# An Estimation Optimization Algorithm of Side Information Based on General Prediction

Min Zhou

School of Intelligence Science and Information Engineering, Peihua University, Xian 710100, China

77440264@qq.com

**Abstract.** The coding efficiency of the Wyner-Ziv video decoder is largely dependent on the quality of the side information obtained in the decoder. The reason why constructing effective side information is difficult is partly caused by unavailability of the original video sequence in the decoder. The traditional motion search method is used to obtain side information, which greatly increases the complexity of Wyner-Ziv video decoding. This paper proposes a new method of extracting side information based on the general prediction idea. This new method is called Wyner-Ziv Video Coding (WZUP) with general prediction. This method does not conduct motion search in the decoder or presuppose a model for the original input video sequence. Instead, side information is estimated based on observations of video data that have been reconstructed in the past. Through testing, this method can greatly reduce the decoder's decoding complexity and achieve good side information estimation performance. This makes it possible to design video encoders and decoders with low computation complexity.

**Keywords:** general prediction, Wyner-Ziv video coding, side information estimation.

## 1. Introduction

Traditional video coding standards, such as MPEG and H.26x, all use asymmetric encoded mode, the main steps of the encoder include transformation, quantization, entropy coding, corresponding decoding process, motion estimation and motion compensation. Therefore, the complexity of the encoder is much higher than that of the decoder; in particular, motion estimation and motion compensation occupy large amounts of resources, the complexity of the encoder is very high. Distributed video coding is a new coding method that is different from traditional coding. Among them, Wyner-Ziv theory is the main theoretical basis of distributed video coding technology. Unlike traditional video coding based on motion compensation prediction (MCP), in Wyner-ziv video coding, side information can only be used them in decoder and not in the encoder side. In video coding based on MCP, side information is analyzed in the encoder to reduce redundancy in the input video sequence. Under normal conditions, this usually need include a motion search process with high computation strength; therefore, it is not suitable for applications that require simple encoders, such as video telephony applications or video surveillance system applications. On the contrary, Wyner-ziv video coding transfers side information analysis to the decoder and attempts to maintain coding performance comparable to coding based on MCP.

The conceptual structure of the source coding can only be used in the decoder for side information, as shown in Figure 1. Two related sources x and y are encoded by two separate encoders a and b, respectively; each of them cannot access another information source. If there is an encoder and decoder for any arbitrary and code rate R, we say that the code rate R can be achieved. If joint decoding is permitted, then the Slepian-Wolf theorem proves that the system in Figure 1, the achievable rate range is:

$$R_X \geq H(X|Y), R_Y \geq H(Y|X), R_X + R_Y \geq H(X,Y) \quad (1)$$

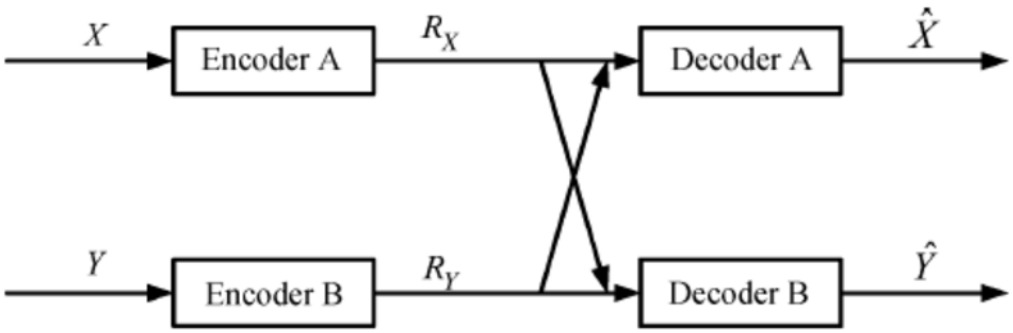


Fig.1 slepian—wolf coding

Therefore, regardless of whether the encoder A can access the side information Y, as long as the decoder A can access Y and the code rate is greater than or equal to H (X/Y), then encoder A can encode X at any high fidelity.

Wyner and Ziv extended this result to the case of lossy compression. Setting R*(d) denote the rate-distortion function when the side information Y is only available in the decoder, and $R_{x|y}(d)$ denote the rate-distortion function when the side information Y is simultaneously available in the encoder and the decoder. Wyner and Ziv prove that, in spite of $R^*(d) \geq R_{X|Y}(d)$, but in certain specific cases, the equation holds, for example, a case where Gaussian sources takes mean-square error as case of distortion measure. Therefore, similar to undistorted coding, under the condition of lossy compression, the side information of the encoder side is not always necessary to achieve the lower limit distortion rate.

