COMPLEXITY-RATE-DISTORTION ANALYSIS
OF BACKWARD CHANNEL AWARE WYNER-ZIV VIDEO CODING

Limin Liu †, Zhen Li † ‡, and Edward J. Delp †

†Video and Image Processing Lab (*VIPER*)
School of ECE, Purdue Univ., West Lafayette, IN 47907

‡Corporate Research, Thomson
Princeton, NJ 08540

ABSTRACT

Many Wyner-Ziv video coding (WZVC) schemes encode a video sequence into two types of frames, key frames and Wyner-Ziv frames. We have previously presented a Wyner-Ziv video coding scheme that uses backward channel aware motion estimation to encode the key frames, where motion estimation was performed at the decoder and motion information was sent back to the encoder. We refer to these backward predictively coded frames as BP frames. In this paper, we extend our previous work and propose three types of motion estimators. A model is presented to examine the complexity-rate-distortion performance of BP frames for the three motion estimators.

Index Terms— Wyner-Ziv Video Coding (WZVC), Backward Channel Aware Motion Estimation (BCAME)

1. INTRODUCTION

The main characteristic of Wyner-Ziv video coding (WZVC) is that side information is available only at the decoder [1, 2]. Sensor networks and mobile video are some of the applications where WZVC may be very useful [3, 4]. These applications are characterized by the need for reduced encoding complexity due to the limited availability of resources (e.g. battery life and memory).

Many WZVC schemes encode a video sequence into two types of frames, key frames and Wyner-Ziv frames. Key frames are encoded using conventional video coding methods such as H.264 and Wyner-Ziv frames are encoded using channel coding techniques [1, 5]. At the decoder, the reconstructed key frames serve as side information used to reconstruct the Wyner-Ziv frames.

We have previously presented a WZVC scheme that uses backward channel aware motion estimation (BCAME) to encode the key frames [6], where motion estimation was performed at the decoder and motion information was sent back to the encoder. We refer to these backward predictively coded frames as BP frames. BP frames generally have lower coding efficiency than standard predictively coded frames (P frames) and also have much lower computational complexity. As shown in [2, 7], the coding efficiency of a Wyner-Ziv video codec significantly depends on the coding efficiency of the key frames. This leads to the need for complexity-rate-distortion (CRD) optimization for applications where the encoder is subject to limited computational resources.

Complexity constrained rate distortion analysis and optimization for video coding has been of interest in the research community. In [8], CRD analysis for a motion-compensated prediction (MCP) based video encoding is developed by modeling the complexity and rate-distortion tradeoff for different encoding parameters. Another MCP-based CRD analysis is proposed in [9] by considering the complexity of different types of mode selection. In [10], CRD analysis is proposed for a wavelet based video encoder by modeling the complexity for different spatiotemporal decomposition structures.

This work was supported by a grant from the Indiana 21^{st} Century Research and Technology Fund. Address all correspondence to E. J. Delp at ace@ecn.purdue.edu.

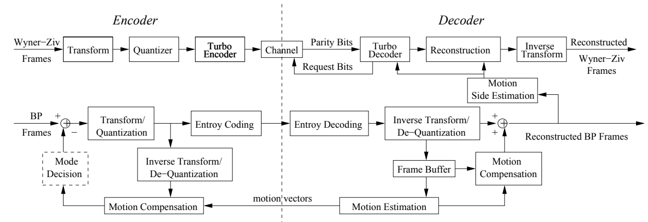


Fig. 1. Backward channel aware Wyner-Ziv video coding.

In this paper we present a model to quantitatively analyze the complexity and rate-distortion tradeoff for BP frames. Three estimators, the minimum estimator, the median estimator and the average estimator, are proposed and analyzed.

