and frame replacement methods. We also explain the developed subjective test used to evaluate the impact of our quality control technique.

## II. Quality Estimation Technique

We consider the video transmission system in Fig. 1 with a low-complexity encoder and a decoder that is not complexity-constrained. The transmitter selects original frames at a fixed interval $K$ and drops all $K-1$ frames between each two selected frames. We refer to the selected frames as *Transmitted frames*. Example with $K=3$ is shown in Fig. 1. The sequence of transmitted frames is encoded using a suitable video coding standard such as H.264 [8]. The encoded sequence is transmitted to the receiver, where it is decoded. The receiver then performs MC interpolation on the decoded transmitted frames to generate an interpolated version of each omitted frame.

Depending on the interpolation method and the number of omitted frames, the quality of the interpolated frames varies. MC interpolation often generates visible artifacts. The receiver estimates the quality of the interpolated frames and suppresses the low-quality ones. To estimate the quality, we allow the transmitter to send a small amount of extra information for each omitted frame. The generation and comparison of this information is performed as follows.

### A. Generation and Comparison of Projection Data

As shown in Fig. 1, we refer to the original omitted video frame as $x$ and the corresponding interpolated frame as $y$. The transmitter projects each omitted frame on a low-dimensional basis to generate the control information that is used to estimate the quality of MC interpolation. The omitted frame is divided into $N \times N$ blocks. The mean of each block is calculated and the projection data comprises the set of the mean values of all the blocks in the omitted frame. We refer to the projection data of the original video frames as $P_x$.

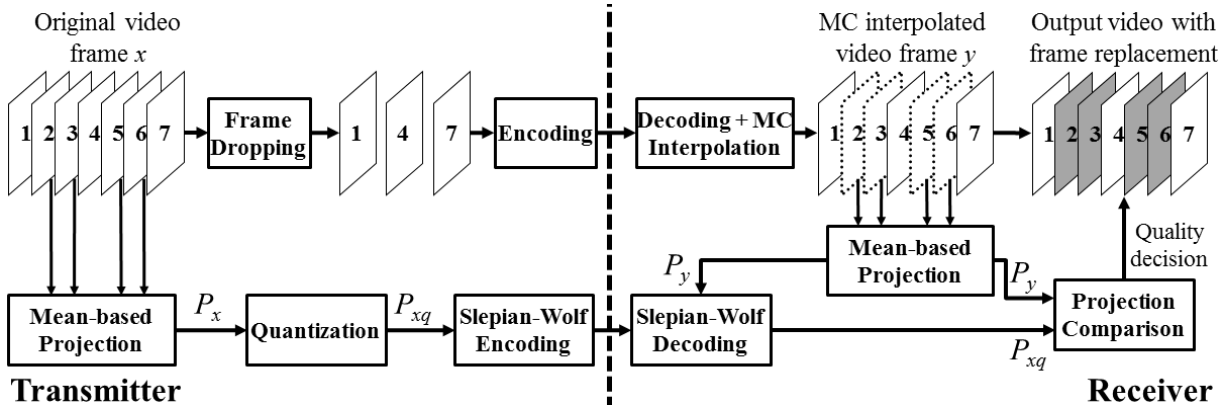


Fig. 1. Block diagram of the proposed quality control technique. On the transmitter side, a projection is calculated for each omitted frame, quantized and encoded with Slepian-Wolf coder. On the receiver side, this projection is compared to the projection of the MC interpolated frame to evaluate its quality. In case of detecting low-quality, the interpolated frame is replaced with the nearest high-quality frame. Dotted lines indicate interpolated frames while the final displayed frames are shown in gray.

At the receiver side, the same mean-based projection is calculated for the interpolated frames. We refer to the projection data of the interpolated frames as $P_y$. Each value in $P_x$ (or a quantized version of it) is compared to the co-located value in $P_y$. If the absolute difference between the two values exceeds a certain threshold $T_{\text{Block}}$, the corresponding $N \times N$ block is considered a low-quality block. Moreover, if the number of low-quality blocks in an interpolated frame exceeds another threshold $T_{\text{Frame}}$, the whole interpolated frame is considered a low-quality frame since the artifacts due to the low-quality blocks are becoming noticeable to the end user.

