J. Vis. Commun. Image R. 23 (2012) 535–548

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Perceptual-based distributed video coding

Yu-Chen Sun, Chun-Jen Tsai *

Dept. of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan

ARTICLE INFO

Article history:
Received 22 July 2011
Accepted 20 January 2012
Available online 2 February 2012

Keywords:
Distributed video coding
Perceptual-based coding
Wyner–Ziv coding
Region-of-interest analysis
Motion consistency analysis
Texture consistency analysis
Visual distortion estimation
Side-information error classification

ABSTRACT

In this paper, we propose a perceptual-based distributed video coding (DVC) technique. Unlike traditional video codecs, DVC applies video prediction process at the decoder side using previously received frames. The predicted video frames (i.e., side information) contain prediction errors. The encoder then transmits error-correcting parity bits to the decoder to reconstruct the video frames from side information. However, channel codes based on i.i.d. noise models are not always efficient in correcting video prediction errors. In addition, some of the prediction errors do not cause perceptible visual distortions. From perceptual coding point of view, there is no need to correct such errors. This paper proposes a scheme for the decoder to perform perceptual quality analysis on the predicted side information. The decoder only requests parity bits to correct visually sensitive errors. More importantly, with the proposed technique, key frames can be encoded at higher rates while still maintaining consistent visual quality across the video sequence. As a result, even the objective PSNR measure of the decoded video sequence will increase too. Experimental results show that the proposed technique improves the R-D performance of a transform domain DVC codec both subjectively and objectively. Comparisons with a well-known DVC codec show that the proposed perceptual-based DVC coding scheme is very promising for distributed video coding framework.

1. Introduction

Distributed video coding (DVC) [1] has received much attention for the past decade because, theoretically, DVC allows flexible distribution of coding complexity between the encoder and the decoder without losing compression efficiency [2,3]. This property makes DVC a potential solution for emerging applications such as surveillance systems, sensor networks, and Mobile 2.0 (e.g., uploading live videos to social networks using mobile phones) where encoders have limited computation capability due to power consumption issues [4]. A practical approach to shift coding complexity is achieved by performing motion prediction only at the decoder side. In many DVC implementations [1,5–8], a video sequence is divided into two interleaving sub-sequences: key frames and Wyner–Ziv (WZ) frames. The encoder uses intra-coding to encode the key frames while the decoder uses the reconstructed key frames to predict the WZ frames. The predictors of WZ frames are called side information (SI). SI prediction errors are then corrected by channel codes [9] to reconstruct the WZ frames. SI prediction, Slepian-Wolf codecs, and WZ reconstruction are the most critical components in a DVC codec.

Motion-compensated frame interpolation methods are often adopted for SI prediction [10,11]. To improve the quality of motion vector estimates, error surface of motion estimation is integrated into WZ decoding iteration to find a MAP solution of the motion field in [12]. A multi-hypothesis overlapped frame interpolation technique is proposed in [13] to increase the quality of the SI predictions. For WZ reconstruction, different channel codes are used to correct the prediction errors of SI frames [9,14]. The correction efficiency of a channel code is highly dependent on the error statistics of SI. In [15] the correlation model between the source and the side information (with single or multiple hypotheses) is explored for optimal reconstruction of quantized samples. In [16], the SI error model is studied, and then a MAP-based decoding method is proposed for better WZ reconstruction.

Even with the progresses in SI prediction, channel codes, and WZ reconstruction techniques, there are many areas in video sequences that simply cannot be efficiently coded using DVC tools. To improve coding performance, some researchers propose to adopt intra coding or skip modes in WZ frames when the SI areas have certain error characteristics. However, in DVC, it is not trivial to get error statistics of SI at the encoder side since SI is only available in the decoder while original video data is available only to the encoder. For encoder-side mode decision methods, [17–19] proposes mode decision scheme based on co-located SAD (i.e., zero-motion SAD). If the SAD between a WZ macrolbock and its co-located previous key frame is small, the WZ macrolbock is excluded from WZ reconstruction. In [20], a mode decision method based on low-complexity motion estimation technology is proposed. However, the

* Corresponding author.
E-mail address: cjtsai@cs.nctu.edu.tw (C.-J. Tsai).

1047-3203/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2012.01.015

performance of these encoder-side decision methods is constrained by encoder complexity. Therefore, in recent years, the possibility of decoder-side mode decision is investigated. A decoder-side skip mode decision policy is derived in [21] based on the error correlation model proposed in [22]. However, all the mode decision policies (either encoder-side or decoder-side) are based on estimates of SI errors of the WZ frames. Unfortunately, the true SI is missing on the encoder side, while the original frame is missing on the decoder side. As a results, it is difficult to estimate true SI errors.

In this paper, we propose a very different approach of perceptual-based decoder-side skip mode strategy. The proposed technique comes from the key observation that SI frames predicted using motion–projection algorithms often contain image areas with large prediction errors (in MSE sense) but small visual distortions. One such example is a video sequence of a low-motion scene taken by a shaky camera. If we remove every other frame and interpolate the missing frames with motion–projection algorithms, the resulting video may become smoother without major visual distortions except at image boundaries. However, to reconstruct the original shaky video from the interpolated frames using channel codes requires significant amount of parity bits, which is not worthwhile from perceptual rate-distortion perspective.

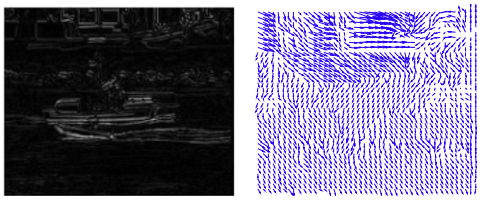
In short, the proposed technique performs perceptual-based analysis to determine the SI regions where visual distortions are noticeable, and only uses channel codes to correct these regions. This paper is organized as follows. Section 2 presents the rationale behind perceptual-based coding for DVC. Some examples comparing perceptual-based and SAD-based WZ coding are shown in this section to shed light on the proposed scheme. The proposed perceptual-based DVC codec is described in Section 3. Experimental results are presented in Section 4. Finally, conclusions and discussions are given in Section 5.

2. Perceptual-based coding for DVC

Fig. 1 shows the SI prediction frame generated by motion–projection using two neighboring key frames. The SI in this example is particularly interesting because the PSNR differences between the key frames (31.2 and 29.9 dB) and the in-between SI frame (17.5 dB) are more than 10 dB, but visually, video quality across the key frames and the SI frame are still consistent. Therefore, if perceptual quality is the coding goal, there is no need to request parity bits to correct the large amount of errors in this SI frame. Note that the main reason that the PSNR of the SI is low is because the scene is shaky due to slight camera motion. The motion field of the SI frame predicted from the two key frames is a smooth field different from the true motion field. The interpolated

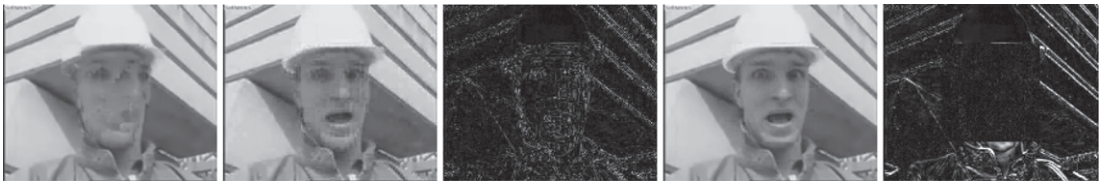
(a) Original frames 37, 38, and 39 (from left to right) of the 15 Hz Coastguard sequence.

(b) The SI frame (middle) and the two key frames (left and right) used for SI prediction.



(c) The error image of SI and the projected motion field between the SI frame and key frame 37.

Fig. 1. An example of SI with good perceptual quality but low PSNR. The PSNR of the SI frame is only 17.5 dB, but its visual quality is comparable to the neighboring key frames (around 30 dB).



(a) SI frame (b) Full correction (c) Error of (b) (d) Partial correction (e) Error of (d)

Fig. 2. Perceptual-based reconstruction of SI prediction errors. The reconstructed WZ frame in (b) tries to correct the prediction errors of the whole SI frame using 12.38 kbits of LDPCA codes. The reconstructed WZ frame in (d) only corrects a rectangular area that contains the face of Foreman using 10.46 kbits of LDPCA codes.

