LOW BIT-RATE VIDEO CODING VIA MODE-DEPENDENT ADAPTIVE REGRESSION FOR WIRELESS VISUAL COMMUNICATIONS

Xianming Liu[1], Xiaolin Wu[2], Xinwei Gao[1], Debin Zhao[1], Wen Gao[3]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[2]ECE Department, McMaster University, Ontario, Canada, L8S 4K1
[3]School of Electronic Engineering & Computer Science, Peking University, Beijing, China

ABSTRACT

In this paper, a practical video coding scheme is developed to realize state-of-the-art video coding efficiency with lower encoder complexity at low bit-rate, while supporting standard compliance and error resilience. Such an architecture is particularly attractive for wireless visual communications. At the encoder, multiple descriptions of a video sequence are generated in the spatio-temporal domain by temporal multiplexing and spatial adaptive downsampling. The resulting side descriptions are interleaved with each other in temporal domain, and still with conventional square sample grids in spatial domain. As such, each side description can be compressed without any change to existing video coding standards. At the decoder, each side description is first decompressed, and then reconstructed to original resolution with the help of the other side description. In this procedure, the decoder recover the original video sequence in a constrained least squares regression process, using 2D or 3D piecewise autoregressive model according to different prediction modes. In this way, the spatial and temporal correlation is sufficiently explored to achieve superior quality. Experiment results demonstrate the proposed video coding scheme outperforms H.264 in rate-distortion performance at low bit-rates and achieves superior visual quality at medium bit-rates as well.

Index Terms— Low bit-rates, mode dependent, adaptive regression, wireless visual communications

1. INTRODUCTION

In recent years, the availability of inexpensive hardware such as CMOS cameras that are able to ubiquitously capture visual content from the environment has fostered the development of Wireless Visual Sensor Networks (WVSNs). It has developed as a new technology with many potential applications, ranging from mobile multimedia to security monitoring.

Unlike PCs or the Internet, which are designed to support all types of applications, WVSNs are usually mission driven and application specific. The explosive growth of diverse visual sensor applications is calling for video coding technology capable of dealing with the associated harsh operating conditions. These contain advances in understanding of resource-deficient wireless communications including constrains on energy and bandwidth, the integration of efficient video coding techniques, and the development of advanced error resilience technique to allow a reliable transmission.

There exists a vast literature on video coding techniques. Traditional video compression standards, such as H.264/AVC, are based on the idea of prediction coding to exploit spatio-temporal correlations. Although achieving the state-of-the-art rate-distortion performance, it may not be suited for low-cost multimedia sensors since predictive coding requires complex encoders and entails high energy consumption. Besides, the prediction coding system also causes inter-frame dependency in decompression, which results in error propagation for an error-prone channel. Another approach in conventional wisdom is distributed video coding (DVC) [1]. Within the framework, the traditional balance of complex encoder and simple decoder can be reversed. Moreover, DVC has an inbuilt robustness to channel losses because there is a duality between distributed source coding and channel coding. Clearly, such algorithms are very promising for WVSNs. However, despite all the research efforts dedicated to DVC, the current systems still fail to meet the compression efficiency of their predictive coding counterparts.

To meet all the strict requirements, new video coding paradigms for wireless visual communications need to be considered. Wisdoms on the emerging compressive sensing (CS) theory [2] gives us some inspiration. On one hand, it shows that it is possible, at least theoretically, to obtain compact signal representation by a greatly reduced number of random samples. Several successful algorithms in image coding demonstrate the validity of this theory [4]. On the other hand, the fact that most natural video frames have strong spatio-temporal correlations suggests the possibility of interpolation-based compact representation of video signals. In this way, video source needed to be compressed and transmitted is significantly reduced, so that it can efficiently ease the burden of the encoder and wireless transmission channel, and extend the life of wireless sensors.

With the above motivation, in this paper, we propose a standard compliance low bit-rate video coding scheme with relatively lower encoder complexity. As illustrated in Fig. 1, by temporal multiplexing and spatial adaptive downsampling we can obtain two side descriptions of the original video signals. The resulting two side descriptions remain a conventional square sample grid, and thus it can be compressed and transmitted without any change to current video coding standards and systems. This property makes the proposed scheme more practical and general. At the decoder, each side description is first decompressed, and then reconstructed to original resolution with the help of the other side description. In this procedure, the decoder recovers the original video sequence in a constrained least squares regression process, using 2D or 3D piecewise autoregressive model depending on different prediction modes. If only one of the two descriptions is received, we also can produce lower, but still acceptable, quality reconstructions. Therefore, the proposed scheme has the error resilience ability to the error-prone

