Resolution-progressive Compression of Encrypted Grayscale Images

Wei Liu, Wenjun Zeng, Lina Dong and Qiuming Yao

Computer Science Department, University of Missouri, Columbia MO 65211, USA

ABSTRACT

Compression of encrypted data can be achieved by employing Slepian-Wolf coding (SWC). However, how to efficiently exploit the source dependency in an encrypted colored signal such as an image remains a challenging issue. Previous works incorporate 2-D Markov models in the SWC, which is not accurate enough for natural grayscale images; as a result, the compression performance is usually poor. In this paper, we propose to compress the image progressively, such that the decoder can observe a low-resolution version of the image, from which local statistics is learned and used for the decoding of the next resolution level. Good performance is observed both theoretically and experimentally.

Index Terms—Compression of encrypted data, Slepian-Wolf coding, resolution-progressive compression.

1. INTRODUCTION

The conventional method to distribute redundant data over an insecure and bandwidth-limited channel is to perform compression before the encryption. However, recent work by Johnson et al. [1] shows that it is possible to reverse the order, and neither the compression performance nor the security will be sacrificed under certain reasonable conditions. For example, suppose the plain text $X$ is an i.i.d. source, and the encryption function is a stream cipher: $Y = X \oplus K$ (1)

where $\oplus$ denotes the bit-wise exclusive OR (XOR) operation, $K$ is the key stream, and $Y$ is the cipher text. In this case, if the decoder knows $K$, a joint decoding-decryption function can be carried out, and the bit rate required to reconstruct $Y$ is

$$H(Y|K) = H(X|K) = H(X)$$ (2)

where $H(\cdot)$ denotes the Shannon entropy.

Slepian-Wolf coding (SWC) [2] plays an important role here. For lossless compression of $Y$, according to the Slepian-Wolf theorem, one can achieve the theoretical bound at $R_Y = H(Y|K) = H(X)$ with vanishing error probability for long sequences, even if the encoder does not know the value of $K$. In other words, the encoder can compress $Y$ to the rate of $H(X)$ by employing SWC; the decoder treats $K$ as the side information (SI) and performs Slepian-Wolf decoding (when decryption is performed jointly [1]).

In the literature, very good Slepian-Wolf codes have been found based on modern channel codes (the reader is referred to [3] for a comprehensive review). Most of them are designed for i.i.d. sources though. For colored signals, a Markov model is usually incorporated in the SWC, e.g. the approach in [4]. Previous works on compression of encrypted binary images [5][6] also employ a 2-D Markov model in the SWC. However, for natural grayscale images, the 2-D Markov model is too simple for the following reasons. First, the 2-D model presented in [6] consists of a horizontal 1st-order Markov chain and a vertical 1st-order Markov chain, and the two chains are processed separately. However, without joint consideration of the 2-D neighboring pixels, geometric features such as edges and textures can not be observed and utilized. Second, real-world image data are usually highly non-stationary, and a global Markov model fails to adapt to the local statistics. As reported in [6], no significant coding gain is observed except for the first two most significant bit-planes (MSB). It is also suggested in [6] that incorporating more neighbors in the Markov model could help. However, maintaining a high-order Markov model is extremely expensive in terms of both memory and computation. How to effectively exploit the correlation of encrypted image data remains a challenging issue.

In this paper, we propose an efficient way to compress encrypted grayscale images through resolution-progressive compression. The encoder starts by sending a downsampled version of the cipher text. At the decoder, the corresponding low-resolution image is decoded and decrypted, from which a higher-resolution image is obtained by interpolation. The interpolated image, together with the secret encryption key, is used as the SI to decode the next resolution level. This process is iterated until the whole image is decoded. By doing so, the task of de-correlating the pixels, which is not possible for the encoder, is shifted to the decoder side. By accessing the low-resolution image, the decoder is able to learn the local statistics, doing much better than “blind” decoding. Moreover, by avoiding exploiting the Markovian property in the SWC, the complexity is significantly reduced.