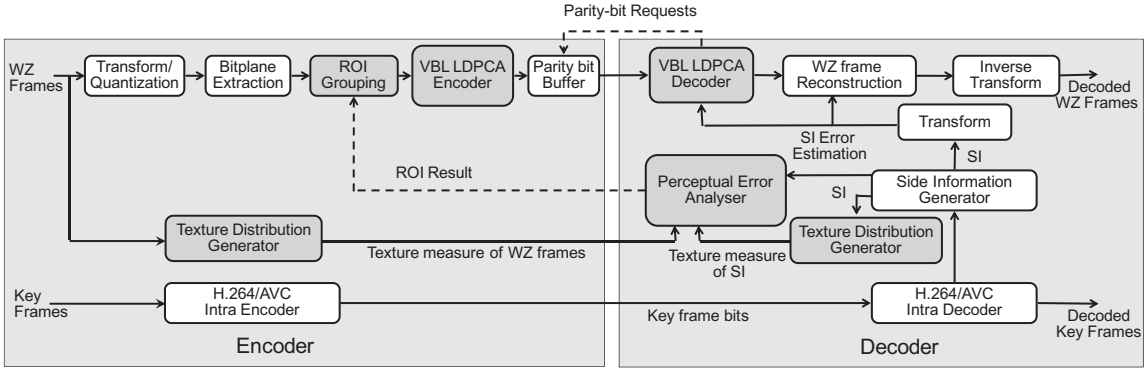


Fig. 3. The proposed DVC architecture. Gray blocks are proposed modules.

SI frame using this predicted motion field is visually appealing, but is different from the original frame.

Another issue with WZ frame reconstruction is that the error distribution of SI prediction is spatially varying [23]. Fig. 2 shows one such example. Fig. 2(a) is a predicted SI frame of the 48th frame of the Foreman sequence (QCIF@15 Hz). The SI error characteristics are affected by both the texture and motion complexities in video contents. Fig. 2(b) shows the reconstructed WZ frame using LDPCA code (the amount of parity bits are 12.38 kbits). The PSNR is 31.22 dB. There are spatially varying burst errors in the SI frame and it is ineffective to assume a spatially invariant i.i.d. SI error model and try to use channel codes to correct these errors. Burst errors usually happen at moving edge boundaries. However, existing DVC techniques [1,5–8] group consecutive macroblocks in scanline order into a coding block without taking into account the texture and motion characteristics of these macroblocks.

If we only correct the facial area (using 10.46 kbits), the partially reconstructed WZ frame looks visually more appealing, as shown in Fig. 2(d). Note that the sharp straight lines of the background building in the SI frame have uniform pixel-shift errors that cause a low PSNR value (28.62 dB). However, visually, there are no perceptible errors. On the other hand, the image in Fig. 2(b) uses part of the bit budget to correct the sharp edges towards the correct pixel position to certain degree. Unfortunately, such halfway correction produces fuzzy edges and degrades visual quality. In short, if a decoder can determine the regions of interest (ROI) automatically, and applies WZ reconstruction only in the ROI, we can achieve better visual quality at lower WZ rates. In addition, more bit budget can be allocated to key frames to further improve overall R-D performance [26]. In this paper, we define ROI as the areas in the SI frame where distortions are perceptually salient.

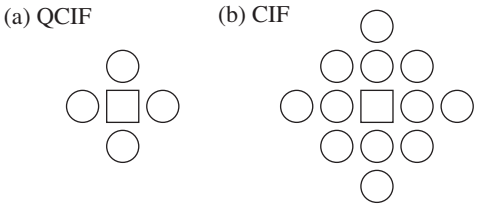


Fig. 6. Neighborhood structure of the ROI refinement process. The squares are the macroblocks under consideration and the circles are their neighbors.

It is important to point out that the main strength of the proposed scheme is not just to deal with the extreme cases illustrated in Fig. 1 or Fig. 2. Since all video frames are captured with noises, as a result, the predicted SI frame usually contains noises inherited (motion-compensated) from the key frames that are different from the noises in the original WZ frames. With the proposed approach, we will not waste syndrome bits on the correction of one set of sample noises to another set of sample noises, unless they are visually significant.

3. The proposed DVC framework

Fig. 3 is the block diagram of the proposed DVC codec. We have added perceptual-based coding tools to a transform-domain DVC

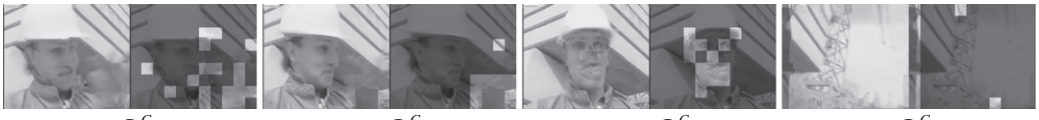


Fig. 4. Examples of SI frames (left) and detected visually distorted macroblocks (right).

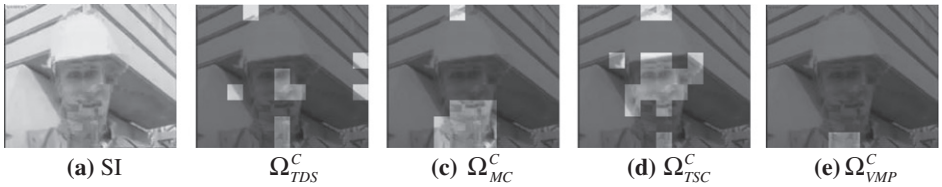


Fig. 5. An example of the 72nd SI frame and the detected visually distorted macroblocks.

Fig. 7. Examples of ROI detection results. Macroblocks with normal gray levels are in Ω_{ROI}.

Table 1
Impact of gop size on the proposed perceptual metrics.

	GOP size	TDS^c (%)	MC^c (%)	TSC^c (%)	VMP^c (%)	ROI (%)
Foreman	2	6.3	14.0	14.0	3.4	36.0
	4	9.4	30.1	45.0	5.0	68.8
	8	15.1	36.5	50.2	6.5	77.7

Table 2
Encoding time comparison for forman, QCIF@15fps.

DISCOVER		Proposed DVC	
(QP,QM)	Encoding time (ms)	(QP,QM)	Encoding time (ms)
(34,1)	3629	(34,1)	3690
(34,5)	3735	(34,5)	3741
(34,8)	3777	(34,8)	3793
AVC Intra		AVC zero-motion	
QP	Encoding time (ms)	QP	Encoding time (ms)
34	7249	34	10236

Table 3
Decoding time comparison for forman, QCIF@15fps.

DISCOVER		Proposed DVC	
(QP,QM)	Decoding time (ms)	(QP,QM)	Decoding time (ms)
(34,1)	1.0×10^7	(34,1)	0.9×10^7
(34,5)	1.6×10^7	(34,5)	$\mathrm{l.l} \times 10^7$
(34,8)	3.4×10^7	(34,8)	1.8×10^7
AVC intra		AVC zero-motion	
QP	Decoding time (ms)	QP	Decoding time (ms)
34	227	34	204

framework. The baseline implementation of the DVC codec is similar to the DISCOVER codec [8], plus prioritized macroblock grouping [24]. An AVC/H.264 intra coder is used to encode key frames and an LDPCA code [9] is used to correct SI frames. For each SI macroblock, the decoder performs perceptual distortion analysis and discriminates whether the macroblock belongs to the ROI. The encoder receives the ROI information from the decoder via a feedback channel and groups ROI macroblocks into the same coding block. Since the coding block size varies from frame to frame, the LDPCA module must handle variable block-length (VBL) coding. The rest of this section describes the proposed perceptual-based error analysis.

3.1. Texture distribution similarity (TDS) analysis

Based on our empirical investigations, bursty SI prediction errors often happen at the boundaries of texture-rich moving objects. Therefore, the first step in the proposed perceptual-based error analysis is to identify whether the distribution of texture-rich macroblocks in the SI frame is the same as that in the quantized original frame. If an SI macroblock is texture-rich while the corresponding quantized original macroblock is not (or vice versa), the SI macroblock should be corrected by parity bits. Since the original video frames are available only to the encoder, the encoder must compute a texture distribution map of the quantized original frame and transmit it to the decoder for analysis.

The texture distribution generator first determines edge pixels in a frame using the Sobel edge operator and a threshold θ_{edge}. A macroblock is considered a texture-rich block if the percentage of edge pixels in the block is larger than a threshold $\theta_{texture}$. Finally, the distribution of texture-rich blocks is recorded using a bit map, one bit per macroblock. In the bit map, a '1' signals a texture-rich macroblock while a '0' signals a regular macroblock.

The selection of the two thresholds, θ_{edge} and $\theta_{texture}$, are described as follows. The thresholds are adaptive to the video contents. To determine the distribution of texture-rich blocks of a WZ frame at t, we compute θ_{edge} and $\theta_{texture}$ of a frame at time t as in Eq. (1):

$$\theta_{edge}(t) = \frac{1}{M}\sum_{\mathbf{p}=1}^{M} s_t(\mathbf{p}),$$
$$\theta_{texture}(t) = \frac{100}{M}\sum_{\mathbf{p}=1}^{M} P_{edge,t}(\mathbf{p}), \quad (1)$$

Table 4
Breakdown of encoding time per frame (in ms) for the proposed codec.

QM	Intra encoding	LDPCA encoding	Computing Sobel edge	Computing texture distribution map
1	47	1.3	0.77	5.0×10^{-2}
5		1.9		
8		2.6		

Table 5
Breakdown of decoding time per frame (in ms) for the proposed codec.