2. BACKWARD CHANNEL AWARE WZVC

WZVC with Backward channel aware motion estimation (BCAME) was proposed in [6]. The coding diagram is shown in Fig. 1. The basic idea of BCAME is to perform motion estimation at the decoder and send the motion information back to the encoder through a backward channel. The motion information is used to encode the key frames at the encoder. The frames coded in this way are referred to BP frames. Wyner-Ziv frames are coded in the transform domain using channel codes, such as turbo codes or low-density-parity-check (LDPC) codes. Only part of the parity bits are sent to the decoder. At the decoder, the side information is derived from the previously decoded key frames. The derived side information is used as the initial estimate to decode the Wyner-Ziv frames with the parity bits received from the encoder. The derivation of side information generally involves motion estimation and the motion information can be used for BCAME, hence the additional computational complexity at the decoder is marginal.

A BCAME encoder can generate the motion compensated frames using the motion vectors received by the backward channel. If several motion vectors are sent to the encoder, the encoder can make a mode decision based on the distortion and choose the best available motion vector. For BP frames, the residual frame between the original frame and motion compensated reference frame is encoded in the same way as an H.264 encoder. Compared to the INTRA coded frames, BP frames can significantly improve the coding efficiency with minimal usage of the backward channel. Compared to the P frames, BP frames moves the computationally intensive task of motion estimation to the decoder and results in a reduction of complexity at the encoder.

The computational complexity of a Wyner-Ziv video encoder includes that of Wyner-Ziv frames and BP frames. Since the complexity of Wyner-Ziv frames are relatively insignificant, we will focus on the analysis of the BP frames. The computational complexity of BP frames can vary depending upon the number of candidate motion vectors received at the encoder. When N motion vectors are received, the encoder will need to compare the distortions resulted from these N motion vectors. In terms of computational complexity, this is comparable to a motion search when there are only N candidate motion vectors in the search area, whereas the traditional P frames require a motion search for every pixel and sub-pixel inside its search window.

3. CRD ANALYSIS OF BP FRAMES

3.1. Problem Formulation

Assume there are N motion vectors estimated at the decoder. Denote these N motion vectors as $MV_1, MV_2, \cdots, MV_N$. We assume that the N motion vectors are 2-D independent and identically distributed (i.i.d.) random variables having the joint probability density function (pdf) $f(x, y)$ and the cumulative probability distribution function (cdf) $F(x, y)$. Denote the true motion vector as $MV_T = (x_n, y_n)$. A true motion vector is an ideal motion vector with minimum mean squared error between the original frame and the reference frame.

As shown in [2, 7] the rate difference between two coders using different motion vectors is,

$$\Delta R_{1,2} = \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\frac{1-e^{-\frac{1}{2}\omega^T\omega\sigma^2_{\Delta_1}}}{1-e^{-\frac{1}{2}\omega^T\omega\sigma^2_{\Delta_2}}}d\omega \quad (1)$$

where $\sigma^2_{\Delta_i}$ $(i = 1, 2)$ are the variances of the error motion vectors. The error motion vector is the difference between the derived motion vector and the true motion vector. In the following we analyze three different motion estimators, namely, the minimum motion estimator, the medium motion estimator, and the average motion estimator.

3.2. The Minimum Motion Estimator

In this case, all N i.i.d. motion vectors are sent to the encoder. At the encoder, each of the N motion vectors is applied to the reference frame. The motion vector leading to the minimum distortion between the original frame and motion compensated reference frame is selected. As in many traditional fast motion search methods, we assume the motion field is homogeneous and unimodal, then the motion search is equivalent to choosing the motion vector nearest to the true motion vector. Hence the corresponding error motion vector is $(X - x_n, Y - y_n)$, where X and Y are the horizontal and vertical motion vectors. We use capital letters to denote a random variable unless otherwise specified.

We introduce a new random variable Z to model the distance between the received motion vector and the true motion vector,

$$Z = \sqrt{(X - x_n)^2 + (Y - y_n)^2} \quad (2)$$

Hence the problem can be formulated as searching for the motion vector with the smallest Z.

