entropy

MDPI

Article

Stochastic Model of Block Segmentation Based on Improper Quadtree and Optimal Code under the Bayes Criterion †

Yuta Nakahara [1,*] and Toshiyasu Matsushima [2]

[1] Center for Data Science, Waseda University, 1-6-1 Nisniwaseda, Shinjuku-ku, Tokyo 169-8050, Japan
[2] Department of Pure and Applied Mathematics, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
[*] Correspondence: yuta.nakahara@aoni.waseda.jp

† This paper is an extended version of our paper published in Nakahara, Y.; Matsushima, T. Stochastic Model of Block Segmentation Based on Improper Quadtree and Optimal Code under the Bayes Criterion. In Proceedings of the 2022 Data Compression Conference (DCC), Snowbird, UT, USA, 22–25 March 2022; pp. 153–162.

Abstract: Most previous studies on lossless image compression have focused on improving preprocessing functions to reduce the redundancy of pixel values in real images. However, we assumed stochastic generative models directly on pixel values and focused on achieving the theoretical limit of the assumed models. In this study, we proposed a stochastic model based on improper quadtrees. We theoretically derive the optimal code for the proposed model under the Bayes criterion. In general, Bayes-optimal codes require an exponential order of calculation with respect to the data lengths. However, we propose an algorithm that takes a polynomial order of calculation without losing optimality by assuming a novel prior distribution.

Keywords: stochastic generative model; quadtree; Bayes code; lossless image compression

check for updates

Citation: Nakahara, Y.; Matsushima, T. Stochastic Model of Block Segmentation Based on Improper Quadtree and Optimal Code under the Bayes Criterion. *Entropy* **2022**, *24*, 1152. https://doi.org/10.3390/e24081152

Academic Editors: Chun-Hung Liu, Jwo-Yuh Wu and Peter Y. Hong

Received: 25 July 2022
Accepted: 17 August 2022
Published: 19 August 2022

1. Introduction

There are two approaches to lossless image compression. (These two approaches are detailed in Section 1 of our previous study [1].) Most previous studies (e.g., [2–4]) adopted an approach in which they constructed a preprocessing function $f : v^{t-1} \mapsto \boldsymbol{p}$ that outputs a code length assignment vector $\boldsymbol{p}$ from past pixel values v^{t-1}. $\boldsymbol{p}$ determines the code length of the next pixel value v_t, or typically, a value v'_t equivalent to v_t in the meaning that there exists a one-to-one mapping $(v'_1, v'_2, \dots v'_t) = g(v_1, v_2, \dots v_t)$ computable for both encoder and decoder. Then, v'_t and $\boldsymbol{p}$ are passed to the following entropy coding process such as [5,6]. In this approach, the elements p_i of the code length assignment vector $\boldsymbol{p}$ satisfy $\sum_i p_i = 1$. Therefore, it appears superficially as a probability distribution. However, it does not directly govern the stochastic generation of original pixel value v_t. Hence, we cannot define the entropy of the source of pixel value v_t, and we cannot discuss the theoretical optimality of the preprocessing function $f(v^{t-1})$ and one-to-one mapping $g(v_1, v_2, \dots v_t)$.

In contrast, we adopted an approach in which we estimated a stochastic generative model $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ with an unknown parameter $\boldsymbol{\theta}^m$ and a model variable m, which is directly and explicitly assumed on the original pixel value v_t [1,7–9]. Therefore, we can discuss the theoretical optimality of the entire algorithm to the entropy defined from the assumed stochastic model $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$. In particular, we can achieve the theoretically optimal coding under the Bayes criterion in statistical decision theory (see, e.g., [10]) by assuming prior distributions $p(\boldsymbol{\theta}^m|m)$ and $p(m)$ on the unknown parameter $\boldsymbol{\theta}^m$ and model variable m. Such codes are known as Bayes codes [11] in information theory. It is known that the Bayes code asymptotically achieves the entropy of the true stochastic model, and its convergence speed achieves the theoretical limit [12]. The Bayes codes have shown remarkable performance in text compression (e.g., [13]). Therefore, we consider this approach.

We assume that the target image herein has non-stationarity, that is, the properties of pixel values are different among the positions in the image. For such an image, researchers have performed quadtree block segmentation as a component of preprocessing $f(v^{t-1})$ and one-to-one mapping $g(v_1, v_2, \dots v_t)$ in the former approach, and its practical efficiency has been reported in many previous studies (e.g., [4,14]). In the latter approach, we proposed a stochastic generative model $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ that contains a quadtree as a model variable m. By assuming a prior distribution $p(m)$ on it, we derived the optimal code under the Bayes criterion, and we constructed a polynomial order algorithm to calculate it without loss of optimality [1]. However, in all these studies [1,4,14], the class of quadtrees is restricted to that of proper trees, whose inner nodes have exactly four children.

In this paper, we propose a stochastic generative model $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ based on an improper quadtree m and derive the code optimal under the Bayes criterion. In general, the codes optimal under the Bayes criterion require a summation that takes an exponential order calculation for the data length. However, we herein construct an algorithm that only requires a polynomial order calculation without losing optimality by applying a theory of probability distribution for general rooted trees [15] to the improper quadtree representing the block segmentation.

2. Proposed Stochastic Generative Model

Let $\mathcal{V}$ denote a set of possible values of a pixel. For example, we have $\mathcal{V} = \{0, 1\}$ for binary images and $\mathcal{V} = \{0, 1, \dots, 255\}$ for grayscale images. Let $h \in \mathbb{N}$ and $w \in \mathbb{N}$ denote a height and a width of an image, respectively. Although our model is able to represent any rectangular images, we assume that $h = w = 2^{d_{\max}}$ for $d_{\max} \in \mathbb{N}$ in the following for the simplicity of the notation. Then, let V_t denote the random variable of the t-th pixel value in order of the raster scan, and let $v_t \in \mathcal{V}$ denote its realization. Note that V_t is at the $x(t)$-th row and $y(t)$-th column, where t divided by w is $x(t)$ with a reminder of $y(t)$. In addition, let V^t denote the sequence of pixel values $V_0, V_1, \dots, V_t$. Note that all the indices start from zero herein.

We assume V_t is generated from a probability distribution $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ depending on an unknown model $m \in \mathcal{M}$ and unknown parameters $\boldsymbol{\theta}^m \in \boldsymbol{\Theta}^m$. (For $t = 0$, we assume V_0 follows $p(v_0|\boldsymbol{\theta}^m, m)$.) We define m and $\boldsymbol{\theta}^m$ in the following.