QM	Intra decoding	SI generation	TDS analysis	MC analysis	TSC analysis	VMP analysis	ROI calculation	LDPCA decoding
1	1.5	$\mathrm{l.l} \times 10^5$	0.67	0.25	1.4	0.23	1.4×10^{-2}	1.2×10^4
5								3.9×10^4
8								1.4×10^5

where $s_t(\mathbf{p})$ is the sum of magnitudes of horizontal and vertical Sobel edge strength of pixel $\mathbf{p}$ of the WZ frame at time t, and M is the number of pixels in a key frame. $P_{edge,t}(\mathbf{p})$ is a binary function of edge map, i.e., $P_{edge,t}(\mathbf{p})$ is 1 if $\mathbf{p}$ is an edge pixel and $P_{edge,t}(\mathbf{p})$ is 0 otherwise. While the encoder generates the texture distribution map of the WZ frame, the decoder uses the same algorithm to generate the texture distribution map of the SI. The decoder can then compare the received texture distribution map with the texture distribution map of SI. The set of SI macroblocks that have the same texture property as the corresponding original macroblocks is denoted by Ω_{TDS}.

3.2. Motion consistency (MC) analysis

Motion behavior is a useful cue for estimating SI prediction quality. For example, the optical flow field [25] between neighboring key

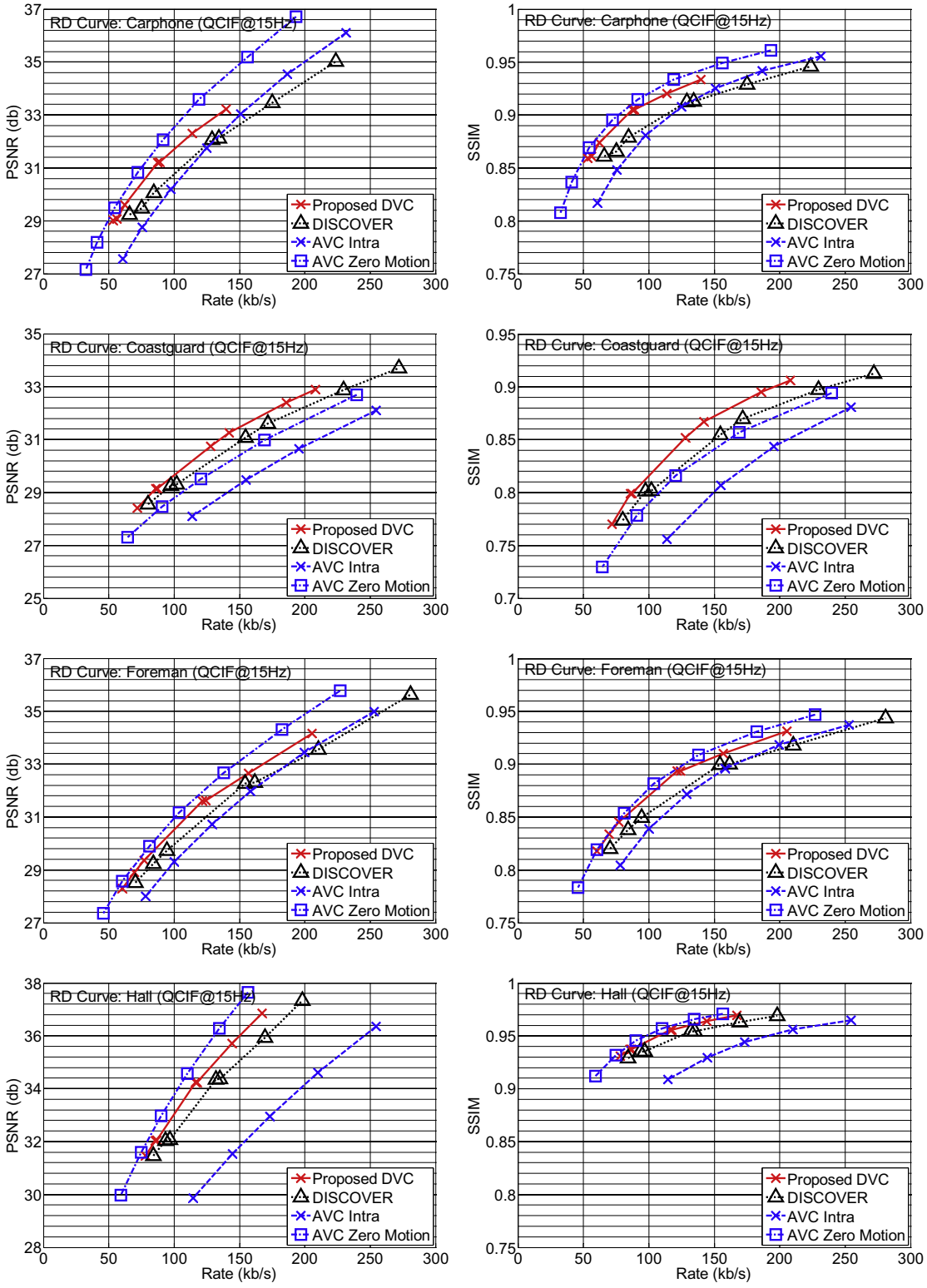


Fig. 8. QCIF sequences R-D performance comparisons using PSNR and SSIM. The average BD PSNR gain over DISCOVER is 0.71 dB, and the average SSIM gain is 0.016.

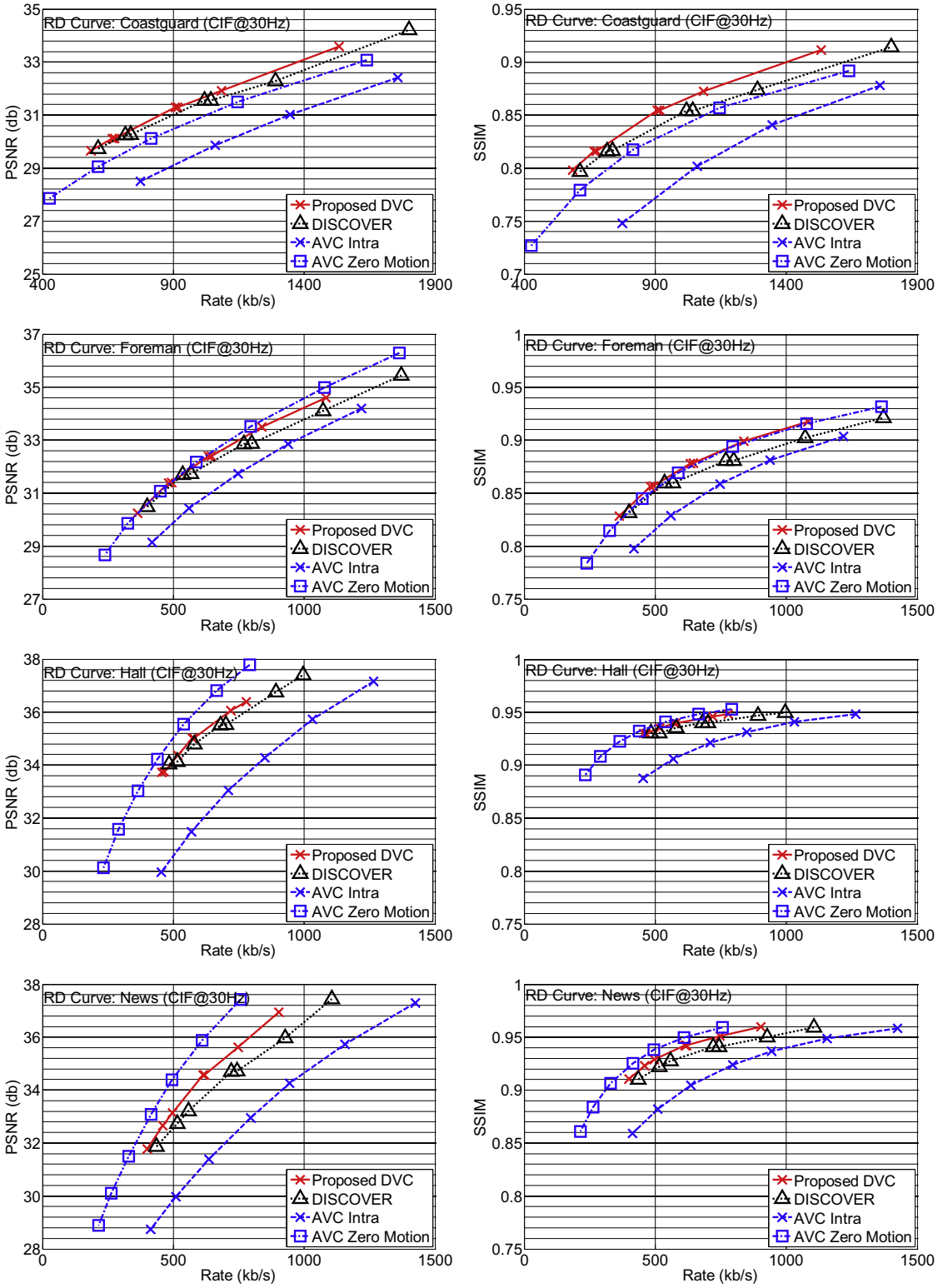


Fig. 9. R-D performance comparisons using PSNR and SSIM. The average PSNR BD gain over DISCOVER is 0.41 dB, and the average SSIM gain is 0.010.

frames is a good indication of how well the true motion field matches the block-based motion model [20]. If the motion field is irregular in a highly textured area, the visual distortion of the predicted SI may be large. For the SI frame at time t, we first calculate the motion smoothness $\delta_t(\mathbf{k})$ measure of a 4×4 block at block position $\mathbf{k}$ using Eq. (2):

$$\delta_t(\mathbf{k}) = \frac{1}{8} \sum_{\mathbf{j} \in \Omega_N(\mathbf{k})} \|\mathbf{v}_f(\mathbf{k}) - \mathbf{v}_f(\mathbf{j})\|_2, \tag{2}$$

where $\Omega_N(\mathbf{k})$ is the set of eight direct neighbors of block $\mathbf{k}$, and $\mathbf{v}_f(\mathbf{k})$ is the estimated forward motion vector of block $\mathbf{k}$ for the SI frame at

Table 6
Quantization Setting (QP,QM) of DVC codecs in the experiments.