This problem can be solved using order statistics and extreme value theory [11]. More specifically, let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables. Denote $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ as the corresponding order statistics, where the first order statistic $X_{(1)}$ is the minimum of $X_1, X_2, \cdots, X_n$. The probability density function of the kth order statistic can be formulated as [11]

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}F(x)^{k-1}[1 - F(x)]^{n-k}f(x) \quad (3)$$

We now consider the case when the received motion vectors have a 2-D Gaussian distribution,

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2}e^{-\frac{(x-x'_n)^2+(y-y'_n)^2}{2\sigma^2}} \quad (4)$$

where σ^2 is the motion vector variance, x'_n and y'_n are the mean of the horizontal and vertical motion vectors respectively. To facilitate our discussion, we denote the deviation from the true motion vector as $\delta_x = x'_n - x_n$ and $\delta_y = y'_n - y_n$.

3.2.1. *Case I: $\delta_x \neq 0$ or $\delta_y \neq 0$*

In this case the means of the motion vectors sent back from the decoder are different from the true motion vector. The random variable Z as defined in (2) is a Rician distribution,

$$f_Z(z) = \frac{z}{\sigma^2}\exp(-\frac{z^2+\nu^2}{2\sigma^2})I_0(\frac{\nu x}{\sigma^2}) \quad z \geq 0 \quad (5)$$

where $I_0(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{t\cos\theta}d\theta$ is a modified Bessel function and $\nu^2 = \delta_x^2 + \delta_y^2$.

The distribution of the first order statistic of Z, or the smallest Z, can be derived from (3) with $k = 1$.

The variance of the error motion vectors with N motion vectors sent back from the decoder is

$$\sigma^2_{\Delta_N} = \int_0^{\infty} z^2 f_{Z_{(1)}}(z)dz = \int_0^{\infty} z^2 N[1 - F_Z(z)]^{N-1}f_Z(z)dz \quad (6)$$

where $f_Z(z)$ is derived in (5) and $F_Z(z)$ is the cumulative probability distribution function of $f_Z(z)$. From (1), the rate difference in using N motion vectors compared to using only one motion vector

$$\Delta R_{1,N} = \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\frac{1-e^{-\frac{1}{2}\omega^T\omega\sigma^2_{\Delta_N}}}{1-e^{-\frac{1}{2}\omega^T\omega\sigma^2_{\Delta_1}}}d\omega \quad (7)$$

The numerical rate difference with $\nu = 1$ and $\nu = \frac{1}{2}$ are shown in Fig. 2-(a) and (b) respectively for various σ^2. Fig. 2 shows that sending more motion vectors through the backward channel can improve the coding efficiency of BP frames. For example, in the case of $\nu = 1, \sigma^2 = 1$, compared to only sending one motion vector to the encoder, sending five motion vectors can lead to a rate saving of 0.2276 bit/sample, or an improvement of roughly $0.2276 \times 6.02 = 1.3702$ dB in peak signal-to-noise ratio (PSNR). It is noted that this rate saving is more significant when σ^2 is smaller. When σ^2 is large (such as $\sigma^2 = 10$), sending more motion vectors to the encoder has very limited impact on rate savings. In other words, when there is fast or large irregular motion in a video sequence, sending more motion vectors is not justified. We also note that when ν is smaller, i.e, when the motion vectors extracted at the decoder from the previous reconstructed frames are closer to the true motion vector on average, the rate saving is more significant.