When the receiver decides that a certain interpolated frame is of low quality, this frame is not displayed. Instead, this frame is replaced with the nearest high-quality frame which is either a decoded transmitted-frame or another interpolated frame. For example, consider the first two transmitted frames which are frame 1 and frame $K+1$. The decision is made in alternating order for frame 2, $K$, 3, $K-1$, 4, ... . In the case of low quality, frame 2 is replaced by a copy of frame 1 and frame $K$ by a copy of frame $K+1$. If frame 3 is low-quality, it is replaced by whatever the receiver decides for frame 2, which is either a copy of frame 1 or a high-quality interpolated frame 2 and so on. Performing the replacement operation in this order ensures the use of the nearest high-quality frame in place of each low-quality interpolated frame and only causes a temporary drop in the effective frame rate of the output video.

An important advantage of the proposed technique is that it does not depend on the video coding standard or the MC interpolation method used. It can be performed as a post-processing step after decoding and interpolating the received video, which allows easy integration with a wide variety of video transmission systems.

### *B. Mean-Based Projection*

The method of projecting each omitted frame on a low-dimensional basis has a great impact on the performance of the proposed technique. The projection process should be performed at very low complexity, and the projection of the original omitted frames $P_x$ should show large distinction between low-quality and high-quality interpolation cases when it is compared with the projection of the interpolated frames $P_y$. We consider a block-wise projection for each $N \times N$ block in the original and the interpolated frames. A block-wise projection allows the projection data to capture the localized interpolation errors and facilitates a low-complexity and parallel implementation of the projection process.

To select a suitable block-wise projection for our application, we analyze the power spectral density (PSD) of the MC interpolation error in the case of low-quality versus high-quality interpolation. This is achieved through the following experiment: We drop frames from CIF size *Foreman* video sequence at $K = 4$. The transmitted frames are encoded at high quality to avoid compression artifacts. The encoded bitstream is decoded and MC interpolation is performed to generate three interpolated frames between each two transmitted frames.

We divide each interpolated frame into $N \times N$ blocks where $N = 32$ in this experiment. We calculate the PSNR of each interpolated block by comparing it to the co-located $N \times N$ block in the original omitted frame. We select the 10000 blocks with the lowest PSNR and consider them low-quality blocks. We also select the 10000 blocks with the highest PSNR and consider them high-quality blocks.

The total MC interpolation error $z$ of the low-quality and the high-quality blocks is calculated by subtracting the interpolated blocks from the corresponding original blocks. Thus, we describe the MC interpolation quality estimation problem in a hypothesis testing formulation as follows,

$$z = \begin{cases} z_L & \text{Low quality} \\ z_H & \text{High quality} \end{cases} \tag{1}$$

We further calculate the autocorrelation function of the MC interpolation error,

$$R_{zz}(k,l) = E\left[z(m,n)\, z(m-k, n-l)\right] \tag{2}$$

and its power spectral density,

$$\Phi_{zz}(\omega_1, \omega_2) = \mathcal{F}\left\{R_{zz}(k,l)\right\} \tag{3}$$

The PSD of the low-quality interpolation error $\Phi_{z_L z_L}(\omega_1, \omega_2)$ is calculated by averaging the PSDs of the interpolation error of 10000 low-quality blocks. Similarly, the PSD of the high-quality interpolation error $\Phi_{z_H z_H}(\omega_1, \omega_2)$ is calculated by averaging the PSDs of the interpolation error of 10000 high-quality blocks. Assuming that the interpolation error follows stationary Gaussian statistics, the ratio between $\Phi_{z_L z_L}(\omega_1, \omega_2)$ and $\Phi_{z_H z_H}(\omega_1, \omega_2)$ at a certain frequency component can be viewed as the ratio between the variances of two Gaussian sources. A higher ratio between the variances allows us to easily decide which of the two sources is generating a certain realization of the interpolation error.

Fig. 2 plots the ratio between the PSDs of the low-quality and the high-quality MC interpolation errors. Since this plot is peaky at low frequencies, this suggests that low frequency components can greatly distinguish between low-quality and high-quality interpolated blocks while high frequency components offer less discrimination. Thus, a mean-based projection that extracts low-frequency information is well suited for the proposed quality control technique. Moreover, the choice of the mean-based projection satisfies all the aforementioned criteria.

