

# Probability Estimation in Arithmetic Coding and Its Application

Feifei Zhou, Rui Yang, Bo Li

Digital Media Laboratory,  
School of Computer Science and Engineering,  
Beihang University, Beijing 100083, China  
Email: zhoufei1120@sina.com

## Abstract

Arithmetic coding is an indispensable part of high performance data compression, which theoretically could encode data close to Shannon entropy. In arithmetic coding, probability estimation is an important step, because it determines coding efficiency directly. The paper discusses the probability estimation of non-stationary sources—a broad class of sources in practical application, such as images and videos. Especially, the paper deeply analyzes the characteristics of encoded bit stream in wavelet image compression. Then an efficient strategy is proposed, which utilizes the correlations between subbands of wavelet coefficients efficiently. Experimental results show that this method could improve image compression efficiency.

**Keywords:** Probability estimation, Arithmetic coding, Image compression

## 1. Introduction

Arithmetic coding is a powerful technique for data compression that theoretically allows to code data at Shannon entropy. In recent years, arithmetic coders have been widely used in image and video compression, such as MQ-coder [1] in JPEG2000 image compression standard and context-based adaptive binary arithmetic coder

(CABAC) [2] in H.264/AVC video compression standard. Without doubt, arithmetic coding has become indispensable component of high performance data compression.

Probability estimation is an important step in arithmetic coding, which determines the compression efficiency directly. Besides, practical implementations generally require adaptive, on-line probability estimation either because sufficient statistical knowledge of the signal is lacking or because its statistics are time varying [3]. How best to make such estimates is a question of much practical importance.

Many adaptive probability estimation methods have been proposed in image and video compression over recent years. The scaled-count adaptive probability estimator [3] is the representation of these ones, which has been used in the arithmetic coder of the JBIG and JPEG image coding standards. It could achieve desired balance between the steady-state quality of probability estimation and the speed of response to changing underlying statistics. So this estimator possesses broad applicability. Based on this, probability estimation with dual symbol sets [4] was proposed to work for the sources that are characterized by large alphabets and alphabet distributions that are skewed and highly non-stationary. It maintained dual symbol set: a primary symbol set that contains all the symbols that are likely to oc-

cur in the near future and a secondary symbol set that contains all other symbols. Especially, binary arithmetic coding has gained rapid development in recent years, because any source could be represented by binary symbol sets. A table lookup operation without multiplier and renormalization-driven probability estimation are the most distinguishing features of this broad class of arithmetic coding algorithms which includes the Q-coder [5], QM-coder [6], MQ-coder [1], Z-coder [7], B-coder [8] and CABAC [2]. Probability estimations of these coders were realized by looking up a pre-determined table. So the design of lookup table is the key and difficulty, which often includes many complex techniques.

To achieve high coding efficiency, wavelet image compression algorithm generally adopts arithmetic coding as the last step, which is utilized to encode the binary symbol stream produced by the prediction of wavelet coefficients, such as JPEG2000 [1]. Due to the non-stationarity of image source, probability estimation in arithmetic coding should be able to track these variabilities. By analyzing the characteristics of encoded binary stream, the paper develops statistical similarities between diagonal subbands, which could be utilized to lead the probability estimation to track the changing well and gain more precise values. Based on this, the paper proposed a modified probability estimation strategy. It resets probability values when coding process jumps from one subband to another. This method enhances precision of probability estimation, and improves the compression efficiency finally.

The paper is organized as follows. Section 2 describes the scaled-count adaptive probability estimation, which is the base of the proposed method. Then probability estimation of non-stationary sources is discussed in section 3. Based on this, in section 4, characteristics of encoded

stream in image compression is analyzed firstly, and then the strategy how to utilize them is proposed, and experimental results are also reported in this section. Finally section 5 gives some conclusions.