Definition 1 ([1])**.** *Let $s_{(x_1y_1)(x_2y_2)\cdots(x_dy_d)}$ denote the following index set called "block."*

$$s_{(x_1y_1)(x_2y_2)\cdots(x_dy_d)} := \left\{ (i,j) \in \mathbb{Z}^2 \;\middle|\; \sum_{d'=1}^{d} \frac{x_{d'}}{2^{d'}} \le \frac{i}{2^{d_{\max}}} < \left(\sum_{d'=1}^{d} \frac{x_{d'}}{2^{d'}} + \frac{1}{2^d} \right), \right.$$
$$\left. \sum_{d'=1}^{d} \frac{y_{d'}}{2^{d'}} \le \frac{j}{2^{d_{\max}}} < \left(\sum_{d'=1}^{d} \frac{y_{d'}}{2^{d'}} + \frac{1}{2^d} \right) \right\}, \tag{1}$$

where $(x_{d'}y_{d'}) \in \{0,1\}^2$ for $d' \in \{1, 2, \dots, d\}$ and $d \le d_{\max}$. In addition, let s_λ be the set of whole indices $s_\lambda := \{0, 1, \dots h-1\} \times \{0, 1, \dots, w-1\}$. Then, let $\mathcal{S}$ denote the set that consists of all the above index sets, that is, $\mathcal{S} := \{s_\lambda, s_{(00)}, \dots, s_{(11)}, s_{(00)(00)}, \dots, s_{(11)(11)}, \dots, s_{(11)(11)\cdots(11)}\}$.

Example 1 ([1])**.** *For $d_{\max} = 2$,*

$$s_{(01)} = \{(i,j) \in \mathbb{Z}^2 \mid 0 \le i < 2, 2 \le j < 4\} = \{(0,2), (0,3), (1,2), (1,3)\}. \tag{2}$$

Therefore, it represents the indices of the upper right region. In a similar manner, $s_{(01)(11)} = \{(i,j) \in \mathbb{Z}^2 \mid 1 \le i < 2, 3 \le j < 4\} = \{(1,3)\}$. It should be noted that the cardinality $|s|$ for each $s \in \mathcal{S}$ represents the number of pixels in the block.

Definition 2. *We define the model m as a quadtree whose nodes are elements of $\mathcal{S}$. Let $\mathcal{M}$ denote the set of the models. Let $\mathcal{S}^m \subset \mathcal{S}$, $\mathcal{L}^m \subset \mathcal{S}$ and $\mathcal{I}^m \subset \mathcal{S}$ denote the set of the nodes, the leaf nodes and the inner nodes of $m \in \mathcal{M}$, respectively. Let $\mathcal{U}^m \subset \mathcal{S}^m$ denote the set of nodes that have less than four children. Then, $\mathcal{U}^m$ corresponds to a pattern of variable block size segmentation, as shown in Figure 1.*

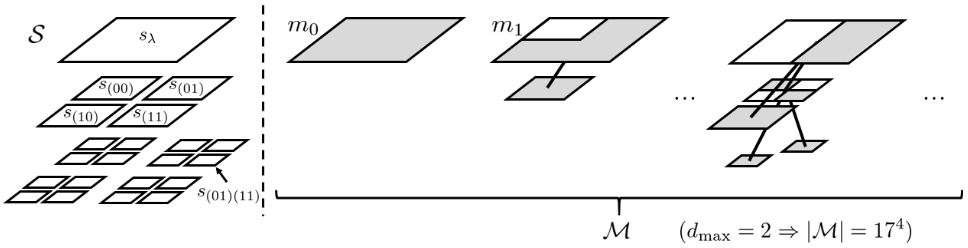

Figure 1. An example of node set $\mathcal{S}$ and models m. The set of blocks with gray region corresponds to $\mathcal{U}^m$, which covers the whole region of the image and represents a block segmentation pattern.

Definition 3. *Each node $s \in \mathcal{U}^m$ of the model m has a parameter θ_s^m whose parameter space is Θ_s^m. We define $\boldsymbol{\theta}^m$ as a tuple of parameters $\{\theta_s^m\}_{s \in \mathcal{U}^m}$, and let $\boldsymbol{\Theta}^m$ denote its space.*

Notably, we can reduce the number of parameters from an equivalent model represented by a proper tree with added dummy child nodes. See the following example.

Example 2. *For $d_{\max} = 2$, consider a model represented by the left-hand side image in Figure 2. It has three parameters: θ_{s_λ}, $\theta_{s_{(00)}}$, and $\theta_{s_{(10)}}$. An equivalent model can be represented by a proper quadtree shown in the right-hand side of Figure 2, if assuming $\theta_{s_{(01)}} = \theta_{s_{(11)}}$ by chance. However, it requires four parameters: $\theta_{s_{(00)}}$, $\theta_{s_{(01)}}$, $\theta_{s_{(10)}}$, and $\theta_{s_{(11)}}$. Therefore, it causes inefficient learning.*

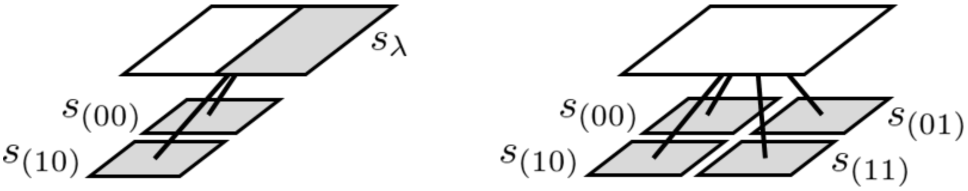

Figure 2. A model with three parameters (**left**) and a model with four parameters (**right**).

Under the model $m \in \mathcal{M}$ and the parameters $\boldsymbol{\theta}^m \in \boldsymbol{\Theta}^m$, we assume that the t-th pixel value V_t is generated as follows.

Assumption 1. *We assume that*

$$p(v_t | v^{t-1}, \boldsymbol{\theta}^m, m) = p(v_t | v^{t-1}, \theta_s^m), \tag{3}$$

where s is the minimal block that satisfies $(x(t), y(t)) \in s \in \mathcal{U}^m$ (in other words, s is the the deepest node that contains $(x(t), y(t))$ in m). For $t = 0$, we assume a similar condition $p(v_0 | \boldsymbol{\theta}^m, m) = p(v_0 | \theta_s^m)$.

Thus, the pixel value V_t given the past sequence V^{t-1} depends only on the parameter of the minimal block s that contains V_t. Note that we do not assume a specific form of $p(v_t | v^{t-1}, \theta_s^m)$ at this point. For example, we can assume the Bernoulli distribution for $\mathcal{V} = \{0, 1\}$ and also the Gaussian distribution (with an appropriate normalization and quantization) for $\mathcal{V} = \{0, 1, \ldots, 255\}$.

3. The Bayes Code for Proposed Model

Since the true m and $\boldsymbol{\theta}^m$ are unknown, we assume prior distributions $p(m)$ and $p(\boldsymbol{\theta}^m | m)$. Then, we estimate the true generative probability $p(v_t | v^{t-1}, \boldsymbol{\theta}^m, m)$ by $q(v_t | v^{t-1})$

under the Bayes criterion in statistical decision theory (see, e.g., [10]). Subsequently, we use $q(v_t|v^{t-1})$ as a coding probability of the entropy code such as [16]. Such a code is known as Bayes codes [11] in information theory. The expected code length of the Bayes code converges to the entropy of $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ for sufficiently large data length, and its convergence speed achieves the theoretical limit [12]. The Bayes code has shown remarkable performances in text compression (e.g., [13]).