3.2.2. *Case II: $\delta_x = 0$ and $\delta_y = 0$*

In this case, the mean of the received motion vector is identical to the true motion vector. This is a special case of (5) with $\nu = 0$. And the distribution of Z as defined in (2) is a Rayleigh distribution with

$$f_Z(z) = \frac{z}{\sigma^2}e^{-\frac{z^2}{2\sigma^2}}, \quad F_Z(z) = 1 - e^{-\frac{z^2}{2\sigma^2}}, \quad z \geq 0 \quad (8)$$

The first order statistic and the variance of the error motion vectors

$$f_{Z_{(1)}}(z) = N[1 - F_Z(z)]^{N-1}f(z) = N\frac{z}{\sigma^2}(e^{-z^2/2\sigma^2})^N$$

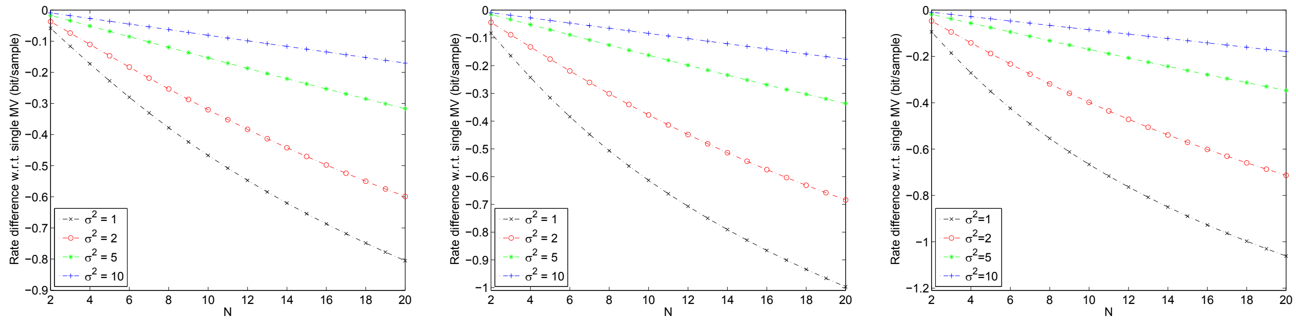
$$\sigma^2_{\Delta_N} = E[Z^2_{(1)}] = \int_0^{\infty} z^2 N\frac{z}{\sigma^2}(e^{-z^2/2\sigma^2})^N dz = \frac{2}{N}\sigma^2 \quad (9)$$

The rate difference with respect to the single motion vector is

$$\Delta R_{1,N} = \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\frac{1-e^{-\frac{1}{2}\omega^T\omega(\frac{2}{N}\sigma^2)}}{1-e^{-\frac{1}{2}\omega^T\omega(2\sigma^2)}}d\omega \quad (10)$$

The results are shown in Fig. 2-(c). Compared to Fig. 2-(a) and (b), the rate saving is higher than Case I when $\nu \neq 0$.

In terms of computational complexity, the motion search complexity of the minimum motion estimator is linear with the number N of the motion vectors sent back from the decoder. Since the encoder complexity in this case depends largely on the the motion search complexity, (7) and (10) describe not only the rate-distortion tradeoff but also the complexity-rate-distortion tradeoff for the minimum estimator. Furthermore, we note that the backward channel bandwidth usage can also assume to be linear with the number N, hence (7) and (10) also describe the tradeoff between the rate distortion performance and the backward channel bandwidth usage.



(a) $\delta_x = \delta_y = \frac{\sqrt{2}}{2}$ $(\nu = 1)$ (b) $\delta_x = \delta_y = \frac{\sqrt{2}}{4}$ $(\nu = \frac{1}{2})$ (c) $\delta_x = \delta_y = 0$ $(\nu = 0)$

Fig. 2. Rate difference of the minimum motion estimator for different values of N.

3.3. The Median Motion Estimator

The minimum motion estimator in Section 3.2 requires that N motion vectors should be sent through the backward channel and the encoder then needs to make a motion search to find the best candidate motion vector. This leads to higher requirement on encoding complexity and backward channel bandwidth. When the application cannot satisfy this requirement, a motion vector needs to be selected at the decoder among the N motion vector candidates and only this motion vector will be sent to the encoder.