## 2. Scaled- Count Probability Estimation

Perhaps the most natural estimation strategy is to count the being coded symbol, and use the appeared frequency of every symbol as the probability estimation to code the following symbol. Scaled-Count probability estimation [3] is the similar approach, which is calculated by Bayes estimation when the source is independent, identically distribution and its prior distribution is uniform distribution. But a lot of experiments show that this estimation method is applicable to most of sources. Suppose  $C(\cdot)$  denotes the observed number of certain symbol, and its initial value is 0. Then probability estimation of the  $N$ -th binary symbol is

$$\begin{cases} p_N'(1) = \frac{C_{N-1}(1)+1}{(C_{N-1}(0)+1)+(C_{N-1}(1)+1)} \\ p_N'(0) = 1 - p_N'(1) \end{cases} \quad (1)$$

Considered the probability distribution of image sources, the following probability estimation has been found quite powerful in JBIG and JPEG application.

$$\begin{cases} p_N'(1) = \frac{C_{N-1}(1)+\Delta}{(C_{N-1}(0)+\Delta)+(C_{N-1}(1)+\Delta)} \\ p_N'(0) = 1 - p_N'(1) \end{cases} \quad (2)$$

Large values of  $\Delta$  reflect a conservative policy in which we are reluctant to estimate highly skewed (i.e., non-uniform) distributions until we have observed a large number of outcomes. Conversely, smaller values of  $\Delta$  reflect a more radical approach [1].

Furthermore, in order to track the probability changing in time, the author designs the threshold  $C_{min}$  and  $C_{max}$ .  $C(0)$  and  $C(1)$  should be shrunk, generally halved, when  $\min\{C(0), C(1)\} > C_{min}$  or  $\max\{C(0), C(1)\} > C_{max}$ . The selection of  $C_{min}$  and  $C_{max}$  mainly depends on the experiment results and the experience.

The above probability estimation obtains the highest compression efficiency in the reported estimation methods. And the improvement proposed by this paper is also based on it.

### 3. Combinatorial Stationary Representation of Non-stationary Sources

Non-stationary sources are a class of most common sources in practical application, whose probability distributions have relationship with the beginning time. Otherwise, there are statistics dependences between random variables of sequence. For instance, image sources are non-stationary. In wavelet image compression, non-stationarity of image sources determines that the binary sequence produced by prediction of wavelet coefficients is also non-stationary. The characteristics of non-stationary sources make it difficult to analyze them. Hence, in this section, a kind of simple source is introduced firstly, and then it will be used to discuss the non-stationary sources.

Suppose that an information source produces a random sequence  $\mathbf{X} = (X_1 X_2 X_3 \dots X_N)$  from a finite set, where  $X_1, X_2, X_3, \dots$  are independent, identically random variables with same probability distributions  $P$ , which is not changed with time. Such an information source is called a *memoryless stationary source* (MSS) [9]. For this kind of source, the Shannon Entropy  $H(\mathbf{X})$  is the low limit of compression in theory.

$$\begin{aligned} H(\mathbf{X}) &= H(X^N) \\ &= NH(X) \\ &= NH(P) \\ &= -N \sum_x P(X) \log P(X) \end{aligned} \quad (3)$$

The MSS is one of the simplest sources. Most of probability estimation methods are applicable to this kind of sources. Hence, it is feasible to utilize this kind of sources to discuss the complex non-stationary sources. As a result, the existing probability estimation method could be utilized to solve the problems in non-stationary sources coding well.

Non-stationary source could be regarded as combination of many MSSs in practical application. For instance, the local data of some sources could be regarded as being stationary, although they possess non-stationary characteristics as a whole, such as the flat area in the image. Treat a non-stationary sequence as many locally stationary sub-sequences, then it can be proved that the entropy  $H(\mathbf{Y})$  of the combinatorial source is not more than the entropy  $H(\mathbf{X})$  which treats the non-stationary source as a single stationary source.

**Theorem** Shannon Entropy calculated by regarding non-stationary source as combinatorial stationary sources is not more than that regarding it as a single stationary source.