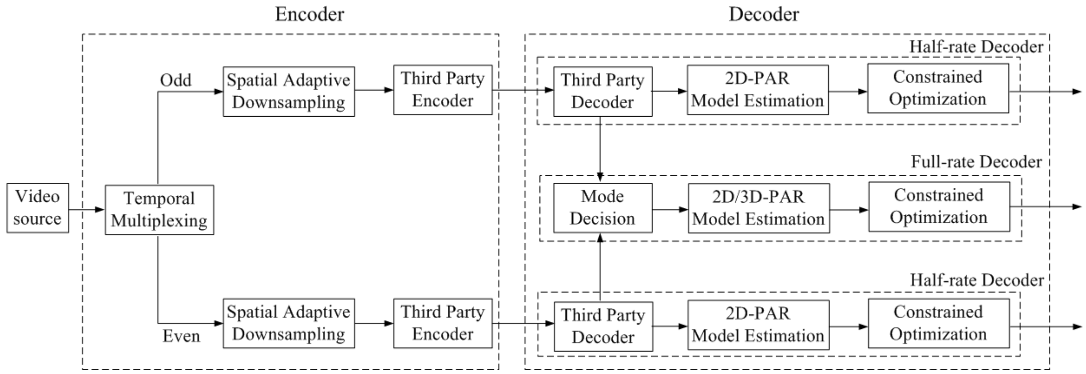


Fig. 1. The framework of the proposed video coding scheme

wireless channels. This manner is similar with multiple description coding [3].

The rest is organized as follows. Section II presents the proposed lightweight encoder. Section III details the design of decoder. Section IV presents some experimental results. Section V concludes the paper.

2. LIGHTWEIGHT ENCODER

In the encoder design, we propose a novel lightweight approach called temporal multiplexing with spatial adaptive downsampling. First, the input video frames are split by a simple multiplexer into odd and even side descriptions, each of which contains odd and even frames respectively. Each frame in two descriptions is then performed spatial downsampling to further compact the video signals. Finally, the two low-resolution (LR) descriptions are compressed by any third-party encoder (e.g., H.264) and transmitted to the destination by wireless channel.

For spatial downsampling, we choose not to perform uneven irregular down sampling of a video frame according to local spatial or frequency characteristics. Instead, we stick to conventional square pixel grid by uniform spatial down sampling of the frame with a factor of two. Yet the simple uniform downsampling scheme is made adaptive by a directional low-pass prefiltering step prior to downsampling. This manner is the same as that in [4]. The other purpose of this preprocessing is to induce a mechanism of collaboration between the spatial uniform downsampling process at the encoder and optimal upconversion process at the decoder. The sampling locations of odd and even descriptions are different, which are carefully designed to make the resulting LR side descriptions interleave with each other on HR sample grid in the temporal domain, as illustrated in Fig. 2. In this way, we can introduce structure correlation for two descriptions.

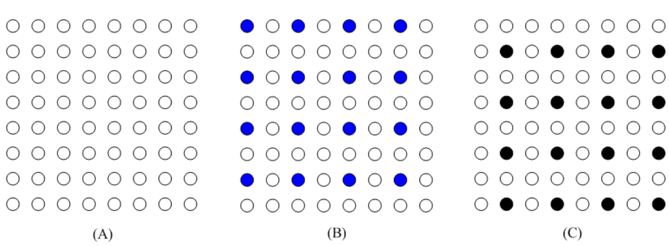


Fig. 2. Uniform downsampling with temporal multiplexing. (A) the original frame, (B) downsampled version (blue points) for odd frames; (C) downsampled version (black points) for even frames.

Now we can see downsampling can efficiently ease the burden of encoder and wireless channel by greatly reducing the amount of data needed to processing. In this way, it naturally extends the lifetime of the WVSNs. Meanwhile, since a uniform downsampling scheme is chosen, the generated two descriptions will constitute two LR video sequences, which can be compressed to further reduce the data rate by using any standard encoder (e.g. H.264). Therefore, the whole system remains standard compliant and practical.

3. MODE-DEPENDENT SOFT DECODING BY MODEL-BASED ADAPTIVE REGRESSION

At the decoder, the two side descriptions are individually decodable and mutually refinable. In this section, we will detail the decoder design.

3.1. The Interpolation Model

Let $\mathbf{y}$ be the low-pass filtered, down-sampled and compressed frame. The vector $\mathbf{y} \in \mathcal{Z}^M$ consists of M LR pixel values in a given lexicographical order, where $\mathcal{Z}$ is an integer alphabet from which the pixel values are drawn. What we want to do is to recover the underlying HR frame $\mathbf{x} \in \mathcal{Z}^N, N = 4M$. The formation of $\mathbf{y}$ from $\mathbf{x}$ is modeled as:

$$\mathbf{y} = \mathbf{DHx} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}$ is the low-pass filtering operation and $\mathbf{D}$ is the downsampling process. The term $\mathbf{n}$ is the quantization noise in compression. In what follows we develop a model-based reconstruction approach to perform up-sampling, inverse filtering and denoising jointly.