The rest of the paper is organized as follows. Section 2 describes the proposed scheme and analyzes its performance on ideal sources. In Section 3, a context-adaptive interpolator is presented for SI generation, and a localized channel estimator is also introduced. Section 4 presents the simulation results and Section 5 concludes the paper.

2. RESOLUTION-PROGRESSIVE COMPRESSION

2.1. System Description

The encoder gets the cipher text $Y$ and decomposes it into several levels. For single-level decomposition, four subbands are generated, as illustrated in Figure 1. Each subband is a shifted and downsampled-by-2 version of $Y$, and $Y$ can be losslessly synthesized from the four subbands. In fact, it is equivalent to perform a 2-D analysis filter bank on $Y$, with $H_d(z) = 1, H(z) = z$. (The corresponding synthesis filter bank is $G_d(z) = 1, H(z) = z^{-1}$.) Hence, some notations are borrowed from conventional DWT. We call the four subbands as the LL, HL, LH and HH subbands, and if multiple-level decomposition is applied, the labels LL$_n$, HL$_n$, LH$_n$ and HH$_n$ are used for the subbands in the $n$-th level. The $(n+1)$-th level decomposition is performed on the LL$_n$ subband. An example of 3-level decomposition of the Lena image is illustrated in Figure 2.
A scaling factor of

\[ \sigma = \pi \omega \] is used as the SI to decode HH

in one level is to improve the qua lity of the SI. Another reason is

known subband(s) for the interpolati on, as illustrated in Figure 1.

the inefficiency of channel codes in achieving the

caused by the inefficiency in removing the redundancy among the pixels.

In this subsection we focus on th e second type of loss. We will

dependency of

source coding loss

\[ \Phi_\omega(\omega) = \frac{2\pi \sigma_\omega^2}{(\omega^2 + \omega^2)^{3/2}} \] (5)

where \( \omega \) is the vector of spatial frequency (in radius). At high rates,

the optimum encoder reaches the R-D bound:

\[ R_{\omega}(D) = \frac{1}{8\pi} \int_{-\infty}^{\infty} \frac{\Phi_\omega(\omega)}{D} d\omega \] (6)

where \( D \) is the distortion level (suppose \( D \) is small).

In our proposed scheme, the rate saving actually comes from the

inter-subband interpolation, because the variance of the residual

is expected to be smaller than the original signal. According to the linear prediction theory [8], if \( t = [t_1, t_2, t_3, t_4] \) is used to predict (interpolate) \( s \), the optimum prediction produces the

residual variance as

\[ E[(s - \hat{s})^2] = \sigma_s^2 - r_{tt}^2 R_{rs} r_{rs} \] (8)

where \( R_{rs} = E[tt^T] \), \( r_{tt} = E[tt^T] \). In different subbands, we can estimate the element values of \( R_{rs} \) and \( r_{tt} \) from Figure 1. Therefore, through some simple calculations, we can get the variances of the interpolation residuals in level \( n \):

\[ \begin{align*}
\sigma_{\omega_{tt}}^2(HL_t) &= \sigma_{\omega_{tt}}^2(LH_t) = F(2n-1)\sigma_s^2, \\
\sigma_{\omega_{tt}}^2(HH_t) &= F(2n)\sigma_s^2 
\end{align*} \] (9)

where the facilitating function \( F(k) \) is defined as

\[ F(k) = 1 + 2p\rho^k - 3p\rho^{2k} + 1 + 2p\rho^k + \rho^{2k}. \] (10)

Thus the overall R-D function for the proposed scheme can be derived by summing-up the bit rates in all subbands:

\[ R_{\omega}(D) = \sum_{n=1}^{\infty} \frac{1}{2\pi} \sum_{(n,i,j) \in \{(0,0),(0,1),(1,0),(1,1)\}} \frac{1}{D} \log_2 \frac{\sigma_{\omega_{tt}}^2(n,i,j)}{D}. \] (11)
Here the factor $1/2^n$ accounts for the fact that the number of pixels in each subband in level $n$ is $1/2^n$ of the full-scale image, and we assume there are infinite resolution levels.