One way to choose such a motion vector is to use the median motion vector among the N motion vectors. The median motion vector is constructed using the median of the horizontal motion vectors and the median of the vertical motion vectors. Without loss of generality, we assume N is odd, i.e., $N = 2l + 1$, and the $m = (l + 1)$th order statistic is used. Consider the Gaussian case in (4), the horizontal and vertical motion vectors are independent with

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x'_n)^2}{2\sigma^2}}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-y'_n)^2}{2\sigma^2}} \tag{11}$$

Since X and Y follow the same distributions, we will discuss X only and the performance of Y can be analyzed similarly.

The cumulative probability distribution function of the Gaussian distribution in (11) can be expressed as

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-x'_n)^2/2\sigma^2} dt = 1 - \frac{1}{2}\text{erfc}(\frac{x - x'_n}{\sqrt{2}\sigma}) \tag{12}$$

where the error function $\text{erfc}(x)$ is defined as $\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$. Since we are interested in the distance to the true motion vector x_n, we define a new variable $X_d = X - x_n$; the distribution of X_d and the variance of the error motion vector $E[X_d^2]$, according to (3), are

$$\begin{aligned} f_{X_d}(x) &= \frac{N!}{(m-1)!(N-m)!}[1 - \frac{1}{2}\text{erfc}(\frac{x - \delta_x}{\sqrt{2}\sigma})]^{m-1} \\ &\times [\frac{1}{2}\text{erfc}(\frac{x - \delta_x}{\sqrt{2}\sigma})]^{N-m} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\delta_x)^2}{2\sigma^2}} \end{aligned} \tag{13}$$

$$E[X_d^2] = \int_{-\infty}^{\infty} x^2 f_{X_d}(x) dx \tag{14}$$

Since X and Y are independent, the total variance of the error motion vector $\sigma_\Delta^2 = E[X_d^2] + E[Y_d^2]$. Table 1 gives a summary of $\sigma_{\Delta_N}^2$ for the case $\nu = 0, \sigma^2 = 1$. The median motion estimator yields larger variance of the error motion vectors than the minimum estimator.

The rate difference with respect to the single motion vector is shown in Fig. 3. Comparing the results of the minimum motion estimator and the median motion estimator, the rate saving of the median motion estimator is smaller than that of the minimum motion estimator. For example, when $\nu = 1, \sigma^2 = 1$ and $N = 5$, using the minimum estimator can achieve a rate saving of 0.2276 bit/sample, while using the median estimator can only achieve a rate saving of 0.0570 bit/sample. It is interesting to note that in Fig. 3-(a), when N is large ($N \geq 12$), the rate saving when $\sigma^2 = 2$ is actually greater than the $\sigma^2 = 1$ case. This is because when the motion vector error is closer to the motion vector variance, while the rate saving will continue to improve with the increase of N, such improvement will grow at a slower pace. 3-(c) shows the rate difference for the case $\delta_x = \delta_y = 0$. In this case, the rate saving is generally more significant than the case $\nu > 0$.

N	1	3	5	7	9
minimum	2.000	0.667	0.400	0.286	0.222
median	2.000	0.897	0.574	0.421	0.332

Table 1. The variance of the error motion vectors ($\nu = 0, \sigma^2 = 1$)