Reconstruction of $\mathbf{x}$ from $\mathbf{y}$ is inherently ill-posed. The performance of the reconstruction algorithm can be greatly improved if a good adaptive model is integrated into estimation. Motivated by the geometric constraint of edges and motion trajectory, we propose to use three-dimensional piecewise autoregressive (3D-PAR) model for video signals:

$$x(i,t) = \sum_{(u,k)\in S(i,t)} \alpha_{i,t}^{u,k} x(i+u, t+k) + n(i,t), \tag{2}$$

where i is spatial location and t is the frame number; $S(i,t)$ is the spatio-temporal support of the 3D-PAR model; $\alpha_{i,t}^{u,k}$ are the model parameters, and $n(i,t)$ is a random perturbation independent of video signal. Specially, we introduce two 6-order 3D-PAR models in our design, as illustrated in Fig.3. One is the diagonal model $AR_\times$ which consists of four 8-connected spatial neighbors and two temporal neighbors, and the other is the axial model AR_+ which consists of four 4-connected spatial neighbors and two temporal neighbors.

3.2. Mode Dependent Soft Decoding

We integrate the 3D-PAR model into the solution of the inverse problem as formulated in Eq. (1). In a local window S, our task is to jointly estimate the parameters of the interpolation model and the block of HR pixels $\mathbf{x} \in S$ such that the estimated model can optimally fit the estimated $\mathbf{x}$. Now the HR frame reconstruction from a compressed LR frame can be stated as the following constrained optimization problem:

$$\begin{aligned} &\min_{\mathbf{x},\alpha} \sum_{i \in W} \|x(i,t) - \sum_{(u,k)\in S(i,t)} \alpha_{u,k} x(i+u, t+k)\|^2, \\ &s.t. \quad \|\mathbf{y} - \mathbf{DHx}\|^2 < \sigma_n^2(r), \end{aligned} \tag{3}$$

where $\sigma_n^2(r)$ is the energy of the quantization noise of the compressed LR frame at bit rate r. Let L be the number of the LR pixels inside S, $\|\mathbf{y} - \mathbf{DHx}\|^2 < \sigma_n^2(r)$ corresponds to L inequality constraints.

If pixels in the local window S have weak temporal correlation with neighboring frames, such as S is a occlusion region, the accuracy of estimation will degrade heavily due to the introduction of uncorrelated temporal neighbors. At such a case, 2D-PAR should be utilized instead of 3D-PAR. How to adaptively choose 2D or 3D model is key to the performance of frame reconstruction. According to the statistical duality between LR frame and its HR counterpart, the prediction mode generated in LR descriptions compression could provide us some useful prior knowledge. In accordance with the H.264 standard, the proposed reconstruction algorithm can be divided into four modes: Intra, Skip, Inter-P, Inter-B.

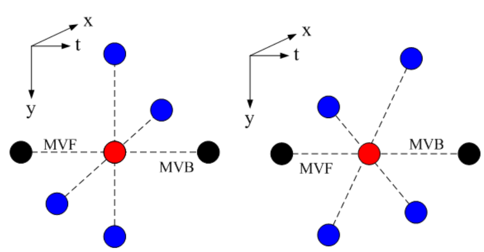


Fig. 3. Two used 6-order 3D-PAR model. The central pixel is in red, the spatial neighbors are in blue, and the temporal neighbors are in black which are aligned by MVs.

For Intra mode, the problem of frame reconstruction degrades to spatial image interpolation. The upconversion is based on the diagonal and axial 2D-PAR image models and on the deconvolution of the directional low-pass prefiltering. Incorporating these two PAR models into the original nonlinear estimation framework, we state the task of upconversion as the following constrained least squares problem:

$$\min_{\mathbf{x},\mathbf{a},\mathbf{b}} \left\{ \begin{array}{l} \zeta^{\times} \sum_{i \in W} \left\| x(i,t) - \mathbf{a}_s^T \mathbf{s}_{\times}(i,t) \right\|^2 + \\ \zeta^{+} \sum_{i \in W} \left\| x(i,t) - \mathbf{b}_s^T \mathbf{s}_{+}(i,t) \right\|^2 + \lambda \|\mathbf{y} - \mathbf{DHx}\|^2 \end{array} \right\} \tag{4}$$

where $\mathbf{s}_{\times}(i,t)$ and $\mathbf{s}_{+}(i,t)$ consist of four 8-connected and four 4-connected spatial neighbors of $x(i,t)$ in the HR image, $\mathbf{a}_s$ and $\mathbf{b}_s$ are model parameters of diagonal and axial models, $\zeta^{\times}$ and ζ^{+} are fusion weights to combine the modeling strength of the two PAR models. Note that the side decoder is performed in the same way as the Intra mode.