Based on this idea, wyner-ziv video coding only utilizes the video source's statistical information in the decoder. In video coding based on MCP, each frame is decoded and reconstructed in the encoder, it is then saved in the frame buffer and used as a reference frame for the next frame encoding. In the decoder, each decoded frame is also saved in the frame buffer and used as a reference frame for decoding the next frame. As long as the same reconstructed frames are saved in the frame buffers of the encoder and the decoder and the motion vectors are correctly passed to the decoder, both the encoder and the decoder can be guaranteed to obtain the same reference information. The reference information can be used as side information when encoding the next frame.

However, it is very difficult to construct reference information without the original video sequence. Existing Wyner-Ziv video decoders usually rely on traditional motion estimators to extract motion information from video frame reconstructed in the decoder. In order to do this point, it is necessary to presuppose a motion model, although this assumption is true for some video sequences, the motion in the natural video sequence cannot be well defined, and a simple model may not be enough. Another drawback of this method is that the video decoder is very complex. Because the task of motion estimation is transferred to the decoder, Wyner-Ziv video coding requires high complexity in the decoder.

In this paper, we propose a new reconstruction method of side information; this method does not rely on a pre-specified video sequence model. This method is called as WZUP; it utilizes the source statistical properties of the reconstructed video sequence and does not presuppose the model of the input sequence. The goal is to construct a video codec that has both low and high complexity in the encoder and decoder, while maintaining good coding efficiency. The rest of the paper is organized as follows: in the second part, we give a side estimation method with general prediction. 2.1 first introduces the concept of general prediction, 2.2 discusses the side decoding process of video side decoder, 2.3 describes the specific algorithm of side estimator when using general prediction

In the third part, we give the test results and discussion. The fourth part is the conclusion of this paper.

## 2. General Predicted W-Z Video Coding

Here we first give an introduction to general prediction and then give a new side estimator when we using general prediction.

### 2.1 General Prediction

The idea of general prediction originated from the practice of predicting the next output of sequence. Can the future of the sequence be predicted by its past. If so, how good can this prediction be.These problems often appear in many applications. Obviously, there is often such a link here, and if it is known in advance, then it may be very useful for prediction. However, in practice, the knowlside or intrinsic model of this connection is usually difficult to obtain or inaccurate, this requires the development of general prediction method, roughly, the general prediction is such a prediction, it does not depend on the unknown intrinsic mode, but still can achieve the same good performance as the model known in advance.

The general prediction problem can be described as follows. Considering that an observer receives a data sequence $x_1, x_2, ..., x_{t-1}$, he wants to predict the next output $x_t$, so that make it obey the loss I of definition in the prediction result $\hat{x}_t$ and the true result $x_t$. If the data source's intrinsic statistical model is known and the predicted goals are well defined, then typical statistical prediction theories can be applied, including maximum likelihood prediction, maximum posterior prediction, and Wiener estimation theory. In these cases, we assume that the data is generated by information source with a statistical model P. However, if the intrinsic statistical characteristics of the information source are unknown, most natural video sequences belong to this situation, and the predictive solution cannot be well defined as in the previous case. Under such circumstance, a general prediction algorithm tries to estimate the intrinsic statistical properties of the data based on observations of past data.

In this paper, we consider the reconstructed video data in the wz video decoder as an observation value, and the decoder tries to predict the result of the next frame without knowing the statistical mechanism of generated video source. Then this prediction will be used as an initial estimation of the wz video decoder (side information).