3.4. The Average Motion Estimator

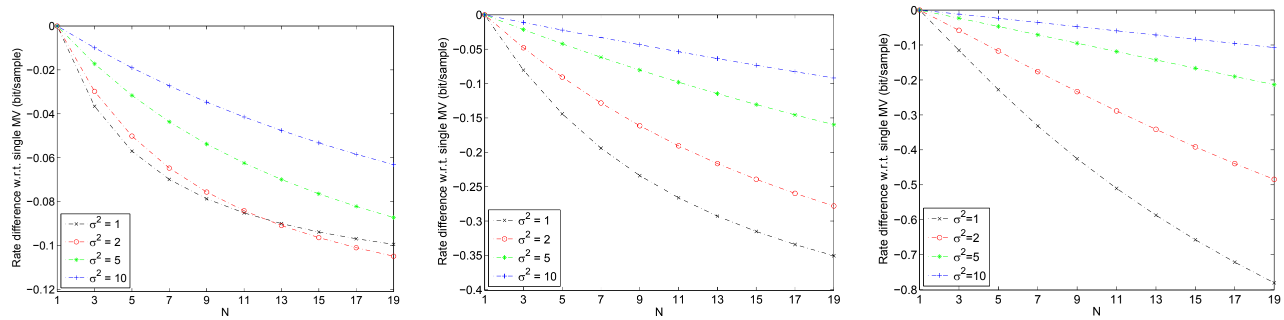
Another way to choose the motion vector candidates at the decoder is to send the average of the motion vectors, which is referred to as the average motion estimator. With the N motion vectors available at the decoder, we send the average motion vector $\bar{X} = \frac{1}{N}(X_1 + X_2 + \cdots + X_N)$ and $\bar{Y} = \frac{1}{N}(Y_1 + Y_2 + \cdots + Y_N)$ to the encoder. The sample average of a Gaussian distribution is also Gaussian,so

$$\bar{X} \sim N(x'_n, \frac{\sigma^2}{N}), \qquad \bar{Y} \sim N(y'_n, \frac{\sigma^2}{N}) \tag{15}$$

The variance of the error motion vector is

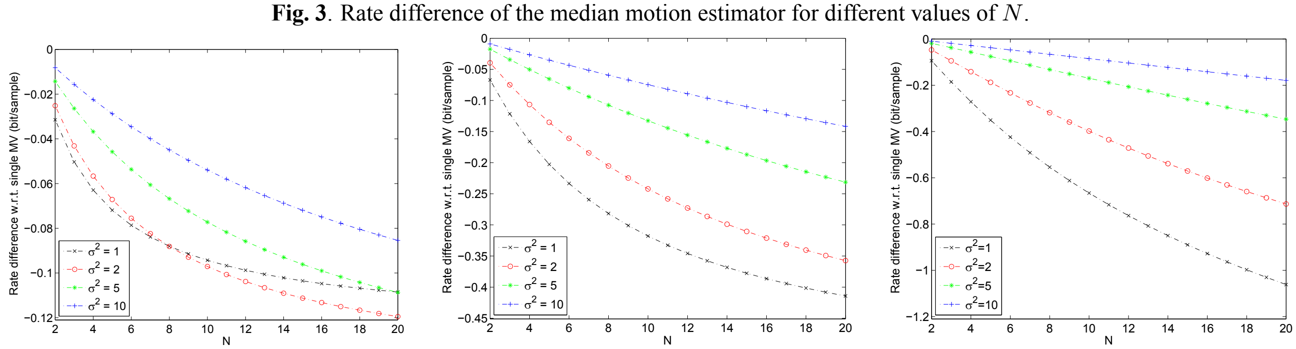
$$\begin{aligned} \sigma_{\Delta_N}^2 &= E[(\bar{X} - x_n)^2] + E[(\bar{Y} - y_n)^2] \\ &= (\delta_x^2 + \frac{\sigma^2}{N}) + (\delta_y^2 + \frac{\sigma^2}{N}) = \nu^2 + 2\frac{\sigma^2}{N} \end{aligned} \tag{16}$$

The result of the variance is shown in Table 2 with comparisons to the minimum and median motion estimators ($\nu = 1, \sigma^2 = 1$). The variance of the error motion vectors using the average motion estimator is higher than the minimum motion estimator. Compared with the median motion vector, the variance using the average motion estimator is lower. The rate difference with respect to the single motion vector using the average motion estimator is shown in Fig. 4. With the same requirement of the encoder complexity and the bandwidth of the backward channel, the average motion estimator gives better rate distortion performance than the median motion estimator. Similar to Fig. 3-(a) and for the same reason mentioned in Section 3.3, we can observe a similar *crossover* effect in Fig. 4-(a) between the curves of $\sigma^2 = 1$ and $\sigma^2 = 2$ when $N \geq 8$. In the case $\nu = 0$, (16) reduces to $\sigma_{\Delta_N}^2 = 2\frac{\sigma^2}{N}$, which is identical to (9). In other words, when $\nu = 0$, the average motion estimator performs as well as the minimum motion estimator. While when $\nu \neq 0$, the minimum motion estimator can achieve higher coding efficiency at the cost of higher encoding complexity and more backward channel bandwidth.