From (6), (7) and (11) we obtain the rate saving over the memoryless coding by using our proposed scheme

$$R_{opt} = R_{opt}(D) - R_{ml}(D) = \sum_{k=1}^{\infty} \frac{1}{2^k} \log_2 F(k).$$

and by using the optimal coding

$$\Delta R_{opt} = R_{opt}(D) - R_{ml}(D) = \frac{1}{8\pi^2} \int_0^\infty \int_0^\infty 2\pi \omega \left( \frac{\omega + \omega'}{\omega^2 + \omega'^2} \right) \frac{d\omega}{d\omega'} = \frac{1}{8\pi^2} \int_0^\infty \left( \frac{1}{\omega^2 + \omega'^2} \right) d\omega.'$$

We can see that both $\Delta R_{opt}$ and $\Delta R_{intp}$ are functions of $\rho$ (or equivalently, functions of $\alpha$). In Figure 3 $\Delta R_{opt}$ and $\Delta R_{intp}$ are compared with respect to different $\rho$ values.

In the previous section, it is assumed that the image signal is wide-sense stationary. In this case, the interpolation coefficients are constant throughout the subband. However, for real-world images, it is desired to have the coefficients adapt to local context. For example, for a pixel on an edge, it is preferred to interpolate along the edge orientation. Similar efforts can be found in conventional lossless image compression, where the median edge detector (MED) [7] and the gradient adaptive predictor (GAP) [9] are two successful context adaptive predictors. However, they process the pixels in a raster-scanning order, thus cannot be directly applied to our scheme.

In this section, a simple, yet effective context adaptive interpolator (CAI) is proposed for our scheme. For example, let’s consider the case using the LL and HL subbands to interpolate the HL subband (the horizontal-vertical interpolator illustrated in Figure 1). The interpolator classifies the local region into 4 types: smooth, horizontally-edged, vertically-edged and other. In smooth regions, a mean filter is applied; in edged regions, the interpolation is done along the edge; otherwise we use a median filter. More specifically, the proposed CAI is formulated as

$$\hat{s} = \begin{cases} 
\text{mean}(t) & \text{(max}(t) - \text{min}(t) \leq 20) \\
\left( t_i + t_j \right)/2 & \left( |t_i - t_j| - |t_l - t_r| \geq 20 \right) \\
\left( t_i + t_j \right)/2 & \left( |t_i - t_j| - |t_l - t_r| > 20 \right) \\
\text{median}(t) & \text{otherwise}
\end{cases}$$

In (14) it can be verified that the first condition contradicts the second (or the third) condition, thus a “smooth” region will never be estimated as “edged” again. The second and the third conditions are adapted from GAP, with an ad hoc threshold. It is also possible that the region is diagonally-edged, but there is no clue about on which side of the edge $s$ lies. Therefore we simply adopt a median filter in this case.

The diagonal interpolator in Figure 1 is the same as (14).

Besides the knowledge of SI, a Slepian-Wolf decoder also needs to know the statistics of the channel between each pixel and its SI. Similar to [10], we assume a Laplacian distribution of $s$, centered at the given SI:

$$p(s|\hat{s}) = \frac{\exp(-\alpha |s - \hat{s}|)}{2\sigma_\hat{s}^2}$$

where $\alpha = \sqrt{2/\sigma_\hat{s}^2}$, and $\sigma_\hat{s}^2$ is the variance of $p(s|\hat{s})$. Hence it is necessary for the decoder to estimate $\sigma_\hat{s}^2$.

In this paper, $\sigma_\hat{s}^2$ is calculated from the prediction residual of the previously decoded level. For example, if $s$ is a pixel in the HL subband, the decoder observes several geometrical neighbors of $s$ in the HL+1 subband. In this paper, the neighborhood is chosen to be a 5x5 window. The mean square error of the CAI results for these pixels is scaled to be used as $\sigma_\hat{s}^2$. The scaling is needed because for interpolation at higher-resolution levels, the correlation between the neighboring pixels is higher, which usually means smaller prediction residual. From (9) we can see, the scaling factor should be $F(2n)/F(2n+2)$ for the HH4 subband, or be $F(2n-1)/F(2n+1)$ for the HL and the LH subbands. Numerical results show that $F(k)/F(k+2)$ typically ranges from 0.5 to 1, depending on the $\rho$ value. In this paper, we simply adopt an empirical scaling factor at 0.75. More sophisticated modeling might further improve the coding performance.