QCIF sequences							
Carphone	(40,1)	(40,2)	(39,3)	(36,4)	(36,5)	(34,6)	(32,7)
Coastguard	(38,1)	(37,2)	(37,3)	(34,4)	(33,5)	(31,6)	(30,7)
Foreman	(40,1)	(39,2)	(38,3)	(34,4)	(34,5)	(32,6)	(29,7)
Hall	(37,1)	(36,2)	(36,3)	(33,4)	(33,5)	(31,6)	(29,7)
CIF sequences							
Coastguard	(37,1)	(36,2)	(36,3)	(34,4)	(34,5)	(33,6)	(30,7)
Foreman	(39,1)	(37,2)	(37,3)	(35,4)	(35,5)	(33,6)	(31,7)
Hall	(35,,1)	(35,2)	(34,3)	(33,4)	(33,5)	(31,6)	(30,7)
News	(38,1)	(37,2)	(36,3)	(34,4)	(34,5)	(32,6)	(30,7)

time t. The motion consistency measure $\Delta_t(i)$ of macroblock i is defined using Eq. (3):

$$\Delta_t(i) = \mathrm{Max}_{\mathbf{k}\in\Omega_B(i)}(\delta_t(\mathbf{k})), \tag{3}$$

where $\Omega_B(i)$ is the set of sixteen 4×4 blocks of macroblock i. Macroblocks whose motion consistency measures are smaller than a threshold θ_{MC} belong to the set Ω_{MC} of macroblocks with consistent motion. The threshold is adaptively calculated using Eq. (4):

$$\theta_{MC}(t) = \mu_{\Delta_t} + \sigma_{\Delta_t}, \tag{4}$$

where μ_{Δ_t} and σ_{Δ_t} are the mean and standard deviation of $\Delta_t()$ of SI frame t, respectively. Note that, the motion vector estimates at low-texture areas are unreliable. Thus, we set θ_{MC} to infinity for macroblocks with no textures so that such macroblocks are always counted as motion-consistent macroblocks. We use the average Sobel edge strengths of a macroblock to determine its texture level. If the total edge strength of a macroblock is larger than a threshold θ_{TL}, it is treated as a texture-rich macroblock. The threshold θ_{TL} is set to 50 in this paper and it is not a sensitive parameter (any values from 50 to 100 produces similar results for all the test sequences).

Statistically speaking, the policy for selecting θ_{MC} will include a fixed percentile of the SI macroblocks into the set Ω_{MC}. In theory, for the detection of macroblocks with irregular motions, a sequence dependent fixed-value threshold, instead of a fixed-percentile threshold, should be used. However, our experiments show that Eq. (4) works quite well for video scenes with distinctive regions of interest, for example, for sensor network-based surveillance videos or head-and-shoulder videos for mobile social networks, etc.

3.3. Texture structure consistency (TSC) analysis

For motion–projection algorithms, the texture structure consistency between the matching blocks in key frames is also an indication of visual quality level of the corresponding SI macroblock. Higher structure consistency could imply better visual quality, even if the true error is high due to uniform shifting of object pixels. For the SI frame at time t, we calculate the correlation coefficient between the edge strength of the forward and backward motion compensated predictor images (i.e., the two hypotheses of SI) using Eq. (5):

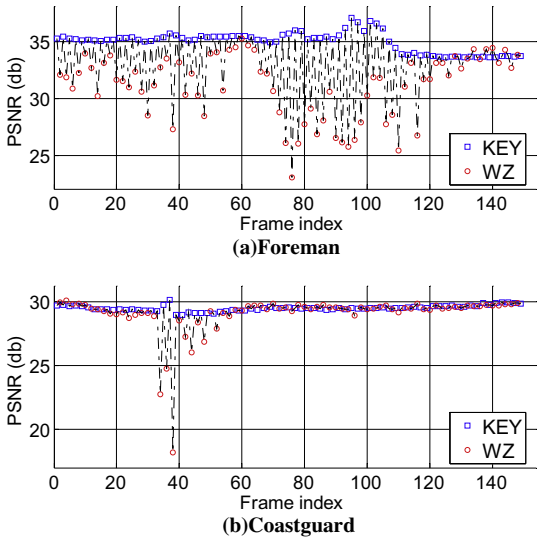


Fig. 10. The PSNRs of reconstructed frames of the Foreman (160 kbps) and Coastguard (100 kbps) sequences using the proposed DVC codec.

$$\rho_t(i) = \frac{\frac{1}{256}\sum_{\mathbf{p}\in\Omega_{MB(i)}}\left[(s_{f,t}(\mathbf{p}) - \mu_{f,t}(i)) \times (s_{b,t}(\mathbf{p}) - \mu_{b,t}(i))\right]}{\sqrt{\sigma^2_{f,t}(i) \times \sigma^2_{b,t}(i)}}. \tag{5}$$

Note that $s_{f,t}()$ and $s_{b,t}()$ are the edge strength images of the forward and backward hypotheses at time t, respectively. The edge strength is computed by the sum of the magnitudes of horizontal and vertical Sobel edge strength at each pixel position. $\Omega_{MB(i)}$ is the set of pixels of macroblock i. $\mu_{f,t}(i)$ and $\sigma^2_{f,t}(i)$ are the mean and variance of $s_{f,t}(\mathbf{p})$, $\mathbf{p}\ \Omega_{MB(i)}$, and $\mu_{b,t}(i)$ and $\sigma^2_{b,t}(i)$ are the mean and variance of $s_{b,t}(\mathbf{p})$, $\mathbf{p}\ \Omega_{MB(i)}$.

SI macroblocks whose $\rho_t(i)$'s are larger than a threshold θ_{TSC} belong to the set Ω_{TSC} of macroblocks with high texture structure consistency. The threshold θ_{TSC} is adaptively calculated by Eq. (6):

$$\theta_{TSC}(t) = \mu_{\rho_t} - \sigma_{\rho_t}, \tag{6}$$

where μ_{ρ_t} and σ_{ρ_t} are mean and standard deviation of the structure consistency measure, respectively. Note that if a pair of macroblocks do not contain any texture structures, comparing their structure correlation is meaningless. Therefore, macroblocks whose average edge strengths are below half of the average edge strength of current frame would be directly included into the set Ω_{TSC}.

Since the formulation of TSC is similar to SSIM [35], it might be possible to use SSIM to replace TSC and achieve similar effects. However, there are two key differences between TSC and SSIM. First, TSC is computed using the edge images, not the original pixels (as in SSIM). We have observed that most visual errors in reconstructed SI frames happen around edge pixels. In other words, TSC is a variant of SSIM that is fine-tuned to capture "texture similarity" around edge pixels (which makes the threshold θ_{TSC} less sensitive to lighting differences between key frames). The second key difference is about computational complexity.

Table 7
BD results of the test sequences.

QCIF				CIF			
Sequence	Δ Rate (%)	Δ PSNR	Δ SSIM	Sequence	Δ Rate (%)	Δ PSNR	Δ SSIM
Carphone	−18.6	0.93	0.021	Coastguard	−5.5	0.23	0.014
Coastguard	−10.5	0.47	0.019	Foreman	−7.8	0.28	0.011
Foreman	−11.3	0.59	0.016	Hall	−6.6	0.31	0.005
Hall	−10.6	0.83	0.008	News	−12.9	0.81	0.009

For SSIM, the computed variance, covariance, and mean images (a total of five images per key frame) are filtered by an 11×11 Gaussian filter. The complexity is quite high for our purposes. For TSC, we use only two 3×3 Sobel filters (horizontal and vertical) per key frame to compute structure correlation.