The optimal coding probability of the Bayes code for v_t is derived as follows, according to the general formula in [11].

Proposition 1. *The optimal coding probability $q^*(v_t|v^{t-1})$ under the Bayes criterion is given by*

$$q^*(v_t|v^{t-1}) = \sum_{m \in \mathcal{M}} p(m|v^{t-1}) \int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m) \mathrm{d}\boldsymbol{\theta}^m. \tag{4}$$

We call $q^*(v_t|v^{t-1})$ the Bayes-optimal coding probability.

Proposition 1 implies that we should use the coding probability that is a weighted mixture of $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ for every block segmentation pattern m and parameters $\boldsymbol{\theta}^m$ according to the posteriors $p(m|v^{t-1})$ and $p(\boldsymbol{\theta}^m|v^{t-1}, m)$. (For $t = 0$, $p(v_0|\boldsymbol{\theta}^m, m)$ is mixed with weights according to the priors $p(m)$ and $p(\boldsymbol{\theta}^m|m)$, which corresponds to the initialization of the algorithm.) Notably, $\mathcal{M}$ is generalized to the set of improper quadtrees from the set of proper quadtrees although (4) has a similar form to Formula (5) in [1].

4. Polynomial Order Algorithm to Calculate Bayes-Optimal Coding Probability

Unfortunately, the Bayes-optimal coding probability (4) contains a computationally hard calculation. (Herein, we assume that $\int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m) \mathrm{d}\boldsymbol{\theta}^m$ is feasible. Examples of feasible settings will be described in the next section.) The summation cost for m exponentially increases with respect to $d_{\max}$. Therefore, we propose a polynomial order algorithm to calculate (4) without loss of optimality by applying a theory of probability distribution for general rooted trees [15] to the improper quadtree m. In this section, we focus on the procedure of the constructed algorithm. Its validity is described in Appendix A.

Definition 4. *Let $\mathrm{Ch}(s) := \{s_{(00)}, s_{(01)}, s_{(10)}, s_{(11)}\}$ be the set of child nodes of s. We define a vector $\boldsymbol{z}_s^m \in \{0,1\}^4$ representing the block division pattern of s in $\mathcal{S}^m$ as $\boldsymbol{z}_s^m := (z_{ss'}^m)_{s' \in \mathrm{Ch}(s)} := (I\{s_{(00)} \in \mathcal{S}^m\}, I\{s_{(01)} \in \mathcal{S}^m\}, I\{s_{(10)} \in \mathcal{S}^m\}, I\{s_{(11)} \in \mathcal{S}^m\})$, where $I\{\cdot\}$ denotes the indicator function. Examples of $\boldsymbol{z}_s^m$ are shown in Figure 3. For leaf nodes, $\boldsymbol{z}_s^m = \boldsymbol{0}$.*

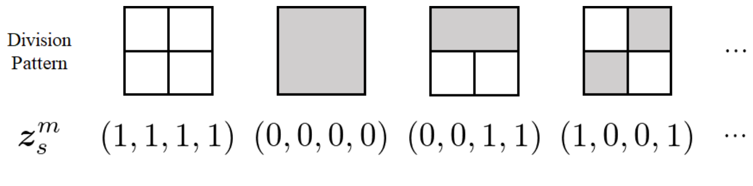

Figure 3. Examples of block division patterns and corresponding $\boldsymbol{z}_s^m$.

First, we assume the following prior distributions as $p(m)$ and $p(\boldsymbol{\theta}^m|m)$.

Assumption 2. *Let $\eta_s(\boldsymbol{z}) \in [0,1]$ be a given hyper parameter of a block $s \in \mathcal{S}$, which satisfies $\sum_{\boldsymbol{z} \in \{0,1\}^4} \eta_s(\boldsymbol{z}) = 1$. Then, we assume that the prior on $\mathcal{M}$ is represented as follows.*

$$p(m) = \prod_{s \in \mathcal{S}_m} \eta_s(\boldsymbol{z}_s^m), \tag{5}$$

where $\eta_s(\boldsymbol{0}) = 1$ for s whose cardinality $|s|$ is equal to 1.

Intuitively, $\eta_s(\boldsymbol{z}_s^m)$ represents the conditional probability that s has the block division pattern $\boldsymbol{z}_s^m$ under the condition that $s \in \mathcal{S}^m$. The above prior actually satisfies the condition $\sum_{m\in\mathcal{M}} p(m) = 1$. Although this is proved for any rooted tree in [15], we briefly describe a proof restricted for our model in the Appendix A to make this paper self-contained. Note that the above assumption does not restrict the expressive capability of the general prior in the meaning that each model m still has possibly to be assigned a non-zero probability $p(m) > 0$.

Assumption 3. *For each model $m \in \mathcal{M}$, we assume that*

$$p(\boldsymbol{\theta}^m|m) = \prod_{s\in\mathcal{U}^m} p(\theta_s^m|m). \tag{6}$$

Moreover, for any $m, m' \in \mathcal{M}$, $s \in \mathcal{U}^m \cap \mathcal{U}^{m'}$, and $\theta_s \in \Theta_s$, we assume that

$$p(\theta_s|m) = p(\theta_s|m') =: p_s(\theta_s). \tag{7}$$

Therefore, each element θ_s^m of the parameters $\boldsymbol{\theta}^m$ depends only on s and they are independent from both of the other elements and the model m.

From Assumptions 1 and 3, the following lemma holds.

Lemma 1. *For any $m, m' \in \mathcal{M}$, let $s_t \in \mathcal{U}^m$ and $s_t' \in \mathcal{U}^{m'}$ denote the minimal node that satisfies $(x(t), y(t)) \in s_t \in \mathcal{U}^m$ and $(x(t), y(t)) \in s_t' \in \mathcal{U}^{m'}$, respectively. If $s_t = s_t' =: s$ and $\boldsymbol{z}_{s_t}^m = \boldsymbol{z}_{s_t'}^{m'} =: \boldsymbol{z}_s$, that is, they are the same block and their division patterns are also the same, then*

$$p(v_t|v^{t-1}, m) = p(v_t|v^{t-1}, m'). \tag{8}$$

Hence, we represent it by $\tilde{q}(v_t|v^{t-1}, s, \boldsymbol{z}_s)$ because it does not depend on m but $(s, \boldsymbol{z}_s)$. Let $\tilde{q}(v_t|v^{t-1}, s) = p(v_t|m) = p(v_t|m')$ for $t = 0$.