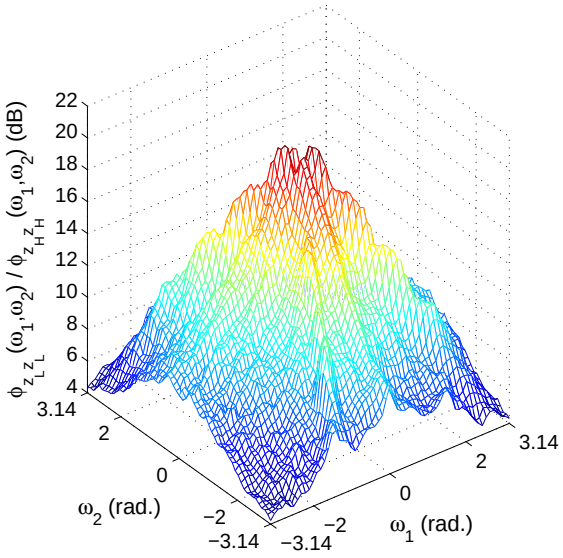


Fig. 2. Ratio between the PSD of the low-quality interpolation error $\Phi_{z_L z_L}(\omega_1, \omega_2)$ and the PSD of the high-quality interpolation error $\Phi_{z_H z_H}(\omega_1, \omega_2)$. $\Phi_{z_L z_L}(\omega_1, \omega_2)$ and $\Phi_{z_H z_H}(\omega_1, \omega_2)$ are calculated by averaging the PSDs of the interpolation error over 10000 low-quality and high-quality blocks respectively. The figure shows that low frequency components offer more distinction between low-quality and high-quality interpolation.

## III. Compression of Projection Data

It is desirable that sending the projection data does not impose a noticeable increase in the transmission bit-rate, thus, the compression of these data is vital. An optional quantization operation can be performed to reduce the number of bits used to represent values in $P_x$. We refer to the quantized version of $P_x$ as $P_{xq}$. The determination of the number of bits sufficient for representing $P_{xq}$ without harming the quality control operation depends on $T_{\text{Block}}$, which is the threshold used to decide whether a certain block is low-quality or high-quality. Larger value of $T_{\text{Block}}$ allows us to coarsely quantize $P_{xq}$ since we can tolerate more quantization error and the quality decision is not much affected.

We exploit the fact that in the case of high-quality MC interpolation, there is high correlation between $P_{xq}$ and $P_y$. Since the projection $P_y$ is already available at the decoder, we can use DSC to efficiently encode $P_{xq}$ at a negligible rate and then use $P_y$ as side-information to decode the Slepian-Wolf coded $P_{xq}$. Each bitplane in $P_{xq}$ is coded at the Slepian-Wolf encoder using *LDPC Accumulate* codes [9] where the choice of the coding bit-rate depends on the worst allowable high-quality interpolated frames. If the encoded projection is correctly decoded, a comparison between $P_y$ and the decoded $P_{xq}$ is performed to decide the interpolation quality. However, if the projection is not decodable, this means that there is a small correlation between $P_{xq}$ and $P_y$ and thus, the quality of the interpolated frame is low. The block diagram in Fig. 1 shows how DSC is incorporated into the whole system.

## IV. Experimental Results

We perform experiments with three CIF size standard video sequences, *Mother&Daughter*, *Foreman* and *Football*. These videos correspond to low, medium and high motion examples respectively, so we test the performance of the proposed technique in different motion conditions. We use 'x264 software' [10] for encoding the transmitted frames and 'YUV Frame Rate Conversion tool' [11] for performing MC interpolation of the decoded videos.

### A. Detection of Low-quality Interpolated Frames

A small value of the block size $N$ increases the required bit-rate for the projection data significantly, while a large value provides less sensitivity to local interpolation artifacts. We find that $N = 16$ or 32 provide the best trade-off for detecting artifacts at a reasonable bit-rate. For $N = 16$, we use larger values for $T_{\text{Block}}$ and $T_{\text{Frame}}$ ($T_{\text{Block}} = 20$ and $T_{\text{Frame}} = 5$) and quantize $P_{xq}$ to 5 bits (instead of the original 8 bits resolution). For $N = 32$, we use smaller thresholds ($T_{\text{Block}} = 8$ and $T_{\text{Frame}} = 4$) and quantize $P_{xq}$ to 6 bits for preserving the sensitivity to artifacts. The choice of the quantization coarseness is based on $T_{\text{Block}}$ and ensures that the final decisions of low-quality interpolated frames do not change much compared to the case of sending the unquantized projection $P_x$.