3.4. Valid motion projection (VMP) analysis

Motion–projection algorithms use neighboring key frames to predict SI frames. For boundary macroblocks, the projected motion vectors are often extrapolated from outside the frame boundaries, which can cause large visual distortions in SI frames. Therefore, we use the error $\varepsilon_t(\mathbf{p})$ between the matching key frame pixels that are projected to pixel $\mathbf{p}$ of SI at time t to determine whether a boundary macroblock i of SI have large errors:

$$\varepsilon_t(\mathbf{p}) = |I_{t-1}(\mathbf{p} - \mathbf{v}_f(\mathbf{p})) - I_{t+1}(\mathbf{p} - \mathbf{v}_b(\mathbf{p}))|, \tag{7}$$

where $I_t()$ is the image function at time t, $\mathbf{v}_f(\mathbf{p})$ and $\mathbf{v}_b(\mathbf{p})$ are the matching forward and backward motion vectors from the key frames to the SI pixel $\mathbf{p}$, respectively. A threshold, θ_ε, is calculated as:

$$\theta_\varepsilon(t) = \mu_{\varepsilon_t} + \sigma_{\varepsilon_t}, \tag{8}$$

where μ_{ε_t} and σ_{ε_t} are the mean and standard deviation of $\varepsilon_t()$ of all SI pixels at time t, respectively. In other words, an invalid motion projection pixel is defined as a pixel $\mathbf{p}$ whose motion vectors projected from key frame falls outside the frame boundary, and its error measure $\varepsilon_t(\mathbf{p})$ is larger than θ_ε. The set Ω_{VMP} of valid motion projection is defined as all SI macroblocks that contain less than $\theta_{VMP} = 10\%$ invalid motion projection pixels. Note that, most interior macroblocks belong to the set Ω_{VMP}, regardless of the magnitude of their error measure $\varepsilon_t()$. The selection of θ_{VMP} is based on empirical analysis. We have computed the percentage of invalid motion projection pixels of all SI macroblock of the test sequences. Setting θ_{VMP} to 10% is a strict threshold that eliminates all visual errors due to wrong boundary motion projection in all the test sequences. In fact, any value of θ_{VMP} below 15% should work fine for all the test sequences. However, as the threshold gets smaller, the coding efficiency may drop accordingly.

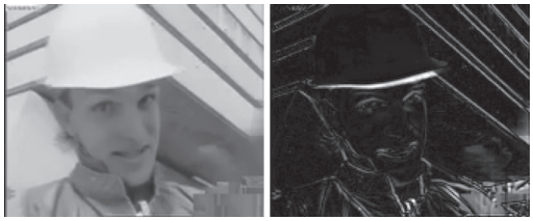


Fig. 12. The SI frame of the proposed codec (left) used in Fig. 11, and its error image (right). The SI frame of the DISCOVER codec is not available.

3.5. Determination of regions of interest (ROI)

The set Ω_{ROI} of macroblocks is composed of the macroblocks that have noticeable visual distortions. If a macroblock belongs to the intersection of the four sets Ω_{TDS}, Ω_{MC}, Ω_{TSC}, and Ω_{VMP}, we can consider this block as a macroblock that has little visual distortion and it does not require WZ reconstruction. Therefore, the initial set Ω'_{ROI} of visually distorted macroblocks is defined as in Eq. (9):

$$\Omega'_{ROI} = (\Omega_{TDS} \cap \Omega_{MC} \cap \Omega_{TSC} \cap \Omega_{VMP})^C, \tag{9}$$

where the superscript C denotes the complementary set.

Figs. 4 and 5 illustrate examples of the complementary sets of several frames in the Foreman sequence. Fig. 4(a)–(d) show Ω^C_{TDS}, Ω^C_{MC}, Ω^C_{TSC}, and Ω^C_{VMP} for different frames of Foreman where each metric capture unique visually distorted blocks. Fig. 5 shows how these metrics complement each other in a specific frame of the Foreman sequence. Note that VMP is not particularly useful in Fig. 5. It is designed to detect artifacts at boundary blocks. Thus, it is useful when there are camera-panning motions (as in Fig. 4(d)). The unions of these sets can capture almost all SI areas

Proposed: 35.7 dB
14.1 kbits
(SSIM=9.4, FSIM=9.6)

Proposed: 23.0 dB
2.1 kbits
(SSIM=7.4, FSIM=7.9)

Proposed: 36.0 dB
13.2 kbits
(SSIM=9.5, FSIM=9.6)

DISCOVER: 33.4 dB
10.4 kbits
(SSIM=9.2, FSIM=9.4)

DISCOVER: 29.5 dB
16.4 kbits
(SSIM=8.6, FSIM=9.0)

DISCOVER: 33.7 dB
9.4 kbits
(SSIM=9.2, FSIM=9.4)

Foreman, left-to-right: 75th (key), 76th (WZ), and 77th (key) frames

Fig. 11. Visual comparisons between the proposed codec (top row) and the DISCOVER codec (bottom row) at frame positions with highest PSNR variations. The bitrate of the proposed codec is 160.0 kbps, and the bitrate of DISCOVER is 161.6 kbps.

with noticeable visual distortions for video scenes with distinctive regions of interest used in our experiments.

Once the initial Ω'_{ROI} is obtained using Eq. (9), we further refine Ω'_{ROI} by removing isolated macroblocks and filling ROI holes. If a macroblock belongs to Ω'_{ROI}, but none of its neighbors belongs to Ω'_{ROI}, it is probably an isolated outlier (unless it is located at the frame boundary). On the other hand, if a macroblock is not in Ω'_{ROI} while the majority of its neighbors are, it probably should belong to Ω'_{ROI} too. To obtain the refined Ω_{ROI}, we first remove isolated non-boundary macroblocks in Ω'_{ROI} and then iteratively include a non-ROI macroblock into Ω'_{ROI} if the majority of its neighbors are in Ω'_{ROI}. The neighborhood structure is a diamond shape area around the macroblock under consideration as shown in Fig. 6. The iterative process continues until it converges. In the worse case, all macroblocks will be included into Ω_{ROI}, which means reliable detection of ROI is not possible and the codec falls back to full-frame WZ reconstruction. However, this situation never happens in our experiments.

Another observation is that, typically, 5–10% of macroblocks belong to the set Ω_{ROI}. However, at a scene change frame, the number of macroblocks in Ω_{ROI} would suddenly become large. Therefore, if the size of Ω_{ROI} is larger than a threshold θ_{SC} = 20% in a frame, full-frame WZ reconstruction would be used. Similar edge-based scene change detection methods have been proposed in [29,30]. Any value of θ_{SC} from 20% to 40% produces similar results in all the test sequences. If we occasionally mis-detect one of the frame as a scene change frame because the threshold is too low, there will not be any visual distortion. We simply suffer slightly on the coding gain. However, if we set the threshold too high, we may fail to detect some scene change frames and causes some visual artifacts. Thus, we set θ_{SC} to 20%. Fig. 7 shows some examples of macroblocks that belong to ROI.

3.6. Impact of large GOP sizes on the proposed perceptual metrics

We have been using the DVC coding structure with GOP size equals two due to the constraint of the SI generation algorithm used in this paper. The motion–projection SI generation algorithm assumes that the motions between two key frames are constant velocity translational motion. For most of the macroblocks, this assumption is valid when GOP size is small. However, as GOP size becomes large, most of the macroblocks will violate the constant velocity translational motion model. As a result, in addition to high SI prediction errors, there will be larger discrepancy between the predicted motion vectors and true motion vectors.

Fig. 14. The SI frame of the proposed codec (left) used in Fig. 13, and its error image (right). The SI frame of the DISCOVER codec is not available.

If the SI prediction error is high, the TDS metrics will include a large portion of macroblocks into the ROI since the original WZ and the predicted SI will have very different texture structure (even if the visual quality of the SI is good). Similar situation may happen to the TSC and VMP metrics because a larger portion of SI macroblocks will now be compensated from mismatched blocks due to translational motion constraint. Furthermore, because we try to use constant velocity translational motion to approximate nonlinear motions across a large period of time (i.e., large GOP), the estimated motion fields would become less regular. The proposed MC metric would also include more macroblocks into the ROI. In Table 1, we list the average percentage of macroblocks captured by each metric to illustrate the effect of GOP size increase on the Foreman sequence to demonstrate the impact of large GOP size on the proposed framework. The Foreman sequence has the most complex motion among four test sequences.