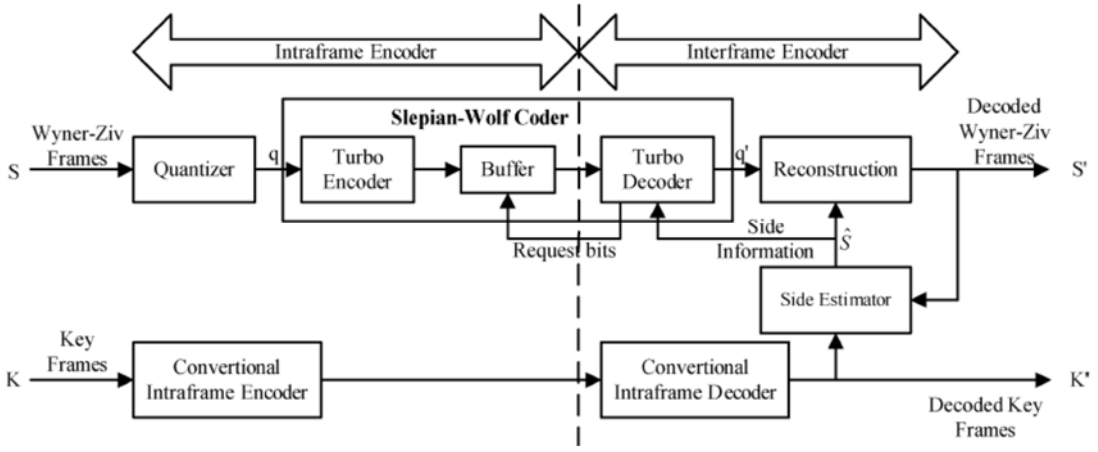


Fig.2 Wyner-Ziv Video Coding

### 2.2 Wz Video Encoding

WZ video encoder as shown in Fig.2, in the encoder side, each frame is either encoded as an INTRA frame or encoded as a WZ frame. INTRA frames are coded in the traditional H264 intra mode. WZ frame coding is conducted in the spatial domain. After all pixels are quantized, these quantized pixels are encoded one by one bit plane by the turbo encoder. Check bits are saved in the encoder side and transmitted to the decoder after the decoder sends a request. After receiving the check, the decoder first constructs an initial estimation for each current decoded frame and then begins decoding. This initial estimation is constructed by side estimator. The turbo decoder uses this initial estimate

and the check bits transmitted from the encoder to decode this frame. The decoder may need to obtain more check bits from the encoder until a predetermined decoding accuracy is reached.

### 2.3 Use Side Estimator of General Prediction

We propose a new side estimator based on general prediction formula. Each video frame is considered as a vector, the pixel values in the same spatial coordinate position are grouped into groups I(k,l), here (k,l) is the spatial coordinates in the video frame. Without loss of generality, consider one such group I(k,l), denoted as X $x = x_1, x_2, \ldots, x_{t-1}, x_t, ,$ here i in the $x_i$ denotes time order in the sequence. At the same time, the loss function $\Lambda : [0, M] \times [0, M] - > [0, M^2]$ is denoted as $\Lambda = \Lambda(i, j), i, j \in [0, M]$ , here $\Delta(i, j)$ denote loss by the pixel value j and estimate the loss of pixel value j. $\lambda_k$ is used to denote the $k$ list in $\Lambda$ , and M is the permitted maximum pixel value, which is set to 255. $Z = z_1, z_2, \ldots, z_{t-1}, z_t$ is used to show the reconstruction of the decoder $\hat{I}$(k，l). $z_t$ is an initial guess for the current reconstruction output. Because this initial guess is arbitrary and often unreliable, the side estimator tries to provide a more accurate estimation for $x_t$ . Therefore, side information estimation can be seen as a noise reduction problem, which can be solved by the method in 29.

$\prod = \prod(i, j)_{i, j \in [0.M]}$ is denoted as transmission matrix from X to Z, $\prod(i, j)$ is the probability when the i is input at in x and the reconstruction in the corresponding z is j.

$P_{X_t|Z}$ is used to denote the conditional probability of $x_t$ , best estimation of $\mathrm{x_t}$ is the value when the loss expectation is minimized, namely:

$$\begin{aligned} &\hat{X}^{\mathrm{opt}}(\mathrm{z})[\mathrm{t}] \\ &= \arg \min_{\hat{\mathrm{x}} \in [0,\ M]} \sum\nolimits_{\alpha \in [0,M]} \Delta(\alpha, \hat{\mathrm{x}}) P\left(\mathrm{x_t} = \alpha \middle| Z = z\right) \\ &= \arg \min_{\hat{x} \in [0,M]} \lambda_{\hat{x}}^T P_{X_t|z} = \arg \min_{\hat{x} \in [0,M]} \lambda_{\hat{\mathrm{x}}}^{\mathrm{T}} P_{X_{\mathrm{t},\ z}} \end{aligned} \tag{2}$$

In the formula,

$$P_{X_{\mathrm{t},\ z}} = P_{X_{\mathrm{t}|z}} \bullet P(Z = \mathrm{z}) \tag{3}$$