For Skip mode, LR pixels in the forward frame $t-1$ can be directly copied to the current sample grid to construct a quincunx lattice. With the quincunx lattice, Skip mode performs spatial interpolation to estimate other missing pixels. The optimization formulation is the same as Eq. (4), while with $2L$ inequality constraints since two times LR samples can be available. The increased constraints can provide more accurate estimation from the solution space.

For Inter-P and Inter-B mode, motion information is available to facilitate the task of resolving intensity uncertainty of video signals by exploiting the fundamental tradeoff between spatial and temporal correlation. The current pixel is approximated as the weighted combination of samples within its spatial neighborhoods as well as the temporal neighbors aligned by motion vectors. The task of upconversion can be stated as the following constrained least squares problem:

$$\min_{\mathbf{x},\mathbf{a},\mathbf{b}} \left\{ \begin{array}{l} \zeta^{\times} \sum_{i \in W} \left\| x(i,t) - (\mathbf{a}_s^T \mathbf{s}_{\times}(i,t) + \mathbf{a}_t^T \mathbf{t}(i,t)) \right\|^2 + \\ \zeta^{+} \sum_{i \in W} \left\| x(i,t) - (\mathbf{b}_s^T \mathbf{s}_{+}(i,t) + \mathbf{b}_t^T \mathbf{t}(i,t)) \right\|^2 \\ +\lambda \|\mathbf{y} - \mathbf{DHx}\|^2 \end{array} \right\}, \tag{5}$$

where $\mathbf{a}_s$ and $\mathbf{b}_s$ are spatial model parameters along diagonal and axial direction; $\mathbf{a}_t$ and $\mathbf{b}_t$ are temporal model parameters along the motion vector; $\mathbf{t}(i,t)$ is temporal reference sample set including forward reference sample for the Inter-P mode and bi-directional reference samples for the Inter-B mode.

3.3. Model Parameters Estimation

The accuracy of model parameters estimation directly influence the quality of reconstructed frames. Let us consider how to estimate the model parameters. For 2D-PAR model, according to the fact that the second order statistics of natural images tends to be invariant across different scales [5][6], we learn model parameters $\mathbf{a}_s$ and $\mathbf{b}_s$ from decoded image by solving the following two least-square estimation problems:

$$\begin{aligned} \mathbf{a}_s^* &= \min_{\mathbf{a}_s} \left\{ \sum_{i \in W} \left\| y(i,t) - \mathbf{a}_s^T \mathbf{s}_y^{\times}(i,t) \right\|^2 \right\}, \\ \mathbf{b}_s^* &= \min_{\mathbf{b}_s} \left\{ \sum_{i \in W} \left\| y(i,t) - \mathbf{b}_s^T \mathbf{s}_y^{+}(i,t) \right\|^2 \right\}. \end{aligned} \tag{6}$$

where $\mathbf{s}_y^{\times}$ and $\mathbf{s}_y^{+}$ are samples along diagonal and axial direction in the LR frame $\mathbf{y}$. Clearly, we can obtain a closed form solution for the above equation. As formulated in Eq. (4), we use $\zeta^{\times}$ and ζ^{+} to combine the modeling strength of the two PAR models. We can exploit the squared errors associated with the solutions of two objective functions in Eq. (6) to determine these two fusion weights:

$$\zeta^{\times} = \frac{e^{+}}{e^{+} + e^{\times}}, \zeta^{+} = \frac{e^{\times}}{e^{+} + e^{\times}}. \tag{7}$$

These weights are optimal in least squares sense if the fit errors of the two models are independent.