**Proof** Suppose an N-length non-stationary code stream sequence  $\mathbf{X}$  consists of  $L$  independent stationary sub-sequences with length  $N_i (i = 1, 2, 3, \dots, L)$  respectively, define  $p_i = N_i/N (i = 1, 2, 3, \dots, L)$ , then

$$\sum_{i=1}^L p_i = 1 \quad (4)$$

Let  $Y$  be combination of sub-sequences. Because every sub-sequence is stationary, all the symbols in a certain sub-sequence have same probability distribution  $P_i (i = 1, 2, 3, \dots, L)$ . Then if regard the non-stationary sequence  $X$  as a stationary one, and let  $P$  be its probability distribution, then have

$$P = \sum_{i=1}^L p_i P_i \quad (5)$$

Because the entropy function is convex  $\cap$  function, according to Jensen's inequality [9] there have

$$\begin{aligned} H(Y) &= \sum_{i=1}^L N_i H(P_i) \\ &= N \cdot \sum_{i=1}^L p_i H(P_i) \\ &\leq N \cdot H\left(\sum_{i=1}^L p_i P_i\right) \\ &= N \cdot H(P) = H(X) \end{aligned} \quad (6)$$

So Theorem is proved.

The Theorem means that better compression result could be obtained if treating a non-stationary sequence as several stationary sub-sequences. In wavelet image compression, non-stationarity of image sources determines that the binary sequence produced by prediction of wavelet coefficients is also non-stationary. Based on conclusion above, if a binary sequence corresponding to certain prediction value could be discriminated further, then better compression performance could be expected.

#### 4. Algorithm Based on Probability Resetting

In application of wavelet image compression, arithmetic coding is generally utilized to code the binary code stream which

is produced by prediction of wavelet coefficients. By scanning and predicting based on bit plane (BP), the wavelet coefficients are classified into several binary sub-streams. Theoretically, if each bit is regarded as a random variable, then all bits in the sub-stream have same probability distributions, but the bits belonging to different sub-stream possess different probability distributions. Hence, prediction technology make the following arithmetic coding be able to gain the high-level entropy of the sources. Subsequently, every sub-stream is sent to one arithmetic coder, which removes the statistical redundancy of the bit stream fully.

In practice, probability distribution in the binary sub-stream is changing. Table 1 gives an example of compressing image "city". Here, prediction technology in [10] is adopted, which is very efficient even better than that in JPEG2000. The data shown in Table 1 are percent of bit "1" corresponding to prediction value "9" of every subband in different bit planes. The sign "-" in table means that there is no data corresponding to this prediction value in current bit plane and subband, or the number of the sample data is too little to have statistical meanings. Obviously, the sub-stream is non-stationary.

But if observing carefully, the below characteristics of the sub-stream could be found from the table.

- The percent of "1" decreases from the lowest subband to the highest subband rapidly, and it increases from the highest bit plane to the lowest bit plane.
- The percent of the diagonal direction have higher similarity.

The above characters are also applicable to other prediction values of other images. Hence, based on conclusion of previous section, higher compression efficiency could be achieved by discriminating sub-segments in different subband and bit plane which belong to one sub-stream.

Table 1: Percentage of bit “1” in Subbands Corresponding to Prediction Value “9”

BP	Subband Number					
	1	2	3	4	5	6
5	5.13	0.0	0.0	0.0	0.0	0.0
4	30.77	8.45	1.81	0.03	0.0	0.0
3	—	16.67	12.34	1.76	0.02	0.0
2	—	—	14.29	11.38	2.32	0.01
1	—	—	—	20.34	10.88	1.54

Based on analysis above, this paper adopts the following strategies to improve the compression efficiency.