To solve the issue of larger GOP sizes, we have to adopt a more complex SI generation algorithm. For example, in [31], the initial SI

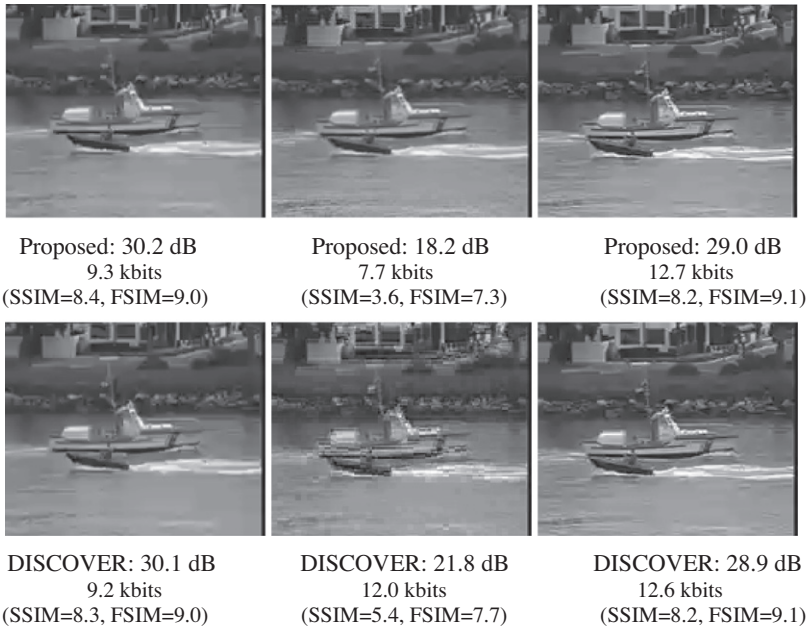


Fig. 13. Visual comparisons between the proposed codec (top row) and the DISCOVER codec (bottom row) at frame positions with highest PSNR variations. The bitrate of the proposed codec is 99.4 kbps, and the bitrate of the DISCOVER codec is 101.4 kbps.

generated using a first-order motion projection algorithm is used only for decoding of most significant bands (e.g., DC bands). The partially reconstructed WZ frame will then be used to help the second phase SI generation. Alternatively, one can use a higher-order motion–projection algorithm that takes into account more reconstructed frames (instead of simply two key frames) and tracks object trajectories for SI generation.

3.7. Complexity analysis of the proposed perceptual DVC codec

The proposed DVC framework has to perform extra computations in both the encoder and the decoder for perceptual analysis. Nevertheless, the overall complexity of the proposed decoder is often less than the complexity of the traditional DVC codecs. Note that the most time-consuming module in a DVC codec is the channel decoder (e.g., the LDPCA decoder in this paper [32]) and the SI generation algorithm. The proposed perceptual analysis technique allows us to perform only partial LDPCA decoding. This scheme reduces the decoder complexity significantly when the ROI is small. On the encoder side, although the complexity does increase slightly, it is negligible compared to the baseline implementation.

To quantify the computational complexity of the encoder and the decoder, we have tested the proposed DVC codec, the DISCOVER codec, the AVC intra codec, and the AVC zero-motion inter codec on an Intel Core2 3 GHz CPU with 4 GB RAM. The AVC codec used is JM 17.2 and the coding structure of the AVC zero-motion codec has GOP size 2 with a B frame between two I frames. The video sequence used is the FOREMAN sequence at 15 frames per second (a total of 149 frames). Table 2 shows the encoding time comparison while Table 3 shows the decoding time comparison.

As one can see from Tables 2 and 3, although the complexity of the proposed encoder is slightly higher than the complexity of the DISCOVER encoder (about 0.76% higher on average), the decoder complexity of the proposed codec is less than that of the DISCOVER codec (about 30.0% lower on average). The breakdown numbers of the execution time per frame of each module of the proposed codec for the same experimental setup are shown in Tables 4 and 5.

Fig. 16. The SI frame of the proposed codec (left) used in Fig. 15, and its error image (right). The SI frame of the DISCOVER codec is not available.

Although the overall complexity is usually lower for the proposed approach, its theoretical coding delay is indeed longer than that of the traditional DVC approaches. When the proposed encoder receives an original WZ frame, it has to wait until the decoder provides the ROI map before it can start LDPCA encoding. This delay is composed of two parts: the computation time of the ROI map and the transmission time of the map back to the encoder. The uncompressed ROI information is one bit per macroblock. For CIF@15fps, the time interval between two video frames is about 66 ms. With a feedback channel bandwidth of 20 kbps, it would take 19.8 ms to transmit 396 bits per frame back to the encoder. Such feedback bandwidth is not difficult for today's wireless access technology.

As for the coding delay caused by the computation of the ROI map, it includes the SI generation time plus the proposed perceptual analysis time. From Table 5, it is obvious that the SI generation time requires hardware acceleration in the decoder in order to fulfill real-time requirement. However, since the SI generation algorithm is very similar to the motion estimation algorithm of traditional video encoders, there are many hardware solutions available.

4. Experimental results

This section presents experiments to demonstrate the efficiency of the proposed framework. The experiments are composed of two

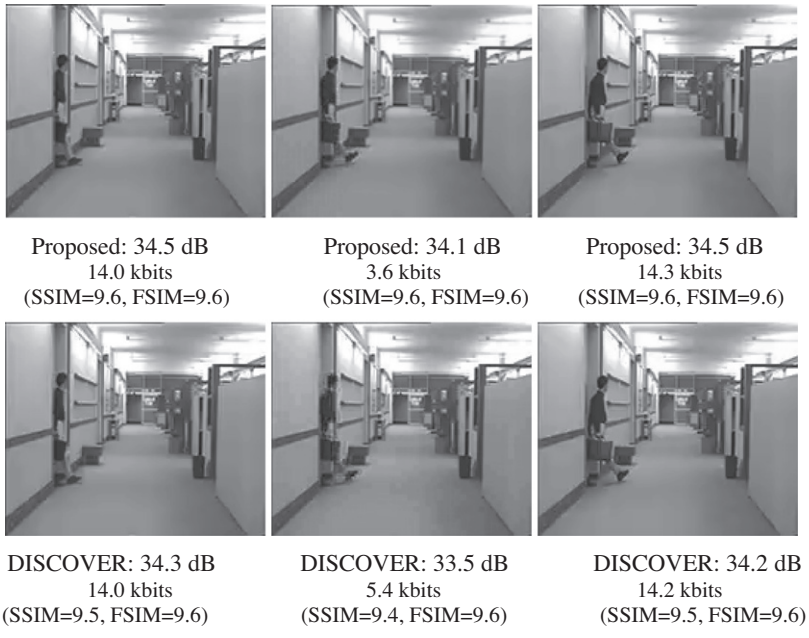


Fig. 15. Visual comparisons between the proposed codec (top row) and the DISCOVER codec (bottom row) at frame positions with noticeable visual improvements. The bitrate of the proposed codec is 127.8 kbps, and the bitrate of the DISCOVER codec is 131.5 kbps.

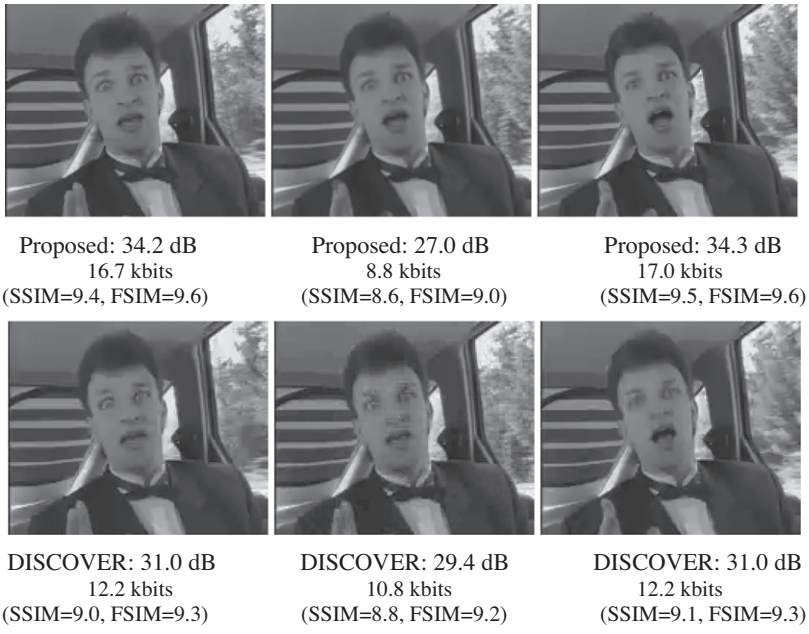


Fig. 17. Visual comparisons between the proposed codec (top row) and the DISCOVER codec (bottom row) at frame positions with noticeable visual improvements. The bitrate of the proposed codec is 134.1 kbps, and the bitrate of the DISCOVER codec is 134.4 kbps.

parts. The first part compares the R-D performance of the proposed framework with the DISCOVER codec [8], AVC/H.264 intra codec, and AVC/H.264 zero-motion inter codec. The GOP size of the DVC codecs and the AVC/H.264 zero-motion inter codec is two. The coding structure of the AVC/H.264 zero-motion inter codec is IBI. The DISCOVER codec is obtained from the DISCOVER website [33]. The second part of experiments provides video snapshots of consecutive frames for visual quality evaluations. Visual results of the proposed DVC codec and the DISCOVER codec are shown side-by-side for comparisons. The 15 Hz, QCIF version of four standard test sequences, (Foreman, Hall Monitor, Coastguard, and Car Phone), and the 30 Hz, CIF version of four standard test sequences, (Foreman, Hall Monitor, Coastguard, and News), are used in the experiments. The key frames are coded using an AVC/H.264 main profile intra coder (JM 17.2). Note that the proposed codec requires transmission of texture distribution bitmap and ROI bitmap in addition to the WZ bits. Both maps are represented using an uncompressed bitmap of 1 bit per macroblock (i.e., 99 bits per map for QCIF images). Although we can use Huffman codes to compress the maps by 68% on average, we do not think it is critical to do so for the proposed framework. All R-D curves of the proposed codec in this section include the data rates required for transmission of the extra information.