Fig. 3 shows the percentage of detected low-quality interpolated frames compared to the total number of interpolated frames for the case of $N = 16$ and 32 and $K = 2$, 3 and 4 at different quantization parameters (QPs). As $K$ increases, the number of omitted frames increases and the quality of MC interpolation degrades and thus, more frames are decided to be low-quality. Very small percentage of low-quality frames is detected for *Mother&Daughter* because of low motion. However, for *Football*, most of the omitted frames are detected

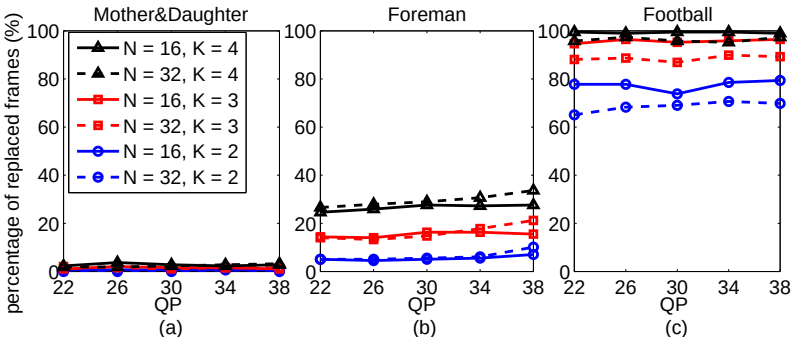


Fig. 3. Percentage of the detected low-quality frames compared to the total number of interpolated frames (a) *Mother&Daughter*, (b) *Foreman* and (c) *Football*.

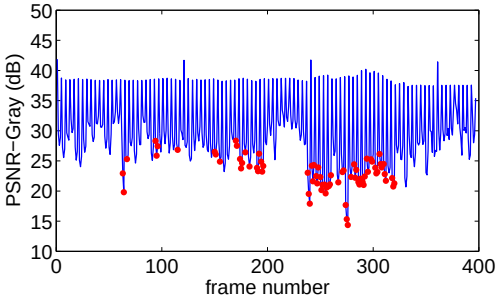


Fig. 4. PSNR trace for interpolated *Foreman*. Red dots represent the detected low-quality frames.

to be low-quality due to very high motion. The results are consistent for different QPs, which suggests the robustness of our technique against compression artifacts. In Fig. 4, we present an example PSNR trace for the interpolated *Foreman* sequence, with QP $= 22$, $N = 16$ and $K = 4$. The frames detected by our technique are marked with red dots and are characterized to have drastic PSNR drop.

Table I shows how efficiently we can compress the projection data using DSC. Values in the table represent the minimum decodable rate for each video sequence. As $K$ increases, the video encoding load and bit-rate decreases. However, the bit-rate for sending the projection data increases because of more omitted frames. The case of $N = 16$ requires nearly 3 times the bit-rate as compared to $N = 32$ and provides more sensitivity to local artifacts. In Table I, we also compare the bit-rate needed for DSC compression to the case of using a fixed length (F. L.) code for sending the projection data (codeword length = 5 bits for $N = 16$ and 6 bits for $N = 32$). Note that the original frame rate of *Foreman* sequence is 25 fps, while the frame rate of the other sequences is 30 fps. Hence, *Foreman* sequence requires less bit-rate in case of using a fixed length code. We find that DSC results in compression ratio of about $60\%$ for $N = 16$ and $50\%$ for $N = 32$. The bit-rate for sending the projection data using DSC is usually negligible when compared to the bit-rate used to encode the transmitted frames at a reasonable quality.

### *B. Subjective Evaluation of Frame Replacement*

PSNR is not a suitable metric to capture the subjective quality improvement that our frame replacement strategy provides. When we compare a low-quality interpolated frame to its corresponding replacing frame, we find that pixel values in the interpolated frame are usually closer to the correct pixel values of the omitted frame due to averaging of two or more transmitted frames. However, the visual quality of the replacing frame is much better due to the absence of interpolation artifacts. We fix $N = 16$ and provide more experiments to demonstrate the improved quality of the proposed replacement strategy. First, we present some visual quality results. In Figs. 5 and 6, example low-quality interpolated frames are presented and the low-quality blocks are highlighted in red. The figures indicate that the replacing frames provide much better visual quality than the low-quality interpolated frames.