It can be seen that, the access to a lower-resolution image enables the decoder to exploit the geometric features in CAI, and to learn the local statistics in channel estimation.

## 4. SIMULATION RESULTS

The images listed in Table I are used for testing. The encoder decomposes each encrypted image into 4 resolution levels. The subbands in the lowest-resolution level are sent without compression. But the decoder still performs inter-subband interpolation on them. The results will be used to estimate the p.d.f. of the pixels in the next level. For the other subbands, we transmit the four LSBs as raw bits, because there is not much gain to employ SWC on them. The four LSBs are sent prior to the MSBs.

<table>
<thead>
<tr>
<th>Image</th>
<th>Bobor</th>
<th>Lena</th>
<th>Peppers</th>
<th>Boats</th>
<th>Goldhill</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.81</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table I. Correlation between neighboring pixels
such that the decoder can have better knowledge about the pixels before starting decoding the MSBs. The four MSBs, on the other hand, are Slepian-Wolf encoded using rate-compatible punctured turbo codes [11] in a bit-plane based fashion. The sending rate of each Slepian-Wolf coded bit-plane is determined by the decoder’s feedback.

Firstly, we focus on the performance of the inter-subband interpolation. The Shannon entropy of the residual is used as the criterion, and the results of MED, GAP and CAI are listed in Table II. For CAI, what we are interested in is the conditional entropy $H(s \mid \hat{s})$. However, we will calculate $H(s - \hat{s})$ as an upper bound, where $[\cdot]$ denotes rounding to the nearest integer. From Table II we can see that CAI provides similar or better performance than MED and GAP.

Secondly, let’s study the coding performance of the resolution-progressive compression (RPC) scheme. In RPC, we estimate the local variance of prediction residual using the method described in Section 3. In RPC, we assume the decoder knows the global variance of the prediction residual in each subband (although this is practically not possible). The results are shown in Table III, from which we can see, RPC outperforms RPC by 0.1 to 0.2 bpp. This means localized channel estimation is more effective than a global one, and this localization is enabled by the resolution progressive compression.

There is no numerical results reported in [6] for grayscale images. The authors just mention that significant compression is only observed in the first 2 MSBs. As a comparison, the average code length for each bit-plane (across different subbands) for “Lena” is plotted in Figure 4. Significant compression is observed in the first 4 MSBs, which is superior to the result reported in [6]. In Figure 4, the entropy for each bit-plane is also plotted. The difference between the entropy and the actual coding rate is the source coding loss, which is summed up to 0.29 bpp for the Lena image.

For comparison purpose, we also list the actual coding rate of CALIC, a very good lossless image codec. The gap between RPC and CALIC is 0.60 bpp on average, which consists of both the source coding loss (0.31 bpp on average) and the image coding loss. The image coding loss is because the codec is still not able to remove the redundancy as efficiently as CALIC. For example, CALIC uses bias cancellation and run length coding to further deal with the redundancy among the prediction residual. These techniques can not be directly applied to our scheme.

### 5. CONCLUSIONS

In this paper, we focus on the design of a practical image codec, where the image data undergoes the stream cipher based encryption before the compression. We propose to compress the image progressively in resolution so that the decoder can have partial access to the image, from which the geometric features and the local source statistics are learned and used for the decoding of the next resolution level. Theoretical analysis shows that, despite the inefficiency of channel codes, our scheme achieves 70% to 90% rate saving of that of the optimal conventional intra-frame coder. The proposed practical system shows better coding performance than the previous approach, which exploits a 2-D Markov model in the SWC.

### ACKNOWLEDGEMENT

This research is supported in part by the Giliom Cyber Security Fellowship, and in part by NSF grant CNS-0423386.

### REFERENCES