(a) $\delta_x = \delta_y = \frac{\sqrt{2}}{2}$ $(\nu = 1)$ (b) $\delta_x = \delta_y = \frac{\sqrt{2}}{4}$ $(\nu = \frac{1}{2})$ (c) $\delta_x = \delta_y = 0$ $(\nu = 0)$

Fig. 3. Rate difference of the median motion estimator for different values of N.



(a) $\delta_x = \delta_y = \frac{\sqrt{2}}{2}$ $(\nu = 1)$ (b) $\delta_x = \delta_y = \frac{\sqrt{2}}{4}$ $(\nu = \frac{1}{2})$ (c) $\delta_x = \delta_y = 0$ $(\nu = 0)$

Fig. 4. Rate difference of the average motion estimator for different values of N.

N	1	3	5	7	9
minimum	3.000	1.052	0.641	0.461	0.360
median	3.000	1.897	1.574	1.421	1.332
average	3.000	1.667	1.400	1.286	1.222

Table 2. The variance of the error motion vectors ($\nu = 1, \sigma^2 = 1$).

4. CONCLUSIONS

In this paper, we presented a model used to study the complexity-rate-distortion coding efficiency of backward channel aware WZVC. The results show the rate-distortion performance of the average estimator is generally higher than that of the median estimator. If the rate-distortion tradeoff is the only concern, the minimum estimator yields better results than the other two estimators. However, for applications with complexity constraints, our analysis shows that the average estimator could be a better choice. The model presented in this paper quantitatively describes the complexity-rate-distortion tradeoff among these estimators. For future work, we will investigate the use of our model to examine the impact of INTRA-coded frames and INTER-coded frames. Based on the CRD model, we feel we can optimize the coding structure with a complexity constraint.

5. REFERENCES

[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE: Special Issue on Advances in Video Coding and Delivery*, vol. 93, pp. 71–83, January 2005.

[2] Z. Li, L. Liu, and E. J. Delp, "Rate distortion analysis of motion side estimation in Wyner-Ziv video coding," *IEEE Transactions on Image Processing*, vol. 16, pp. 98–113, January 2007.

[3] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, pp. 80– 94, september 2004.

[4] R. Puri, A. Majumbar, P. Ishwar, and K. Ramchandran, "Distributed video coding in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, pp. 94–106, July 2006.

[5] L. Liu and E. J. Delp, "Wyner-Ziv video coding using LDPC codes," in *Proceedings of the IEEE 7th Nordic Signal Processing Symposium (NORSIG 2006)*, Reykjavik, Iceland, June 2005.

[6] L. Liu, Z. Li, and E. J. Delp, "Backward channel aware Wyner-Ziv video coding," in *Proceedings of the IEEE International Conference on Image Processing*, Atlanta, GA, October 2006.

[7] Z. Li and E. J. Delp, "Wyner-Ziv video side estimator: conventional motion search methods revisited," in *Proceedings of the IEEE International Conference on Image Processing*, Genova, Italy, September 2005, vol. 1, pp. 825 – 828.

[8] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 645– 658, May 2005.

[9] H. F Ates, B. Kanberoglu, and Y. Altunbasak, "Rate-distortion and complexity joint optimization for fast motion estimatin in H.264 video coding," in *Proceedings of the IEEE International Conference on Image Processing*, Atlanta, GA, October 2006.

[10] D. S. Turaga, M. van der Schaar, and B. Pesquet-Popescu, "Complexity scalable motion compensated wavelet video encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 982– 993, August 2005.

[11] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.