We design a subjective test for performance evaluation. The original video is encoded after omitting frames with different QPs and different values of $K$. After decoding, we generate two output videos. The first is the result of MC interpolation and the second is the result of applying the proposed quality control method. Each pair of test videos are displayed side-by-side, on a 21-inch monitor at a distance of $60 - 70$ cm. from the viewer (6 times the picture height). The position of the video processed with the proposed method versus the MC interpolated video is randomized.

The test proceeds as follows. For each value of QP and $K$, the subject is asked to give his/her opinion score according to Table II. The scores are then mapped, so that positive values indicate that users prefer the video processed with the proposed method, while negative values indicate that users prefer the videos with MC interpolated frames. In Fig. 7, mean opinion score (MOS) statistics (averaged over 11 subjects) are used to describe the performance of the proposed technique for *Foreman* and *Football* sequences. For $K = 2$, the subjects usually assign similar quality to both videos due the efficiency of the interpolation process. For higher values of $K$, subjects tend to much prefer videos generated with our quality control mechanism in spite of having slight temporal artifacts due to frame repeating.

## V. Conclusions

We present a reduced reference quality metric for measuring the quality of frames omitted at a video transmitter and interpolated at the video receiver for lowering the encoding complexity. This is achieved by transmitting a small amount of error control information in lieu of an omitted frame. Typically, this information is obtained from a low-dimensional mean-based projection of the frame. The projection coefficients can be efficiently compressed by Slepian-Wolf coding. We also propose a simple frame replacement strategy to account for the low-quality interpolated frames.

Experimental results prove the efficiency of the proposed technique in terms of lowering the complexity of the video encoder, detecting the correct low-quality interpolated frames and significantly improving the visual quality of the resulting video at a negligible increase in bit-rate.

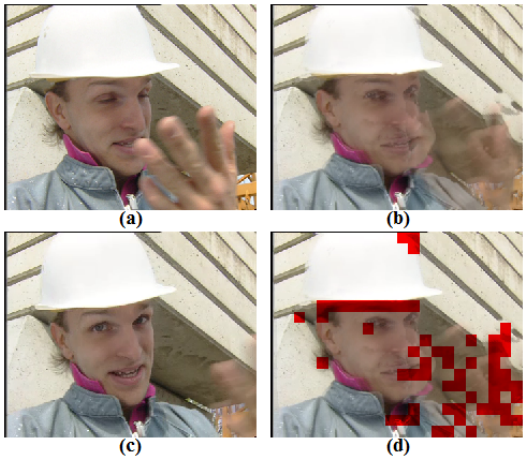

Fig. 5. Low-quality interpolated frame detected using the proposed method. *Foreman* sequence, frame 255, QP = 22, $N = 16$ and $K = 4$. (a) original frame, (b) interpolated frame, (c) frame used to replace low-quality interpolated frame and (d) interpolated frame with low-quality blocks highlighted in red.

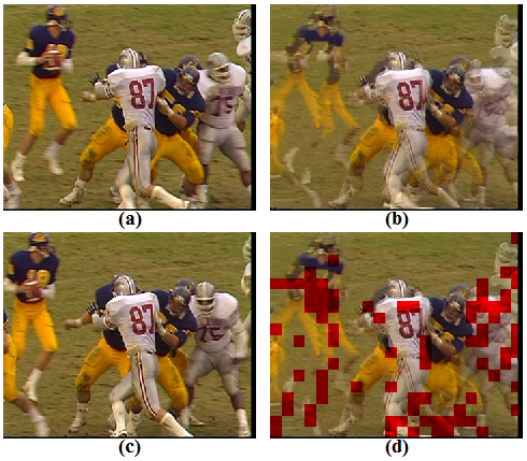

Fig. 6. Low-quality interpolated frame detected using the proposed method. *Football* sequence, frame 63, QP = 22, $N = 16$ and $K = 4$. (a) original frame, (b) interpolated frame, (c) frame used to replace low-quality interpolated frame and (d) interpolated frame with low-quality blocks highlighted in red.