Lemma 1 means that the optimal coding probability for v_t depends on the minimal block s that contains v_t and its division pattern $\boldsymbol{z}_s$. Therefore, it could be calculated as $\tilde{q}(v_t|v^{t-1}, s, \boldsymbol{z}_s)$ if $(s, \boldsymbol{z}_s)$ was known.

At last, the Bayes-optimal coding probability $q^*(v_t|v^{t-1})$ can be calculated by a recursive function for nodes on a path of the perfect quadtree on $\mathcal{S}$. The definition of the path is the same as [1].

Definition 5 ([1])**.** *Let $\mathcal{S}_t$ denote the set of nodes which contain $(x(t), y(t))$. They construct a path from the leaf node $s_{(x_1y_1)(x_2y_2)\cdots(x_{d_{\max}}y_{d_{\max}})} = \{(x(t), y(t))\}$ to the root node s_λ on the perfect quadtree whose depth is $d_{\max}$ on $\mathcal{S}$, as shown in Figure 4. In addition, let $s_{\mathrm{ch}} \in \mathcal{S}_t$ denote the child node of $s \in \mathcal{S}_t$ on that path.*

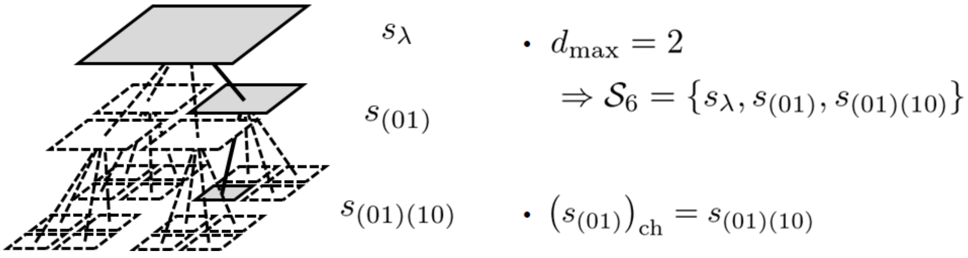
s_λ
$s_{(01)}$
$s_{(01)(10)}$
• $d_{\max} = 2$
$\Rightarrow \mathcal{S}_6 = \{s_\lambda, s_{(01)}, s_{(01)(10)}\}$
• $\left(s_{(01)}\right)_{\mathrm{ch}} = s_{(01)(10)}$

Figure 4. An example of a path constructed from $\mathcal{S}_t$.

Definition 6. *We define the following recursive function $q(v_t|v^{t-1},s)$ for $s \in \mathcal{S}_t$.*

$$q(v_t|v^{t-1},s) := \begin{cases} \tilde{q}(v_t|v^{t-1},s,\mathbf{0}), & |s|=1, \\ \sum_{z_s:z_{ss_{\mathrm{ch}}}=0} \eta_s(z_s|v^{t-1})\tilde{q}(v_t|v^{t-1},s,z_s) & \\ \quad +\left(\sum_{z_s:z_{ss_{\mathrm{ch}}}=1} \eta_s(z_s|v^{t-1})\right) q(v_t|v^{t-1},s_{\mathrm{ch}}), & \text{otherwise}, \end{cases} \tag{9}$$

where $\eta_s(z_s|v^t)$ is also recursively updated for $s \in \mathcal{S}_t$ as follows:

$$\eta_s(z_s|v^t) := \begin{cases} \eta_s(z_s), & t=-1, \\ \frac{\eta_s(z_s|v^{t-1})\tilde{q}(v_t|v^{t-1},s,z_s)}{q(v_t|v^{t-1},s)}, & t \geq 0 \wedge z_{ss_{\mathrm{ch}}}=0, \\ \frac{\eta_s(z_s|v^{t-1})q(v_t|v^{t-1},s_{\mathrm{ch}})}{q(v_t|v^{t-1},s)}, & t \geq 0 \wedge z_{ss_{\mathrm{ch}}}=1. \end{cases} \tag{10}$$

Consequently, the following theorem holds.

Theorem 1. *The Bayes-optimal coding probability $q^*(v_t|v^{t-1})$ for the proposed model is calculated by*

$$q^*(v_t|v^{t-1}) = q(v_t|v^{t-1},s_\lambda). \tag{11}$$

Although Theorem 1 is proved by applying Corollary 2 of Theorem 7 in [15], we briefly describe a proof restricted to our model in the Appendix A to make this paper self-contained. Theorem 1 means that the summation with respect to $m \in \mathcal{M}$ in (4) is able to be replaced by the summation with respect to $s \in \mathcal{S}_t$ and $z_s \in \{0,1\}^4$, which costs only $O(2^4 d_{\max})$. The proposed algorithm recursively calculates a weighted mixture of coding probabilities $\tilde{q}(v_t|v^{t-1},s,z_s)$ for the case where block s is not divided at s_{ch} (i.e., $z_{ss_{\mathrm{ch}}}=0$) and the coding probability $q(v_t|v^{t-1},s_{\mathrm{ch}})$ for the case where block s is divided at s_{ch} (i.e., $z_{ss_{\mathrm{ch}}}=1$).

5. Experiments

In this section, we perform four experiments. Three of them are similar to the experiments in [1]. The fourth one is newly added. In Experiments 1, 2, and 3, we assume $\mathcal{V}=\{0,1\}$, which is the simplest setting, to focus on the effect of the improper quadtrees. In Experiment 4, we assume $\mathcal{V}=\{0,1,\dots,255\}$ to show our method is also applicable to grayscale images. The purpose of the first experiment is to confirm the Bayes optimality of $q(v_t|v^{t-1},s_\lambda)$ for synthetic images generated from the proposed model. The purpose of the second experiment is to show an example image suitable to our model. The purpose of the third experiment is to compare average coding rates of our proposed algorithm with a current image coding procedure on real images. The purpose of the fourth experiment is to show our method is applicable to grayscale images.

In Experiments 1 and 2, $p(v_t|v^{t-1},\boldsymbol{\theta}^m,m)$ is Bernoulli distribution $\mathrm{Bern}(v_t|\theta_s^m)$ for the minimal s that satisfies $(x(t),y(t)) \in s \in \mathcal{U}^m$. Each element of $\boldsymbol{\theta}^m$ is i.i.d. distributed with the beta distribution $\mathrm{Beta}(\theta|\alpha,\beta)$, which is the conjugate prior distribution of Bernoulli distribution. Therefore, the integral in (4) has a closed form. The hyperparameter $\eta_s(z)$ of the model prior is $\eta_s(z)=1/2^4$ for every $s \in \mathcal{S}$ and $z \in \{0,1\}^4$, and the hyperparameters of the beta distribution are $\alpha=\beta=1/2$. For comparison, we used the previous method based on proper quadtrees, whose hyperparameters are the same as the experiments in [1], and the standard methods known as JBIG [17] and JBIG2 [18].

5.1. Experiment 1

The setting of Experiment 1 is as follows. The width and height of images are $w=h=2^{d_{\max}}=64$. We generate 1000 images according to the following procedure.