4.1. R-D performance evaluation

Figs. 8 and 9 show the R-D performance of different codecs using both PSNR metric and the perceptual metric SSIM [35]. For DVC codecs, each rate point assumes a constant key frame QP and a constant WZ frame quantization matrix QM. For the proposed codec, we have adopted the same (QP, QM) setting as in the DISCOVER codec, obtained by minimizing the PSNR variance of the sequence [10]. The (QP, QM) settings in the experiments are listed in Table 6. For the AVC/H.264 intra codec and the AVC/H.264 zero-motion inter codec, we choose the QP parameters from 29 to 43. For QCIF version of the sequences, the proposed DVC codec has better performance in all sequences comparing to the AVC-intra and the DISCOVER codecs. When compared against the AVC zero-motion inter codec, the proposed codec has better performance for Coastguard, and worse for Carphone, Foreman, and Hall Monitor. The main reason that the proposed perceptual-based codec out-performs DISCOVER in objective evaluation is that, with the proposed perceptual coding model, only a small portion of WZ bits is required to maintain consistent visual quality across the sequence. In other words, more data bits can be allocated to key frames to improve overall PSNR in the proposed scheme. As a result, the BD PSNR gain [37] over DISCOVER is 0.71 dB on average, and the SSIM gain is 0.016 on average.

For CIF version of the sequences, the result is similar. When compared against the AVC zero-motion inter codec, the proposed codec has better performance for Coastguard, slightly worse for Foreman, but worse for News and Hall Monitor. The overall PSNR gain over DISCOVER is 0.41 dB on average, and the SSIM gain is 0.010 on average. The BD Rate and BD PSNR results are listed in Table 7.

4.2. Visual quality comparisons

Since objective measures, such as the PSNR, do not fully reflect the visual quality of video sequences, subjective evaluation is often required for practical purposes. Therefore, in this section, we show some reconstructed frames from the proposed codec and the

Fig. 18. The SI frame of the proposed codec (left) used in Fig. 16, and its error image (right). The SI frame of the DISCOVER codec is not available.

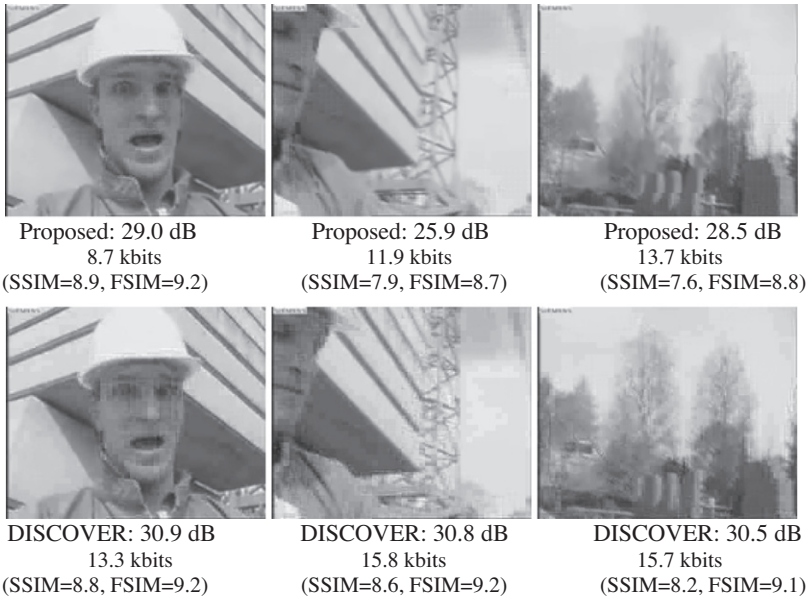


Fig. 19. Visual comparisons between the proposed codec (top row) and the DISCOVER codec (bottom row) at frame positions with poorest SI quality. The key frames for both codecs are the same. The target WZ rate for the corresponding frame is the same too. The bitrate of the proposed codec is 150.4 kbps, and the bitrate of the DISCOVER codec is 161.6 kbps.

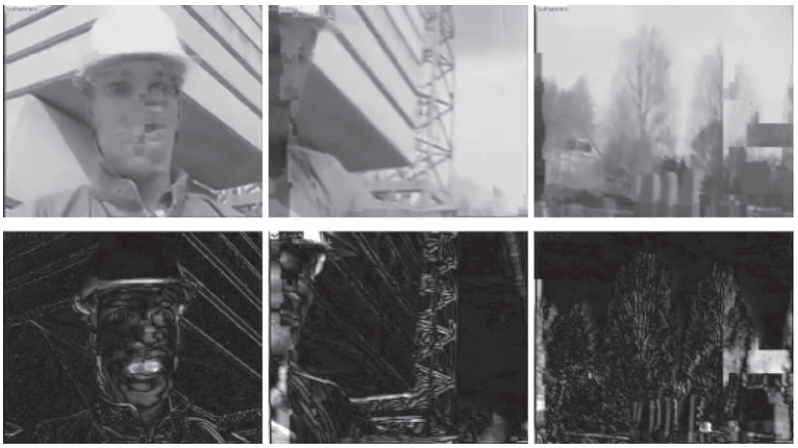


Fig. 20. The SI frames of the proposed codec (top row) used in Fig. 19, and their error images (bottom rows).

DISCOVER codec for visual quality evaluation. Since we have to match the target bitrate of a reconstructed sequence by the DISCOVER codec, we cannot use the same QP/QM selected using DISCOVER's algorithm. The reason is that, given same QP/QM, the encoded bit rate using the proposed codec would be much lower than that of DISCOVER's. Therefore, for each video sequence, we will use the rate ratio obtained using DISCOVER's QP/QM, and then find a set of finer QP/QM quantizers that maintains this rate ratio and produces an initial rate that is close to the target bitrate of DISCOVER's. To match the rate exactly, for each sequence, the following rate-allocation policy is applied. The bitrate of key frames are deducted from the target bitrate of DISCOVER's. The remaining bits are allocated to WZ frames. The target bit budget for each WZ frame is linearly proportional to the total sum of the errors $\varepsilon_t(\mathbf{p})$ in its ROI (see Eq. (7)). Note that such process is not a general policy for rate control of the proposed codec. Rate control of DVC codecs is a difficult problem [27,28,34]. We simply use the aforementioned process to match DISCOVER's bitrates for visual comparisons. To evaluate visual quality, in Figs. 11 and 13, we show the snapshots of consecutive frames where the proposed codec produces largest PSNR differences between key frames and the in-between WZ frame for the Foreman and Coastguard sequences. As a reference, the SI frame and its error image corresponding to the reconstructed WZ frame in Fig. 11 by the proposed algorithm are shown in Fig. 12. Similarly, the SI frame and its error image corresponding to the reconstructed WZ frames in Fig. 13 by the proposed algorithm are shown in Fig. 14. The PSNR values across frames of Foreman and Coastguard are shown in Fig. 10. It is quite evident from Figs. 11 and 13 that PSNR, as well as SSIM [35] and FSIM [36], do not precisely reflect visual quality. For Hall Monitor and Carphone, we show the snapshots where there are noticeable visual improvements between the proposed method and the DISCOVER codec in Figs. 15 and 17. The SI frame and its error image corresponding to the reconstructed WZ frame in Fig. 15 by the proposed algorithm are shown in Fig. 16. Finally, the SI frame and its error image corresponding to the reconstructed WZ frames in Fig. 17 by the proposed algorithm are shown in Fig. 18. When

the sequences are played back in real time, all four test sequences reconstructed by the proposed codec look sharper and have better visual quality than those reconstructed by the DISCOVER codec.

Since key frame quality has direct impact on the visual quality of the reconstructed DVC video, we also conduct another experiment where the same key frames are used for both the DISCOVER codec and the proposed codec. The key frames are encoded using the DISCOVER codec. In addition, same amount of WZ bits are used to correct WZ frames for each method. However, in the proposed codec, WZ bits are used to reconstruct only the ROI region. Some reconstructed WZ frames are shown in Fig. 19 and their SI frames and error images are shown in Fig. 20. In Fig. 19, snapshots are chosen at frame positions with poorest SI quality. It is clear that the visual quality of the ROI-only decoding is much better than that of the full-frame decoding method because the WZ bits are devoted to error corrections of the ROI areas where the visual errors are estimated to be large.