For 3D-PAR, according to 2D-3D duality between edge contour and motion trajectory, the model parameters are adaptively estimated within a localized spatio-temporal window in the current side description video sequence. Similarly, the derivation $\mathbf{a} = [\mathbf{a}_s, \mathbf{a}_t]$ and $\mathbf{b} = [\mathbf{b}_s, \mathbf{b}_t]$ follows the standard LS formulation:

$$\mathbf{a}^* = \min_{[\mathbf{a}_s, \mathbf{a}_t]} \left\{ \sum_{i \in W} \left\| y(i,t) - (\mathbf{a}_s^T \mathbf{s}_y^{\times}(i,t) + \mathbf{a}_t^T \mathbf{t}_y(i,t)) \right\|^2 \right\}$$
$$\mathbf{b}^* = \min_{[\mathbf{b}_s, \mathbf{b}_t]} \left\{ \sum_{i \in W} \left\| y(i,t) - (\mathbf{b}_s^T \mathbf{s}_y^{+}(i,t) + \mathbf{b}_t^T \mathbf{t}_y(i,t)) \right\|^2 \right\} \tag{8}$$

We also can obtain a closed form solution for the above equation, and can obtain $\zeta^{\times}$ and ζ^{+} in a similar way with Eq. (7).

Once the PAR model is constructed, soft decoding can be performed efficiently by constrained linear least-square estimation. For convenient representation we rewrite Eq. (4) and Eq. (5) in matrix form:

$$\mathbf{x}^* = \min_{\mathbf{x}} \left\{ \begin{array}{l} \zeta^{\times} \sum_{i \in W} \|\mathbf{x} - \mathbf{C}_1 \mathbf{x}\|^2 + \zeta^{+} \sum_{i \in W} \|\mathbf{x} - \mathbf{C}_2 \mathbf{x}\|^2 \\ +\lambda \|\mathbf{y} - \mathbf{D}\mathbf{H}\mathbf{x}\|^2 \end{array} \right\} \tag{9}$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are two matrixes containing PAR model parameters. The objective function can be further written in quadratic form as:

$$\min_{\mathbf{x}} \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x}), \tag{10}$$

where the residue vector $\mathbf{r}(\mathbf{x})$ is defined as:

$$\mathbf{r}(\mathbf{x}) = \begin{bmatrix} \sqrt{\zeta^{\times}}(\mathcal{I} - \mathbf{C}_1)\mathbf{x} \\ \sqrt{\zeta^{+}}(\mathcal{I} - \mathbf{C}_2)\mathbf{x} \\ \sqrt{\lambda}(\mathbf{y} - \mathbf{D}\mathbf{H}\mathbf{x}) \end{bmatrix}. \tag{11}$$

And the objective function in Eq.(10) is a linear least square problem that can obtain a close-form solution as

$$\mathbf{x} = (\mathcal{F}^T \mathcal{F})^{-1} \mathcal{F}^T \mathcal{G} \tag{12}$$

with

$$\mathcal{F} = \begin{bmatrix} \sqrt{\zeta^{\times}}(\mathcal{I} - \mathbf{C}_1) \\ \sqrt{\zeta^{+}}(\mathcal{I} - \mathbf{C}_2) \\ \sqrt{\lambda}\mathbf{D}\mathbf{H} \end{bmatrix}. \tag{13}$$

and

$$\mathcal{G} = \begin{bmatrix} 0 \\ 0 \\ -\sqrt{\lambda}\mathbf{y} \end{bmatrix}. \tag{14}$$

4. EXPERIMENTAL RESULTS

In this section, experimental results are presented to verify the performance of the proposed video coding scheme with respect to rate-distortion performance and subjective quality. For thoroughness and fairness of our comparison study, we selected six video sequences as test ones, including *Akiyo* (CIF), *Foreman* (CIF), *Mother&Daughter* (CIF), and *City* (4CIF), *Crew* (4CIF), *Ice* (4CIF). These sequences are with frame rate 30Hz, and each sequences contains 100 frames.

The following video codecs will be used as benchmarks to evaluate the performance of the proposed codec.

- H.264-Motion: This codec is performed on JM16.0 [7] in main profile exploiting spatial and temporal redundancy (i.e., intra and inter prediction are both selected). The GOP size is 6 with IBPBPB structure. The RD optimization is done in the high-complexity mode. The loop filter is enabled. Entropy coding is performed in CABAC mode, and the search range of motion estimation is set to 32. It can be considered as the state-of-the-art encoder-centralized video codec.
- H.264-Zero Motion: We use JM16.0 main profile to code the GOP of 24 frames with the first frame coded as I frame and all other frames coded as predictive frames, for which the rest settings are the same as H.264-Motion except that the range of motion search is set to 1. It is often used as a benchmark in comparison for non-ME based low complexity video codecs.

Since DVC and H.264-Intra give too poor RD performances, we have not included their results into comparison. The performance comparison among our method and other methods in the literature can be indirectly reflected by H.264/AVC standard codecs. Through comparisons with these two standard codecs motioned above, results are shown that serve to support the efficiency of the proposed scheme for low bit-rate coding.