1. Generate m according to (5).
2. Generate θ_s^m according to $p(\theta_s^m|m)$ for $s \in \mathcal{U}^m$.

3. Generate pixel value v_t according to $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ for $t \in \{0, 1, \ldots, hw-1\}$.
4. Repeat Steps 1 to 3 for 1000 times.

Examples of the generated images are shown in Figure 5. Subsequently, we compress these 1000 images. The size of the image is saved in the header of the compressed file using 4 bytes. The coding probability calculated by the proposed algorithm is quantized in 2^{16} levels and substituted into the range coder [16]. Table 1 shows the coding rates (bit/pel) averaged over all the images. Our proposed code has the minimum coding rate as expected by the Bayes optimality.

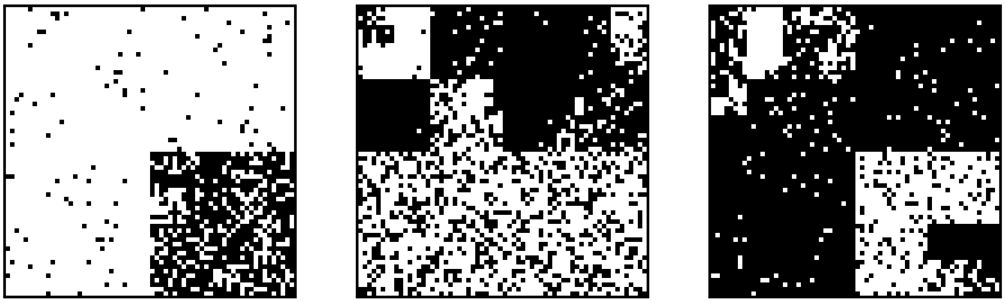

Figure 5. Examples of the generated images in Experiment 1.

Table 1. The average coding rates (bit/pel).

Improper Quadtree (Proposal)	Proper Quadtree [1]	JBIG [17]	JBIG2 [18]
0.619	0.624	1.811	0.962

5.2. Experiment 2

In Experiment 2, we compress `camera.tif` in [19], which is binarized with the threshold of 128. The setting of the header and the range coder is the same as those of Experiment 1. Figure 6 visualizes the maximum a posteriori (MAP) estimation $m^{\mathrm{MAP}} = \arg\max_m p(m|v^{hw-1})$ based on the improper quadtree model and the proper quadtree model [1], which are by-products of the compression. They are obtained by applying Theorem 3 in [15] and the algorithm in Appendix B in the preprint of the full version of [15], which is uploaded on arXiv. The improper quadtree represents the non-stationarity by a fewer number of regions (i.e., fewer parameters) than that of the proper quadtree [1]. Table 2 shows that the coding rate of our proposed model for `camera.tif` is lower than the previous one based on the proper quadtree [1] and JBIG [17] without any special tuning. However, JBIG2 [18] showed the lowest coding rate. The improvement of our method for real images will be described in the next experiment.

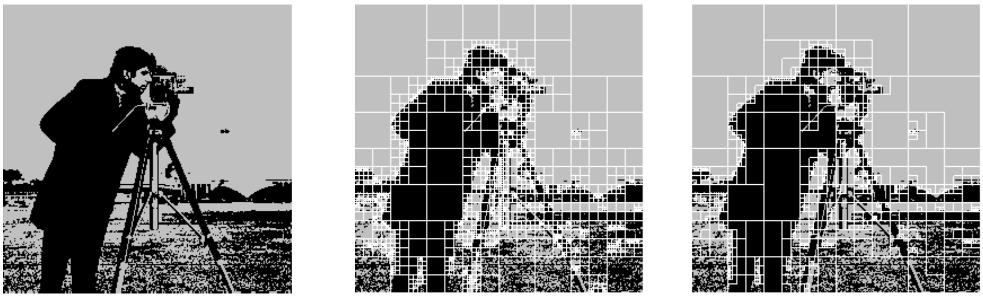

Figure 6. The original image (**left**), the MAP estimated model m^{MAP} based on the proper quadtree [1] (**middle**), and that based on the improper quadtree (**right**).

Table 2. The coding rates for the `camera.tif` in [19] (bit/pel).

Improper Quadtree (Proposal)	Proper Quadtree [1]	JBIG [17]	JBIG2 [18]
0.318	0.323	0.348	**0.293**

5.3. Experiment 3

In Experiment 3, we compare the proposed algorithm with the proper-quadtree-based algorithm [1], JBIG [17], and JBIG2 [18] on real images from [19]. They are binarized in a similar manner to Experiment 2. The setting of the header and the range coder is the same as those of Experiments 1 and 2. A difference from Experiments 1 and 2 is in the stochastic generative model $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ assumed on each block s. We assume another model $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ represented as the Bernoulli distribution $\mathrm{Bern}(v_t|\theta^m_{s;v_{t-w-1}v_{t-w}v_{t-w+1}v_{t-1}})$ that depends on the neighboring four pixels. (If the indices go out of the image, we use the nearest past pixel in Manhattan distance.) Therefore, $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ has a kind of Markov property. In other words, there are 16 parameters $\theta^m_{s;0000}, \theta^m_{s;0001}, \dots, \theta^m_{s;1111}$ for each block s of model m, and one of them is chosen by the observed values v_{t-w-1}, v_{t-w}, v_{t-w+1}, and v_{t-1} in the past. Each parameter is i.i.d. distributed with the beta distribution whose parameters are $\alpha = \beta = 1/2$. The results are shown in Table 3. The algorithms labeled as Improper-i.i.d. and Proper-i.i.d. are the same as those in Experiments 1 and 2. The algorithms labeled as Improper-Markov and Proper-Markov are the aforementioned ones.

Table 3. The coding rates for the binarized images from [19] (bit/pel).

Images	Proper-i.i.d	Improper-i.i.d.	JBIG [17]	Proper-Markov	JBIG2 [18]	Improper-Markov
`bird`	0.121	0.113	0.149	0.099	0.090	**0.067**
`bridge`	0.390	0.382	0.386	0.373	0.353	**0.300**
`camera`	0.323	0.318	0.348	0.310	0.293	**0.255**
`circles`	0.100	0.090	0.102	0.060	0.045	**0.030**
`crosses`	0.140	0.132	0.083	0.110	0.027	**0.027**
`goldhill1`	0.371	0.364	0.359	0.353	0.321	**0.280**
`horiz`	0.075	0.070	0.078	0.022	0.018	**0.004**
`lena1`	0.254	0.243	0.217	0.216	0.169	**0.141**
`montage`	0.176	0.165	0.164	0.163	0.114	**0.087**
`slope`	0.091	0.083	0.096	0.056	0.038	**0.021**
`squares`	0.005	0.004	0.076	0.010	0.016	**0.003**
`text`	0.468	0.465	0.301	0.468	**0.229**	0.280
avg.	0.209	0.202	0.197	0.187	0.143	**0.125**

Improper-Markov outperforms the other methods from the perspective of average coding rates. The effect of the improper quadtree is probably amplified because the number of parameters for each block is increased. However, JBIG2 [18] still outperforms our algorithms only for `text`. We consider it is because JBIG2 [18] is designed for text images such as faxes in contrast to our general-purpose algorithm. Note that our algorithm has room for improvement by tuning the hyperparameters α and β of the beta distribution for each of $\theta^m_{s;0000}, \theta^m_{s;0001}, \dots, \theta^m_{s;1111}$.