5. Conclusions

In this paper, we propose a perceptual-based WZ coding technique for DVC codecs. In the proposed framework, the decoder estimates the visual distortion levels of SI macroblocks and marks the macroblocks that require WZ reconstruction as the macroblocks in the region-of-interest (ROI). The ROI information is then transmitted back to the encoder using a bitmap so that parity bits can be generated to correct only these macroblocks. Experimental results show that the proposed perceptual-based coding technique improves coding efficiency of DVC both subjectively and objectively.

Although the proposed technique works well for video sequences that have distinctive regions of interest, it does not detect all the visually distorted regions for sequences with multiple complex moving objects. For example, for the QCIF Soccer sequence, the proposed technique misses 128 macroblocks (out of 7326) when adaptive thresholds of θ_{MC} and θ_{TSC} are used. Although we can select fixed threshold values such that all visually distorted blocks in Soccer are included into the ROI, the size of ROI will become large (contains 62% of macroblocks on average) and makes the proposed techniques less effective. More sophisticated visual distortion detection techniques will be investigated to deal with complex scenes such as the Soccer sequence.

Currently, the proposed technique has been tested using DVC coding structure of GOP size equals two. When GOP size becomes larger, both the SI prediction errors and the discrepancy between true and estimated motion fields will become large. As a result, the majority of the macroblocks will be included into the ROI. If larger GOP size is to be used, a more sophisticated SI generation algorithm has to be used to maintain efficiency of the proposed framework. For example, in current implementation, we only use the texture distribution map for perceptual-based analysis. It is possible to also use the texture map to constraint the motion–projection algorithm so that predicted SI and motion field are closer to the original WZ frame and true motion fields, respectively.

Finally, in the proposed framework, LDPCA is used for WZ reconstruction in the ROI. Since the size of the ROI is only composed of 20–30% of macroblocks in a frame, the LDPCA coding efficiency may suffer due to short coding block length. More detail analysis on the error characteristics of ROI macroblocks will be conducted in the future for the design of a more efficient WZ reconstruction algorithm.

References

[1] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, Distributed Video Coding Procss. 93 (1) (2005) 71–83. Special Issue on Video Coding and Delivery.

[2] D. Slepian, J.K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inform. Theor. 19 (1973) 471–480.

[3] A.D. Wyner, J. Ziv, The rate-distortion function for source coding with side information at the decoder, IEEE Trans. Inform. Theor. 22 (1976) 1–10.

[4] L. Liu, Z. Li, E.J. Delp, Efficient and low-complexity surveillance video compression using backward-channel aware Wyner–Ziv video coding, IEEE Trans. Circ. Syst. Video Technol. 19 (4) (2009).

[5] R. Puri, K. Ramchandran, PRISM: a reversed multimedia coding paradigm, in: Proceedings of International Conference on Image Processing, Barcelona, Spain, 2003.

[6] A. Aaron, S. Rane, E. Setton, B. Girod, Transform domain Wyner–Ziv codec for video, in: Proceedings of Visual Communications and Image Processing, San Jose, USA, 2004.

[7] A. Aaron, R. Zhang, B. Girod, Wyner–Ziv coding of motion video, in: Proceedings of Asilomar Conference on Signals and Systems, vol. 1, 2005.

[8] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, The DISCOVER codec: architecture, techniques and evaluation, in: Proceedings of Picture Coding Symposium, Lisbon, Portugal, 2007.

[9] D. Varodayan, A. Aaron, B. Girod, Rate-adaptive codes for distributed source coding, Signal Process. 86 (11) (2006) 3123–3130.

[10] J. Ascenso, F. Pereira, Advanced side information creation techniques and framework for Wyner–Ziv video coding, J. Vis. Commun. Image Represent. 19 (8) (2008) 600–613. Special Issue Resource-Aware Adaptive Video Streaming.

[11] Z. Li, L. Liu, E.J. Delp, Rate distortion analysis of motion side estimation in Wyner–Ziv video coding, IEEE Trans. Image Process. 16 (11) (2006) 98–113.

[12] D. Varodayan, D. Chen, M. Flierl, B. Girod, Wyner–Ziv coding of video with unsupervised motion vector learning, EURASIP Signal Process. Image Commun. 23 (5) (2008) 369–378. Special Issue on Distributed Video Coding.

[13] C. Dikici, T. Maugey, M.A. Agostini, O. Crave, Efficient frame interpolation for Wyner–Ziv video coding, in: SPIE Electronic Imaging, Visual Communications and Image Processing Conference (VCIP), San Jose, USA, 18–22 January, 2009.

[14] A. Aaron, B. Girod, Compression with side information using turbo codes, in: Proceedings of IEEE Data Compression Conference, Snowbird, UT, 2002, pp. 252–261.

[15] D. Kubasov, J. Nayak, C. Guillemot, Optimal reconstruction in Wyner–Ziv video coding with multiple side information, in: Proceedings of International Workshop on Multimedia Signal Processing, Crete, Greece, 2007.

[16] X. Huang, S. Forchhammer, Improved virtual channel noise model for transform domain Wyner–Ziv video coding, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.

[17] G. Hua, C.-W. Chen, Distributed video coding with zero motion skip and efficient DCT coefficient encoding, in: Proceedings of International Conference on Multimedia and Expo, Hannover, Germany, 2008.

[18] J.-C. Chiang, K.-L. Chen, C.-J. Chou, C.-M. Lee, W.-N. Lie, Block-based distributed video coding with variable block modes, in: Proceedings of International Symposium on Circuits and Systems, Paris, France, 2010.

[19] B.-R. Chiou, Y.-C. Shen, H.-P. Cheng, J.-L. Wu, Performance improvement of distributed video coding by using block mode selection, in: Proceedings of the 18th ACM International Conference on Multimedia, Firenze, Italy, 2010.

[20] Y.-C. Sun, C.-J. Tsai, Low complexity motion model analysis for distributed video coding, in: Proceedings of International Symposium on Multimedia over Wireless, Crete Island, Greece, 2008.

[21] J. Slowack, S. Mys, J. Skorupa, N. Deligiannis, P. Lambert, A. Munteanu, R. Van de Walle, Rate-distortion driven decoder-side bitplane mode decision for distributed video coding, Signal Process. Image Commun. 25 (9) (2010) 660–673.

[22] J. Slowack, S. Mys, J. Skorupa, P. Lambert, C. Grecos, R. Van de Walle, Accounting for quantization noise in online correction noise estimation for distributed video coding, in: Proceedings of Picture Coding Symposium, Chicago, IL, USA, 2009.

[23] V. Toto-Zarasoa, A. Roumy, C. Guillemot, Hidden Markov model for distributed video coding, in: Proceedings of International Conference on Image Processing, Hong Kong, 2010.

[24] Y.-C. Sun, S.-Y. Lian, C.-J. Tsai, Prioritized side information correction for distributed video coding, in: Proceedings of Picture Coding Symposium, Chicago, IL, USA, 2009.

[25] J.L. Barron, D.J. Fleet, S.S. Beauchemin, System and experiment performance of optical flow techniques, Int. J. Comput. Vis. (1994) 43–77.

[26] J.D. Areia, F. Pereira, W.A.C. Fernando, Impact of the key frames quality on the overall Wyner–Ziv video coding performance, in: Proceedings of 50th International Symposium on ELMAR, 2008.

[27] D. Kubasov, F. Lajnef, C. Guillemot, A hybrid encoder/decoder rate control for Wyner–Ziv video coding with a feedback channel, in: Proceedings of International Workshop on Multimedia Signal Processing, Crete, Greece, 2007.

[28] C. Brites, F. Pereira, Encoder rate control for transform domain Wyner–Ziv video coding, in: Proceedings of International Conference on Image Processing, San Antonio, Texas, 2007.

[29] R. Zabih, J. Miller, K. Mai, A feature-based algorithm for detecting and classifying production effects, Multimedia Syst. 7 (2) (1999) 119–128.

[30] H.S. Song, I.K. Kim, N.I. Cho, Scene change detection by feature extraction from strong edge blocks, in: Proceedings of Visual Communications and Image Processing, USA, 2002.

[31] R. Martins, C. Brites, J. Ascenso, F. Pereira, Statistical motion learning approach for improved transform domain Wyner–Ziv video coding, IET Image Process. J. 4 (1) (2010) 28–41.

[32] Y.-S. Pai, Y.-C. Shen, J.-L. Wu, High efficient distributed video coding with parallelized design for LDPCA decoding on CUDA based GPGPU, J. Vis. Commun. Image Represent. 23 (1) (2012) 63–74.
[33] DISCOVER codec. Available at: <http://www.discoverdvc.org>.
[34] S. Sofke, F. Pereira, E. Muller, Dynamic quality control for transform domain Wyner–Ziv video coding, EURASIP J. Image Video Process. 2009 (2009). Article ID: 978581.
[35] Z. Wang, A.C. Bovik, H.R. Sheikh, Image quality assessment: form error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[36] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[37] G. Bjontegaard, Improvement of the BD-PSNR model, VCEG document VCEG-AI11, ITU-T SG16/Q6, in: 35th VCEG Meeting, July 2008.