4.1. Encoder Complexity Comparison

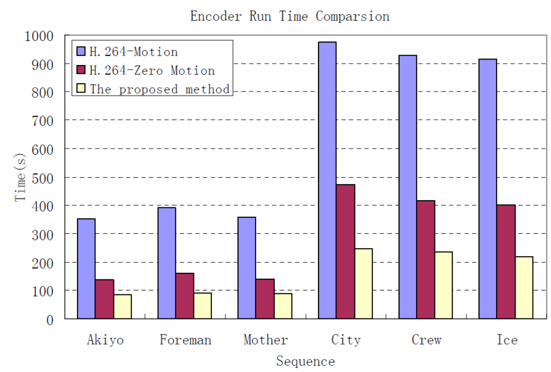


Fig. 5. Encoder complexity comparison

We first give the encoder running time comparison of the compared three codecs on six test sequences. The compared codecs are run on a typical computer (2.5GHz Intel Dual Core, 4G Memory). For each sequence, we keep the bit-rates of three codecs almost the same. As illustrate in Fig. 5, it is easy to find the proposed method achieves lowest encoder complexity, the running time is even lower than H.264-Zero Motion. The running time is about 1/4 of that of H.264-Motion, because the video data needed to be compressed reduces to 1/4 of original ones by downsampling. These results demonstrate our method can provide a lightweight encoder, which is attractive for resource-deficient wireless video communications.

4.2. Rate-distortion Performance Comparison

To verify the performance improvement of the propose scheme at low bit rates, we use coarse quantization parameters (QP) to obtain rate-distortion curves shown below. And the comparisons shown here are all for approximately the same average bit rate over the entire sequence and therefore can be readily compared in terms of the PSNR values. The RD curves of six video sequences are plotted in Fig. 4. We can find for all six test sequences the proposed scheme can achieve better RD performance at low bit-rate compared with other two codecs. The gain is up to 1.5dB for CIF sequences and 1.2dB for 4CIF sequences compared with H.264-Motion, which is regarded as the state-of-the-art video codec. Compared with H.264-Zero Motion, which is also with a low-complexity encoder, the gains are obvious and is up to 1.5dB for CIF sequences and 3dB for 4CIF sequences. According to the trend of RD curves, the proposed method outperform H.264-Zero Motion at a wide range of bit-rate.