- Reset probability estimation value of the symbol between subbands each bit plane, and use the scaled frequency value of the diagonal direction as the resetting value if exists.
- Otherwise, adopt the following measures: use the scaled frequency value of the lower subband as the initial probability estimation value of the current subband in the highest bit plane; use the higher bit plane initializes the lower bit plane in the lowest subband, use 50% as the initialization probability estimation value for the lowest subband in the highest bitplane.
- Pay attention to make sure that the probability estimation value is between 0 and 1 but not equal to them, otherwise there may appears errors in decoding process.

Table 2 shows the compression results of four images which use arithmetic coding with scaled-count (SC) probability estimation, SC with probability reset (SCPS) and MQ-coder (adopted in JPEG2000). The image size is 512×512, and the PSNR of the reconstructed image is 30dB. Compared with SC, saved bits and percentage of the SCPS are given in

the last two columns. The SCPS could track the probability change more quickly, so it gains the better results than the SC and MQ.

Table 2: Compression Results for SCPS and Other Methods

Image	Length of Compression Stream( <i>bit</i> )				Saved Pct.
	MQ	SC	SCPS	Saved	
Lena	23628	23046	22717	329	1.4%
Goldhill	53324	51910	50841	1069	2.1%
Barbara	89964	87382	85251	2131	2.4%
City	309320	303936	299166	4770	1.6%

## 5. Conclusion

The major contribution of this work is developing the probability distribution similarities in diagonal direction subbands of wavelet coefficients. Based on this, a simple strategy is proposed to enhance precision of probability estimation and improve compression efficiency finally. In this paper, the scaled-count probability estimation is adopted as the basis of the proposed technology. In fact, most of probability estimations can utilize this technology.

Complexity is not discussed in this paper. In fact, the proposed method itself is very simple, so as to complexity of the algorithm is mainly from the probability estimation method. So combining the proposed method with other simple probability estimations is the future work.

## 6. Acknowledgment

This work was supported by the NSFC (60775018) and 863 Program (HI-Tech Research and Development Program of China). The research was made in the State Key Laboratory of Virtual Reality Technologies.

## 7. References

- [1] D. S. Taubman, M. W. Marcellin, JPEG2000 Image Compression Fundamentals, Standard and Practice Kluwer Academic Publishers, 2001.
- [2] D. Marpe, H. Schwarz, T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," IEEE Trans. Circ. Syst. Video Tech., vol. 13, no. 7, pp. 620-636, Jul. 2003.
- [3] D. L. Duttweiler, C. Chamzas, "Probability estimation in arithmetic and adaptive-huffman entropy coders," IEEE Trans. on Image Processing, vol. 4, no. 3, pp. 237-246, Mar. 1995.
- [4] B. Zhu, E. H. Yang, A. H. Tewfik, "Arithmetic coding with dual symbol sets and its performance analysis," IEEE Trans. on Image Processing, vol.8, no. 12, pp. 1667-1676, Dec. 1999.
- [5] W. Pennebaker, J. Mitchell, G. Langdon, and R. Arps, "A overview of the basic principles of the Q-coder adaptive binary arithmetic coder," IBM J. Res. Develop., vol. 32, no. 6, pp. 717-726, Nov. 1988.
- [6] W. Pennebaker, J. Mitchell, JPEG: Still Image Data Compression Standard, Van Nostrand Reinhold, New York, 1992.
- [7] L. Bottou, P. G. Howard, Y. Bengio, "The Z-coder adaptive binary coder," Proc. IEEE Data Compression Conf., pp 13-22, 1998.
- [8] B. Kelly, D. Brailsford, "The B-coder: an improved binary arithmetic coder and probability estimator," IEEE Data Compression Conference, 28-30 Mar. 2006.
- [9] R. J. McEliece, The Theory of Information and Coding, Cambridge University Press, 2002.
- [10] B. Li, H. Wang, "Bit plane predicting image compression algorithm based on wavelet packet transform," Journal of Computers, vol. 22, no. 7, pp. 685-691, Jul. 1999 (*in Chinese*).