If we know $\mathrm{x_t}$ and reconstruct the joint distribution $P_{X_{\mathrm{t}}|z}$ of the context z, it is easy to get the optimal estimate by the Lagrange optimization and solving the root. But because we assume that the statistical knowledge of the video sequence model in the decoder is not available, this probability distribution is not available. Therefore, we need to find a good estimation of p. Because

$$\begin{aligned} &P\left(X_{\mathrm{t}} = x_t, Z_t = z_t, Z^{n\backslash t} = z^{n\backslash t}\right) \\ &= P\left(X_t = x_t, Z^{n\backslash t} = z^{n\backslash t}\right) \bullet \prod\left(x_t, z_t\right) \end{aligned} \tag{4}$$

Here $z^{n\backslash z} = \mathrm{z}_1, \mathrm{z}_2 \cdots, \mathrm{z}_{t-1}$. Using vector to denote, we get:

$$P_{X_{t,z}} = \pi_{z_t} \odot P_{X_t, z^{n\backslash t}} \tag{5}$$

Here $(u \odot v)[i] = u_i v_i$ , $X_{\mathrm{t}}$ in the formula (5) is eliminated by finding the marginal probability, and all possible $\mathrm{z_t} \in [0, \mathrm{M}]$ cycle are obtained.

$$P_{Z_{\mathrm{t}}, z^{n\backslash t}} = \prod^T P_{X_{t, z^{n\backslash t}}} \tag{6}$$

So,

$$P_{X_{t,z^{n\backslash t}}} = \pi_{z_t} \odot \left[ \prod^{-T} P_{Z_{t,\ z^{n\backslash t}}} \right] \tag{7}$$

The optimal estimate at this time is:

$$\hat{X}^{opt}\left(z^n\right)[t] = \arg\min_{\hat{x}\in[0,M]} \lambda_{\hat{x}}^T \pi_{z_t} \odot \left[ \prod^{-T} P_{Z_{t,z^{n\backslash t}}} \right]$$

$$= \arg\min_{\hat{x}\in[0,M]} \left[ P_{Z_{t,z^{n\backslash t}}} \right]^T \prod^{-1} \left[ \lambda_{\hat{x}} \odot \pi_{z_t} \right] \tag{8}$$

We now consider the case where the root mean square error is used to measure the distortion, namely

$$\Delta(i,j) = (i,j) **2 \tag{9}$$

And consider the simplest case, namely, the transmission probability $\pi\left(x_t, z_t\right) = 1$ when $x_t = z_t$ , and the transmission probability is 0 when $x_t \neq z_t$. Under this condition, the optimal estimator in formula (8) is the minimum mean square error estimator (MMSE). Using the Lagrange optimization, the MMSE estimator is obtained by formula 8:

$$\hat{x}^{opt}\left(z^n\right)[t] = \left[ P_{Z_{t,z^{n\backslash t}}} \right]^T \odot Z_t \tag{10}$$

Namely, the optimal estimator is a previously appearing (output) weighted average under the same scenario. The weighting factor is determined by the number of events that occur.

## 3. Test Results

Here, we evaluate the performance of wzup, we compare it with the wz video encoding of the side estimator based on MCP.

The side estimation in wzup are built based on (10). For each pixel that is decoded in the decoder, we first collect its context in the previous frame. This context is the pixel value when the previous N frames in the same spatial coordinates. In the test, we first need to consider how many frames to predict the current frame. Considering less prediction of the current frame may not achieve better prediction result, and use too many current frames will increase the decoding complexity. The two factors are synthesized, we set N=5, five frames is used to predict the next frame. Assume that the sequence model is IPIPIPIPIPIP.... It is obvious that odd frames are key frames and do not need to be predicted, Even frames need to be predicted. The optimal estimator is a weighted average of the previously occurring (output) in the same scenario. Its weighting factor is determined by the number of events that occur. We can count the number of occurrences of the same pixel value on the same spatial coordinate of five frames; consider its occurrence probability to calculate its weight as an estimation of the next frame. We only evaluate the PSNR (Power Signal-Noise Ratio) of the reference frame and do not evaluate the final coding efficiency. We use akiyo and salesmen two sequences to do the experimental and obtain PSNR, the results are shown in Figure 3-4.