5.4. Experiment 4

Through Experiment 4, we show our method is applicable to grayscale images. Herein, we assume two types of stochastic generative models $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ for the block of the proper quadtree and the improper quadtree. The first one is the i.i.d. Gaussian distribution $\mathcal{N}(v_t|\mu^m_s, (\lambda^m_s)^{-1})$. In this case, $\boldsymbol{\theta}^m_s$ can be regarded as $\{\mu^m_s, \lambda^m_s\} \in \mathbb{R} \times \mathbb{R}_{>0}$. The second one is the two-dimensional autoregressive (AR) model [7] of the neighboring four pixels, i.e., $\mathcal{N}(v_t|\tilde{\boldsymbol{v}}_{t-1}^\top \boldsymbol{w}^m_s, (\tau^m_s)^{-1})$, where $\tilde{\boldsymbol{v}}_{t-1} = (v_{t-w-1}, v_{t-w}, v_{t-w+1}, v_{t-1})^\top$. (If the indices go out of the image, we use the nearest past pixel in Manhattan distance.) In this case, $\boldsymbol{\theta}_s^m$ can be regarded as $\{\boldsymbol{w}_s^m, \tau_s^m\} \in \mathbb{R}^4 \times \mathbb{R}_{>0}$. For both models, v_t is normalized and quantized into $\mathcal{V} = \{0, 1, \ldots, 255\}$ in a similar manner to [7]. The prior distributions for each model are assumed to be the Gauss–gamma distributions $\mathcal{N}(\mu_s^m | \mu_0, (\kappa_0 \lambda_s)^{-1})\mathrm{Gam}(\lambda_s^m | \alpha_0, \beta_0)$ and $\mathcal{N}(\boldsymbol{w}_s^m | \boldsymbol{\mu}_0, (\tau_s^m \boldsymbol{\Lambda}_0)^{-1})\mathrm{Gam}(\tau_s^m | \alpha_0, \beta_0)$, where $\mu_0 = 0$, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\kappa_0 = 0.01$, $\boldsymbol{\Lambda}_0 = 0.01\boldsymbol{I}$, $\alpha_0 = 1.0$, $\beta_0 = 0.0001$. Here, $\boldsymbol{I}$ is the identity matrix. The results are shown in Table 4. (The values for previous studies [2,4,20,21] are cited from [21].)

Table 4. The coding rates for the grayscale images from [19] (bit/pel).

Images	JPEG2000 [20]	JPEG-LS [2]	MRP [4]	Vanilc [21]	Proper-Gaussian	Improper-Gaussian	Proper-AR	Improper-AR
`bird`	3.630	3.471	3.238	**2.749**	4.086	4.055	3.461	3.422
`bridge`	6.012	5.790	**5.584**	5.596	6.353	6.294	5.696	5.678
`camera`	4.570	4.314	3.998	**3.995**	4.651	4.589	4.163	4.121
`circles`	0.928	0.153	0.132	**0.043**	1.190	0.915	1.030	0.826
`crosses`	1.066	0.386	0.051	**0.016**	1.603	1.240	0.898	0.625
`goldhill1`	5.516	5.281	5.098	**5.090**	5.796	5.738	5.220	5.196
`horiz`	0.231	0.094	0.016	**0.015**	1.091	0.922	0.279	0.216
`lena1`	4.755	4.581	4.189	**4.123**	5.312	5.259	4.433	4.394
`montage`	2.983	2.723	**2.353**	2.363	3.818	3.734	2.940	2.850
`slope`	1.342	1.571	**0.859**	0.960	3.721	3.683	1.728	1.602
`squares`	0.163	0.077	0.013	**0.007**	0.335	0.205	0.323	0.202
`text`	4.215	1.632	3.175	**0.621**	4.310	3.691	4.176	3.732
Whole avg.	2.951	2.506	2.392	**2.132**	3.522	3.360	2.862	2.739
Natural avg.	4.897	4.687	4.421	**4.311**	5.240	5.187	4.595	4.562
Artificial avg.	1.561	0.948	0.943	**0.575**	2.295	2.056	1.625	1.436

The coding rates of the proper-quadtree-based algorithm are improved by our proposed method for all the images in this data set and for both settings of the stochastic generative model assumed within blocks. This indicates the superiority of the improper-quadtree-based model to the proper-quadtree-based model. The method labeled by Improper-AR showed an average coding rate lower than JPEG2000, averaging for the whole images. It also showed an average coding rate lower than JPEG-LS, averaging for the natural images. Although it does not outperform recent methods such as MRP and Vanilc, we consider this is because of the suitability of the stochastic generative model within blocks, which is out of the scope of this paper.

6. Conclusions

We proposed a novel stochastic model based on the improper quadtree, so that our model effectively represents the variable block size segmentation of images. Then, we constructed a Bayes code for the proposed stochastic model. Moreover, we introduced an algorithm to implement it in polynomial order of data size without loss of optimality. Some experiments both on synthetic and real images demonstrated the flexibility of our stochastic model and the efficiency of our algorithm. As a result, the derived algorithm showed a better average coding rate than that of JBIG2 [18].

Author Contributions: Conceptualization, Y.N. and T.M.; methodology, Y.N.; software, Y.N.; validation, Y.N. and T.M.; formal analysis, Y.N. and T.M.; investigation, Y.N. and T.M.; resources, Y.N.; data curation, Y.N.; writing—original draft preparation, Y.N.; writing—review and editing, Y.N. and T.M.; visualization, Y.N.; supervision, T.M.; project administration, T.M.; funding acquisition, T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by JSPS KAKENHI Grant Numbers 17K06446, JP19K04914 and 22K02811.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: http://links.uwaterloo.ca/Repository.html (accessed on 18 August 2022).

