IEICE TRANS. FUNDAMENTALS, VOL.E85–A, NO.1 JANUARY 2002

PAPER

Edge-Based Image Synthesis Model and Its Synthesis Function Design by the Wavelet Transform

Makoto NAKASHIZUKA[†a)], **Hidetoshi OKAZAKI**[†], *and* **Hisakazu KIKUCHI**[††], ***Regular Members***

SUMMARY In this paper, a new image synthesis model based on a set of wavelet bases is proposed. In the proposed model, images are approximated by the sum of synthesis functions that are translated to image edge positions. By applying the proposed model to sketch-based image coding, no iterative image recovery procedure is required for image decoding. In the design of the synthesis functions, we define the synthesis functions as a linear combination of wavelet bases. The coefficients for wavelet bases are obtained from an iterative procedure. The vector quantization is applied to the vectors of the coefficients to limit the number of the synthesis functions. We apply the proposed synthesis model to the sketch-based image coding. Image coding experiments by eight synthesis functions and a comparison with the orthogonal transform methods are also given.

key words: *edge detection, multi-scale image analysis, wavelet transform, image coding*

1. Introduction

Sketch-based image coding is one of the very low-bit rate image compression techniques [1]–[5]. In the sketch-based image coding [1], [3], [4], images are represented in the form of edge geometry and intensity differences across edges. In decoding, images are recovered by an iterative procedure. The iteration minimizes a cost function that is defined by a constraint to the smoothness in intensity changes on planar regions.

The wavelet transform [13] is employed for another approach to image coding based on image edges [5], [7]. In Ref. [5], the wavelet maxima representation is defined for characterization of image edges. If the basic wavelet function corresponds to the first-order derivative of a smoothing function, the wavelet maxima indicate the positions of edges and describe multiscale behaviors of edges. In Ref. [7], the two-dimensional wavelet transform is sampled at the watershed and watercourse lines which appear around image edges for image coding. In image recovery from adaptive sampling of wavelet transforms, POCS (Projection Onto Convex Sets) algorithm [9] is applied. The procedure of image recovery corresponds to a sequence of successive projections between two convex sets that are defined by given sampling points of wavelet transforms. Iterative procedures are required to recover the original image in the above two sketch-based image coding methods.

In this paper, we propose a new edge-based image representation [8]. By the representation, an approximation $g(m,n)$ of the original image is obtained from two-parts: a low-passed image and high-frequency components around image edges by the following form

$$g(m,n) = q(m,n) + \sum_{(m_c,n_c)\in C} h_{(m_c,n_c)}(m-m_c, n-n_c) \quad (1)$$

where $q(m,n)$ denotes the low-passed image that is obtained from the original image. The high-frequency components are recovered from the sum of synthesis functions $h_{(m_c,n_c)}$ that are located on edge positions (m_c,n_c). The set of two-dimensional coordinates C consists of all edge positions that are detected from the original image. The approximation of the original image is reconstructed by only additions of the low-passed image and the synthesis functions along image edges. We refer to the proposed image synthesis model as an *edge-based image synthesis model*. Since the approximation of the original image can be reconstructed by a linear combination of the functions, the computational cost for image decoding can be smaller than that of the other edge-based image coding.

In Sect. 2, the wavelet transform and wavelet maxima representation that has been employed to previous works of the sketch based image coding are explained. Next, we propose the design of synthesis functions $h_{(m_c,n_c)}$ in Eq. (1) from the wavelet bases. The synthesis function design is split into two stages. The first stage is the iterative procedure to get the coefficient vector that can reconstruct the approximation of the original image by single inverse wavelet transform. The second stage is the quantization of the coefficients to wavelet bases to reduce the number of the synthesis functions. The first and second stage are explained in Sect. 3 and Sect. 4 respectively. In Sect. 5, the proposed model is applied to sketch-based image coding and is compared with some transform coding methods. The simple image processing in the code domain that is ob-

Manuscript received February 5, 2001.
Manuscript revised July 27, 2001.
Final manuscript received September 5, 2001.