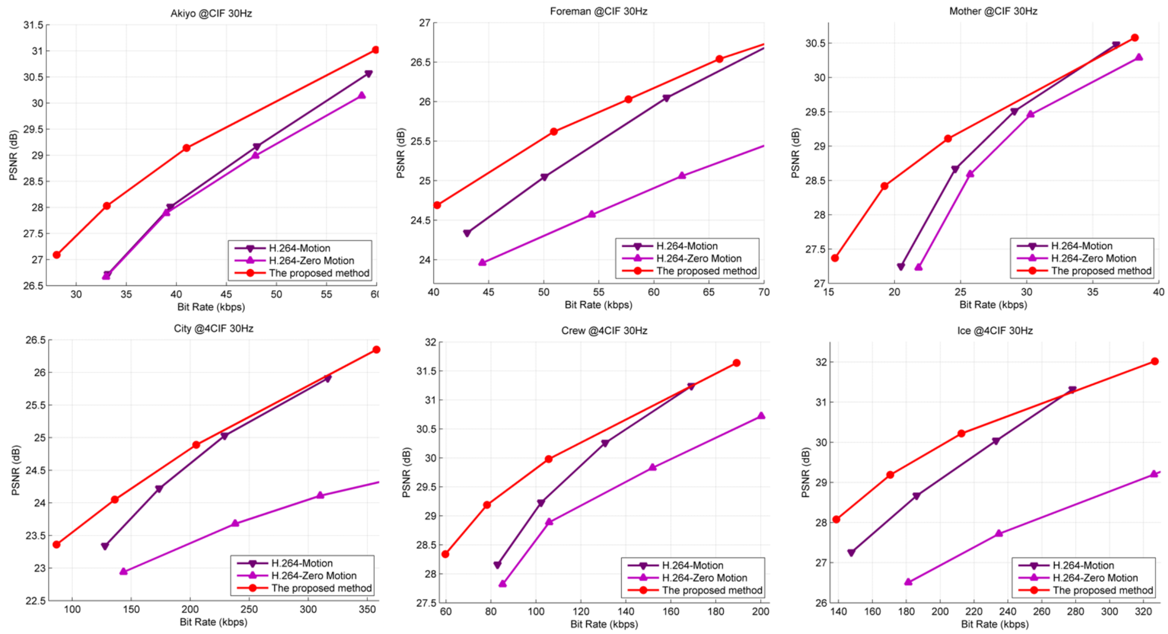


Fig. 4. Rate-distortion performance comparison

4.3. Subjective Performance Comparison

The advantage of the proposed scheme is not only limited to low bit-rates. We also give the subjective comparison results at medium bit-rates. As illustrate in Fig. 6 and Fig. 7, we show the decoded frames of two 4CIF sequences by H.264-Zero Motion, H.264-Motion and the proposed side and central decoders. For clearly comparison, the corresponding PSNR values and bit-rates are also given. From the results, we can find H.264-Zero Motion produces objectionable visual artifacts (e.g., jaggies and ringings) in edge areas, H.264-Motion performs better but still suffers from annoying blurring artifacts along the edges. The proposed schemes on side and central decoder are both largely free of those defects. Even when the bit rate gets higher and H.264-Motion starts to have higher PSNR than the proposed method, its visual quality still appears inferior, as demonstrated by examples in Fig. 6 and Fig. 7. This is due to the fact quantization in H.264/AVC standard is uniform and there is no special mechanism to preserve edges that are important for human visual perception. The proposed side decoder achieves a lower but still acceptable subjective quality, and the quality is significantly enhanced at the central decoder. The results are visually compelling in reconstructing edges and textures. The produced edge and texture are clean and sharp, and most visual artifacts appeared in the results of H.264-Motion are eliminated in the proposed method. These results demonstrate the proposed method can efficiently favors the reconstruction of edges. The superior visual quality of the proposed method is due to the good fit of the piecewise autoregressive model to edge structures and the fact that human visual system is highly sensitive to phase errors in reconstructed edges.

5. CONCLUSION

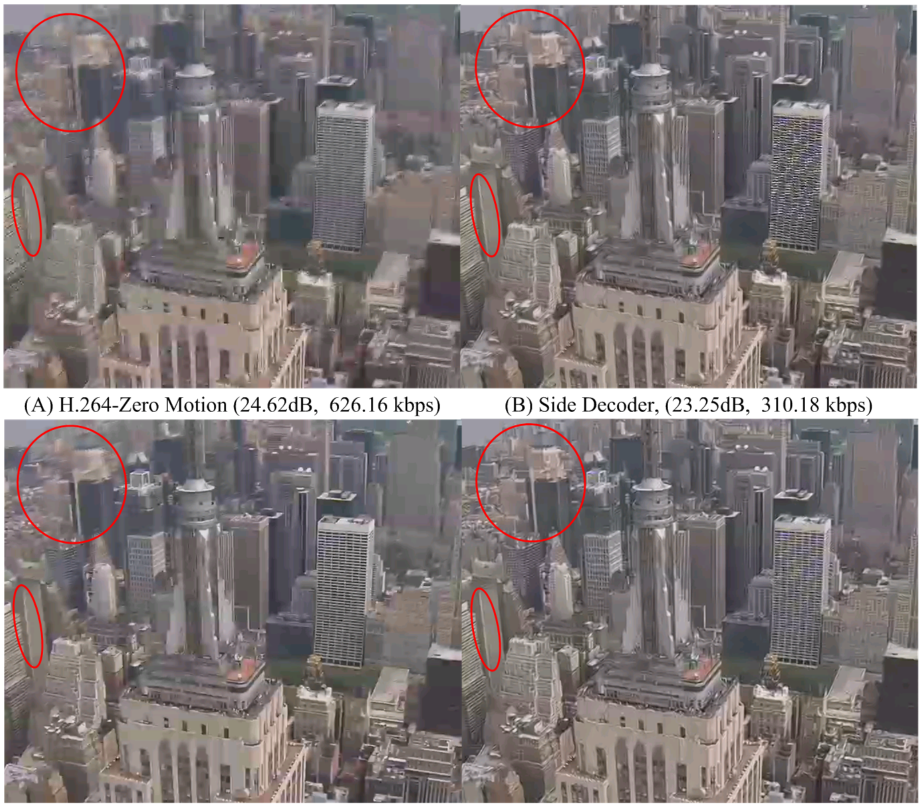
In this paper, we presented an efficient and standard-compliant low bit-rate video coding scheme for wireless visual communications. The encoder is designed to be relative simple, where multiple low-resolution descriptions are generated by temporal multiplexing and spatial adaptive downsampling. At the decoder, the mode-dependent soft-decoding is performed to jointly estimate the model and the original frame. Simulations results show the proposed method outperforms H.264/AVC at low bit-rates.

6. ACKNOWLEDGEMENT

This work was supported in part by the Major State Basic Research Development Program of China (973 Program 2009CB320905), the Program for New Century Excellent Talents in University (NCET) of China (NCET-11-0797), and the National Science Foundation of China (NSFC) under grants 61100096 and the Fundamental Research Funds for the Central Universities (Grant No. HIT.BRETIII.201221).