Acknowledgments: We would like to thank the members of Matsushima laboratory for their meaningful discussions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Validity of Proposed Algorithm

Validity of Prior Distribution for Models

Although a general proof for any rooted trees is described in [15] (please see also a preprint for the full version of [15] uploaded on arXiv.), in the following, we briefly describe a proof restricted for our model to make this paper self-contained.

$$\sum_{m\in\mathcal{M}} p(m) = \underbrace{\sum_{m\in\mathcal{M}} \prod_{s\in\mathcal{S}^m} \eta_s(\boldsymbol{z}_s^m)}_{(a)} = \sum_{\boldsymbol{z}_{s_\lambda}\in\{0,1\}^4} \sum_{m\in\mathcal{M}:\boldsymbol{z}_{s_\lambda}^m=\boldsymbol{z}_{s_\lambda}} \prod_{s\in\mathcal{S}^m} \eta_s(\boldsymbol{z}_s^m) \tag{A1}$$

$$= \sum_{\boldsymbol{z}_{s_\lambda}\in\{0,1\}^4} \eta_{s_\lambda}(\boldsymbol{z}_{s_\lambda}) \sum_{m\in\mathcal{M}:\boldsymbol{z}_{s_\lambda}^m=\boldsymbol{z}_{s_\lambda}} \prod_{s\in\mathcal{S}^m\setminus\{s_\lambda\}} \eta_s(\boldsymbol{z}_s^m) \tag{A2}$$

$$= \sum_{\boldsymbol{z}_{s_\lambda}\in\{0,1\}^4} \eta_{s_\lambda}(\boldsymbol{z}_{s_\lambda}) \prod_{s'\in\mathrm{Ch}(s_\lambda)} \left(\underbrace{\sum_{m\in\mathcal{M}^{s'}} \prod_{s\in\mathcal{S}^m} \eta_s(\boldsymbol{z}_s^m)}_{(b)}\right)^{z_{s_\lambda s'}} \tag{A3}$$

In (A3), $\mathcal{M}^{s'}$ denotes the set of subtrees whose root node is s'. The factorization from (A2) to (A3) is because m in (A2) is determined by the subtrees m' whose root nodes are in $\mathrm{Ch}(s_\lambda)$. The same idea is also detailed in Figure 4 in the preprint of the full version of [15], which is uploaded on arXiv. The underbraced parts (a) and (b) have the same structure except for the depth of the root node. We represent them by $\phi(s)$, which is a function of the root node s of the subtree.

Subsequently, we have

$$\phi(s) = \begin{cases} \sum_{\boldsymbol{z}\in\{0,1\}^4} \eta_s(\boldsymbol{z}) = 1, & |s| = 1, \\ \sum_{\boldsymbol{z}\in\{0,1\}^4} \eta_s(\boldsymbol{z}) \prod_{s'\in\mathrm{Ch}(s)} (\phi(s'))^{z_{ss'}}, & \text{otherwise.} \end{cases} \tag{A4}$$

Therefore, the following holds by recursively substituting $\phi(s)$ from the leaf nodes.

$$\sum_{m\in\mathcal{M}} p(m) = \phi(s_\lambda) = 1 \tag{A5}$$

Proof of Lemma 1. Let $R(s, \boldsymbol{z}_s)$ denote $\bigcup_{s'\in\mathrm{Ch}(s):z_{ss'}=0} s'$, which is a region where v_t is generated according to $\eta_s(\boldsymbol{z}_s)$ when $(x(t), y(t)) \in s$. Then,

$$p(v_t|v^{t-1}, m) \propto \int p(v_t|v^{t-1}, \theta_s^m) \int p(\boldsymbol{\theta}^m|m) p(v^{t-1}|\boldsymbol{\theta}^m, m) \mathrm{d}\boldsymbol{\theta}_{\setminus s}^m \mathrm{d}\theta_s^m \tag{A6}$$

$$\propto \int p(v_t|v^{t-1}, \theta_s^m) p_s(\theta_s^m) \prod_{i\in\{i'\leq t|(x(i'),y(i'))\in R(s,\boldsymbol{z}_s)\}} p(v_i|v^{i-1}, \theta_s^m) \mathrm{d}\theta_s^m, \tag{A7}$$

where $\propto$ means that the left-hand side is proportional to the right-hand side, regarding the variables except for v_t as constant, and $\boldsymbol{\theta}^m_{\setminus s}$ denotes the parameters $\boldsymbol{\theta}^m$ except for θ^m_s. Formula (A7) does not depend on m but $(s, \boldsymbol{z}_s)$. □

Proof of Theorem 1. Although Theorem 1 is proved by applying Corollary 2 of Theorem 7 in [15] (please see also the preprint for the full version of [15] uploaded on arXiv), in the following, we briefly describe a proof restricted to our model to make this paper self-contained.

Theorem 1 will be proved by induction. First, we assume

$$p(m|v^{t-1}) = \prod_{s \in \mathcal{S}^m} \eta_s(\boldsymbol{z}^m_s|v^{t-1}), \tag{A8}$$

which is true for $t = 0$ because of Assumption 2 and will be proved later for $t > 0$. In addition, we define the following function to simplify the notation.

$$f(v_t|v^{t-1}, s, \boldsymbol{z}_s) := \begin{cases} \tilde{q}(v_t|v^{t-1}, s, \boldsymbol{0}), & s = \{(x(t), y(t))\}, \\ \tilde{q}(v_t|v^{t-1}, s, \boldsymbol{z}_s), & {}^{\exists}s' \in \mathrm{Ch}(s) \text{ s.t. } (s' \ni (x(t), y(t))) \wedge (z_{ss'} = 0), \\ 1, & \text{otherwise.} \end{cases} \tag{A9}$$

Using this notation, we can represent $p(v_t|v^{t-1}, m)$ as follows.

$$p(v_t|v^{t-1}, m) = \prod_{s \in \mathcal{S}^m} f(v_t|v^{t-1}, s, \boldsymbol{z}^m_s). \tag{A10}$$

(Equations (A9) and (A10) correspond to Conditions 4 and 3 in [15] , respectively. If we accept this fact, Theorem 1 is immediately proved by applying Corollary 2 in [15].) By using (A10), we have

$$p(v_t|v^{t-1}) = \sum_{m \in \mathcal{M}} p(m|v^{t-1}) p(v_t|v^{t-1}, m) = \sum_{m \in \mathcal{M}} \prod_{s \in \mathcal{S}^m} \eta_s(\boldsymbol{z}^m_s|v^{t-1}) f(v_t|v^{t-1}, s, \boldsymbol{z}^m_s). \tag{A11}$$

Since the right-hand side of (A11) has a similar form to the underbraced part (a) in (A2), we can define a recursive function $q(v_t|v^{t-1}, s)$ that satisfies

$$p(v_t|v^{t-1}) = q(v_t|v^{t-1}, s_\lambda), \tag{A12}$$

where

$$q(v_t|v^{t-1}, s) := \begin{cases} f(v_t|v^{t-1}, s, \boldsymbol{0}), & |s| = 1 \\ \sum_{\boldsymbol{z}_s \in \{0,1\}^4} \eta_s(\boldsymbol{z}_s|v^{t-1}) f(v_t|v^{t-1}, s, \boldsymbol{z}_s) & \\ \qquad \times \prod_{s' \in \mathrm{Ch}(s)} q(v_t|v^{t-1}, s')^{z_{ss'}}, & \text{otherwise} \end{cases} \tag{A13}$$

By substituting (A9), $q(v_t|v^{t-1}, s) = 1$ holds for $s \not\ni (x(t), y(t))$ (or equivalently for $s \notin \mathcal{S}_t$). Therefore, we need not calculate (A13) for $s \notin \mathcal{S}_t$ and (9) will be derived by substituting (A9) again for $s \in \mathcal{S}_t$.