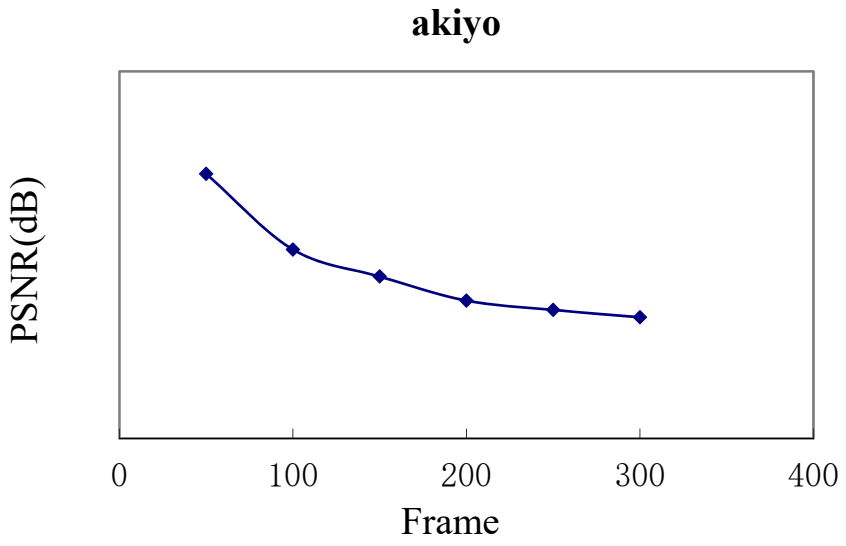


Fig.3 akiyo sequences

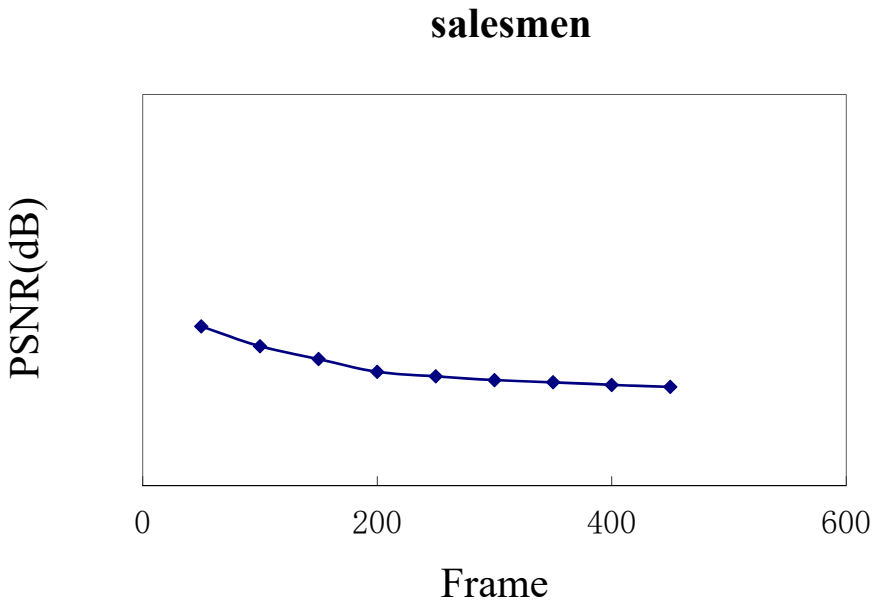


Fig.4 salesmen sequences

## 4. Conclusion

This paper proposes a new side estimator for Wyner-Ziv video coding. It constructs the original estimation in the Wyner-Ziv decoder by using non-motion search. This Wyner-Ziv video coding based on general prediction does not rely on a preset video sequence model, and only utilizes the reconstructed video sequence statistical properties. Our test results show that it can achieve good coding efficiency. Moreover, this new method can greatly reduce the decoding complexity of the decoder and save storage space. This makes it possible to design a codec with low computational complexity at both the encoder and decoder.

## References

[1]. Zhen Li, Limin Liu ,Edward J. Delp. Wyner–Ziv Video Coding With Universal Prediction[J]. IEEE Transactions on Circuits and Systems for Video Technology. 2006, 11, 16, (11).

[2]. N. Merhav and M. Feder, Universal prediction[J]. IEEE Trans. Inf.Theory, vol. 44, no. 6, pp. 2124–2147, Oct. 1998.

[3]. Wang Yumin, Liang Chuanjia, "Information and Coding Theory".

[4]. A. Aaron, S. Rane, and B. Girod, Wyner-Ziv video coding with hash based motion compensation at the receiver [J]. Proc. IEEE Int. Conf.Image Process., Singapore, Oct.