TABLE I
BIT-RATE FOR SENDING PROJECTION DATA.

| Sequence | Bit-rate for projection data (kbps) | | | | | |
|---|---|---|---|---|---|---|
| | $N = 16$ | | | $N = 32$ | | |
| | $K = 2$ | $K = 3$ | $K = 4$ | $K = 2$ | $K = 3$ | $K = 4$ |
| *Mother& Daughter* | | | | | | |
| DSC | 11.25 | 15.00 | 16.88 | 4.59 | 6.12 | 6.89 |
| F. L. | 29.70 | 39.60 | 44.55 | 8.91 | 11.88 | 13.37 |
| *Foreman* | | | | | | |
| DSC | 10.88 | 14.50 | 16.31 | 3.60 | 4.80 | 5.40 |
| F. L. | 24.75 | 33.00 | 37.13 | 7.43 | 9.90 | 11.14 |
| *Football* | | | | | | |
| DSC | 12.15 | 16.20 | 18.23 | 4.32 | 5.76 | 6.48 |
| F. L. | 29.70 | 39.60 | 44.55 | 8.91 | 11.88 | 13.37 |

TABLE II
INTERPRETATION OF SUBJECTIVE TEST SCORES.

| Score | $-2$ | $-1$ | 0 | 1 | 2 |
|---|---|---|---|---|---|
| Interpretation | Left much better | Left better | Same quality | Right better | Right much better |

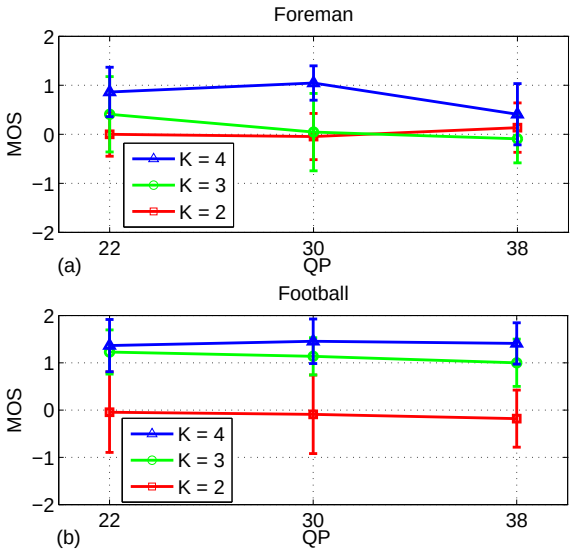


Fig. 7. MOS results for subjective test. Error bars represent the standard deviation of MOS values. Positive values indicate that users prefer the video processed with the proposed method, while negative values indicate that users prefer the videos with MC interpolated frames. (a) *Foreman* sequence and (b) *Football* sequence.

## REFERENCES


[1] "Cisco Visual Networking Index: Forecast and Methodology, 2008-2013," June 2009.

[2] Y.-C. Lin, D. Varodayan, and B. Girod, "Image Authentication Based on Distributed Source Coding," *Proc. IEEE International Conference on Image Processing*, vol. 3, 2007, pp. 5-8, San Antonio, TX, Sep. 2007.

[3] —, "Image Authentication and Tampering Localization Using Distributed Source Coding," *Proc. IEEE Multimedia Signal Processing Workshop*, Crete, Greece, Oct. 2007.

[4] K. Chono, Y.-C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, "Reduced-Reference Image Quality Estimation Using Distributed Source Coding," *Proc. International Conference on Multimedia and Expo*, Hannover, Germany, June 2008.

[5] Y.-C. Lin, D. Varodayan, and B. Girod, "Video Quality Monitoring for Mobile Multicast Peers Using Distributed Source Coding," *Proc. 5th International Mobile Multimedia Communications Conference*, London, U.K., Sep. 2009.

[6] Z. Li, Y.-C. Lin, D. Varodayan, P. Baccichet, and B. Girod, "Distortion-Aware Retransmission and Concealment of Video Packets Using a Wyner-Ziv-Coded Thumbnail," *Proc. IEEE Multimedia Signal Processing Workshop*, Cairns, Australia, Oct. 2008.

[7] D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471-480, July 1973.

[8] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol.13, no.7, pp. 560-576, July 2003.

[9] D. Varodayan, A. Aaron, and B. Girod, "Rate-Adaptive Codes for Distributed Source Coding," *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, vol. 86, no. 11, pp. 3123-3130, Nov. 2006.

[10] "x264 encoder."
Website: http://www.videolan.org/developers/x264.html.

[11] "YUV Frame Rate Conversion tool."
Website: http://www.yuvsoft.com/technologies/frame_rate/index.html.