Lastly, we will prove (A8). Using (A9), the updating Formula (10) can be generally represented as follows.

$$\eta_s(\boldsymbol{z}_s|v^t) = \begin{cases} \eta_s(\boldsymbol{z}_s), & t = -1, \\ \eta_s(\boldsymbol{z}_s|v^{t-1}), & t \geq 0 \wedge |s| = 1, \\ \frac{\eta_s(\boldsymbol{z}_s|v^{t-1}) f(v_t|v^{t-1}, s, \boldsymbol{z}_s) \prod_{s' \in \mathrm{Ch}(s)} q(v_t|v^{t-1}, s')^{z_{ss'}}}{q(v_t|v^{t-1}, s)}, & \text{otherwise.} \end{cases} \tag{A14}$$

By substituting the above general updating formula,

$$\prod_{s\in\mathcal{S}^m}\eta_s(z_s|v^t)=\prod_{s\in\mathcal{I}^m}\frac{\eta_s(z_s|v^{t-1})f(v_t|v^{t-1},s,z_s)\prod_{s'\in\mathrm{Ch}(s)}q(v_t|v^{t-1},s')^{z_{ss'}}}{q(v_t|v^{t-1},s)}$$

$$\times\prod_{s\in\mathcal{L}^m}\frac{\eta_s(z_s|v^{t-1})f(v_t|v^{t-1},s,z_s)\prod_{s'\in\mathrm{Ch}(s)}q(v_t|v^{t-1},s')^{0}}{q(v_t|v^{t-1},s)} \tag{A15}$$

$$=\frac{1}{q(v_t|v^{t-1},s_\lambda)}\prod_{s\in\mathcal{S}^m}\eta_s(z_s|v^{t-1})\prod_{s\in\mathcal{S}^m}f(v_t|v^{t-1},s,z_s) \tag{A16}$$

$$=\frac{p(m|v^{t-1})p(v_t|v^{t-1},m)}{p(v_t|v^{t-1})}=p(m|v^t) \tag{A17}$$

In the above operation, (A15) was a telescoping product, i.e., $q(v_t|v^{t-1},s)$ appeared at once in each of the denominator and the numerator. Therefore, we canceled them except for $q(v_t|v^{t-1},s_\lambda)$. (A16) is because of (A8), (A10) and (A11), where (A8) and (A11) are the induction hypotheses. □

References

1. Nakahara, Y.; Matsushima, T. A Stochastic Model for Block Segmentation of Images Based on the Quadtree and the Bayes Code for It. *Entropy* **2021**, *23*, 991. [CrossRef] [PubMed]
2. Weinberger, M.J.; Seroussi, G.; Sapiro, G. The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS. *IEEE Trans. Image Process.* **2000**, *9*, 1309–1324. [CrossRef] [PubMed]
3. Wu, X.; Memon, N. Context-based, adaptive, lossless image coding. *IEEE Trans. Commun.* **1997**, *45*, 437–444. [CrossRef]
4. Matsuda, I.; Ozaki, N.; Umezu, Y.; Itoh, S. Lossless coding using variable block-size adaptive prediction optimized for each image. In Proceedings of the 2005 13th European Signal Processing Conference, Antalya, Turkey, 4–8 September 2005; pp. 1–4.
5. Huffman, D.A. A Method for the Construction of Minimum-Redundancy Codes. *Proc. IRE* **1952**, *40*, 1098–1101. [CrossRef]
6. Rissanen, J.; Langdon, G. Universal modeling and coding. *IEEE Trans. Inf. Theory* **1981**, *27*, 12–23. [CrossRef]
7. Nakahara, Y.; Matsushima, T. Autoregressive Image Generative Models with Normal and t-distributed Noise and the Bayes Codes for Them. In Proceedings of the 2020 International Symposium on Information Theory and Its Applications (ISITA), Kapolei, HI, USA, 24–27 October 2020; pp. 81–85.
8. Nakahara, Y.; Matsushima, T. Hyperparameter Learning of Stochastic Image Generative Models with Bayesian Hierarchical Modeling and Its Effect on Lossless Image Coding. In Proceedings of the 2021 IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021.
9. Nakahara, Y.; Matsushima, T. Bayes code for two-dimensional auto-regressive hidden Markov model and its application to lossless image compression. In Proceedings of the International Workshop on Advanced Imaging Technology (IWAIT) 2020, Yogyakarta, Indonesia, 1 June 2020; SPIE: Bellingham, WA, USA, 2020; Volume 11515, pp. 330–335. [CrossRef]
10. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
11. Matsushima, T.; Inazumi, H.; Hirasawa, S. A class of distortionless codes designed by Bayes decision theory. *IEEE Trans. Inf. Theory* **1991**, *37*, 1288–1293. [CrossRef]
12. Clarke, B.S.; Barron, A.R. Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Inf. Theory* **1990**, *36*, 453–471. [CrossRef]
13. Matsushima, T.; Hirasawa, S. Reducing the space complexity of a Bayes coding algorithm using an expanded context tree. In Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 719–723. [CrossRef]
14. Sullivan, G.J.; Ohm, J.; Han, W.; Wiegand, T. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1649–1668. [CrossRef]
15. Nakahara, Y.; Saito, S.; Kamatsuka, A.; Matsushima, T. Probability Distribution on Rooted Trees. In Proceedings of the 2022 IEEE International Symposium on Information Theory, Espoo, Finland, 26 June–1 July 2022.
16. Martín, G. Range encoding: An algorithm for removing redundancy from a digitised message. In Proceedings of the Video and Data Recording Conference, Southampton, UK, 24–27 July 1979; pp. 24–27.
17. Kuhn, M. JBIG-KIT. Available online: https://www.cl.cam.ac.uk/~mgk25/jbigkit/ (accessed on 24 July 2022).
18. Langley, A. jbig2enc. Available online: https://github.com/agl/jbig2enc (accessed on 24 July 2022).
19. Image Repository of the University of Waterloo. Available online: http://links.uwaterloo.ca/Repository.html (accessed on 8 November 2021).

20. Skodras, A.; Christopoulos, C.; Ebrahimi, T. The JPEG 2000 still image compression standard. *IEEE Signal Process. Mag.* **2001**, *18*, 36–58. [CrossRef]
21. Weinlich, A.; Amon, P.; Hutter, A.; Kaup, A. Probability Distribution Estimation for Autoregressive Pixel-Predictive Image Coding. *IEEE Trans. Image Process.* **2016**, *25*, 1382–1395. [CrossRef] [PubMed]