7. REFERENCES

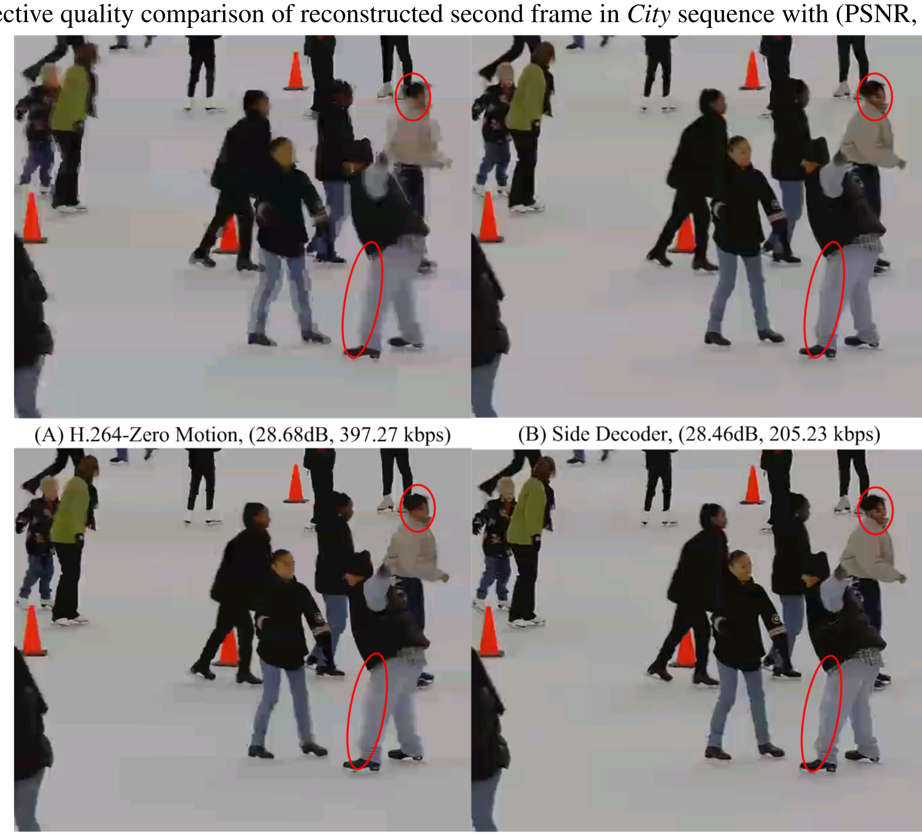
[1] B. Girod, A. Aaron, S. Rane, D. Monedero, “Distributed video coding,” Proceedings of the IEEE, vol.93, no.1, pp.71-83, 2005.
[2] E. Cands, “Compressive sampling,” in Proc. Int. Congr. Mathematics, Madrid, Spain, 2006, pp. 1433-1452.
[3] Y. Wang, A. Reibman, and S. Lin, “Multiple description coding for video delivery ,” Proc. IEEE, vol. 93, pp. 57-70, 2005.
[4] X. Wu, X. Zhang, and X. Wang, “Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion,” IEEE Trans. Image Process., vol. 18, no. 3, pp. 552-561, Mar. 2009.
[5] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao and H. Sun, “ Image Interpolation Via Regularized Local Linear Regression,” IEEE Trans. on Imag. Process., vol.20, no.12, pp.3455-3469, Dec.2011.
[6] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao and H. Sun,, “Transductive Regression with Local and Global Consistency for Image Super-Resolution,” in Proceedings of IEEE Data Compression Conference, DCC2011, Snowbird, Utah, USA, Mar.29-31, 2011.
[7] H.264/AVC Reference Software, JM16.0, Online Available: http://iphome.hhi.de/suehring/tml/



(A) H.264-Zero Motion (24.62dB, 626.16 kbps) (B) Side Decoder, (23.25dB, 310.18 kbps)

(C) H.264 Motion, (27.87dB, 610.19 kbps) (D) Central Decoder, (26.46dB, 625.14 kbps)

Fig. 6. Subjective quality comparison of reconstructed second frame in *City* sequence with (PSNR, Bit Rate) pairs.



(A) H.264-Zero Motion, (28.68dB, 397.27 kbps) (B) Side Decoder, (28.46dB, 205.23 kbps)

(C) H.264 Motion, (33.77dB, 401.01 kbps) (D) Central Decoder, (33.36dB, 413.44 kbps)

Fig. 7. Subjective quality comparison of reconstructed second frame in *Ice* sequence with (PSNR, Bit Rate) pairs.