[†]The authors are with the Graduate School of Bio-application and Systems Engineering, Tokyo University of Agriculture and Technology, Koganei-shi, 184-0012 Japan.
[††]The author is with the Faculty of Engineering, Niigata University, Niigata-shi, 950-2181 Japan.
a) E-mail: nkszk@cc.tuat.ac.jp

tained by the edge-based image synthesis model is also demonstrated.

2. Wavelet Maxima Representation

The discrete dyadic wavelet transform [5], [13] is defined by the inner products between the wavelet bases and an image. Usually, the one-dimensional discrete wavelet transform is computed from the discrete-time filterbanks. In two-dimensional case, the filterbanks are applied to the original image along horizontal and vertical direction. The two-dimensional wavelet transform has the property of spatial orientation selectivity. We select the two-dimensional wavelet transform that can be computed by the discrete filterbank in Fig. 1 for the edge detection and analysis [5]. In this filter bank, the horizontal and vertical edges are detected from the filter outputs separately. The J-th scale discrete dyadic wavelet transform of the original image $\{f(m,n)\}_{0\le m<N_x,0\le n<N_y}$ is represented as set of the sequences:

$$\{\{W_j^H f\}_{1\le j\le J},\{W_j^V f\}_{1\le j\le J},\{S_J f\}\} \tag{2}$$

where we denote $W_j^H f$ and $W_j^V f$ as

$$W_j^H f=\{W_j^H f(m,n)\}_{0\le m<N_x,0\le n<N_y} \tag{3}$$

for the horizontal direction wavelet transform and

$$W_j^V f=\{W_j^V f(m,n)\}_{0\le m<N_x,0\le n<N_y}, \tag{4}$$

for the vertical direction wavelet transform respectively. $S_J f$ denotes the lowest frequency component that is complement to the wavelet transform as

$$S_j f=\{S_j f(m,n)\}_{0\le m<N_x,0\le n<N_y}. \tag{5}$$

For convenient, we represent the two-dimensional discrete dyadic wavelet transform as the vector that is defined by the multiplication between the transform matrix $\mathbf{W}$ and the vector $\mathbf{f}$ which represents the image pixel by the form as

$$\begin{aligned}\mathbf{f} \;=\; &[f(0,0)\cdots f(N_x-1,0)\\ &f(0,1)\cdots f(N_x-1,N_y-1)]^T.\end{aligned} \tag{6}$$

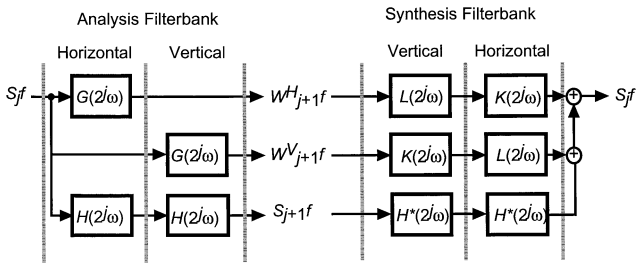


Fig. 1 Filter bank structure for two-dimensional discrete dyadic wavelet transform for two-directional decomposition (* denotes complex conjugate).

The wavelet transform is denoted as

$$\mathbf{Wf}. \tag{7}$$

In the form of (7), the elements of $\mathbf{Wf}$ is represented as:

$$\begin{aligned}\mathbf{Wf}=[&\mathbf{w}_{(0,0)}^H\cdots\mathbf{w}_{(N_x-1,0)}^H,\mathbf{w}_{(0,1)}^H\cdots\mathbf{w}_{(N_x-1,N_y-1)}^H,\\ &\mathbf{w}_{(0,0)}^V\cdots\mathbf{w}_{(N_x-1,0)}^V,\mathbf{w}_{(0,1)}^V\cdots\mathbf{w}_{(N_x-1,N_y-1)}^V,\mathbf{s}]^T\end{aligned} \tag{8}$$

where

$$\mathbf{w}_{(m,n)}^H=[W_1^H f(m,n)\cdots W_J^H f(m,n)] \tag{9}$$

for the horizontal wavelet transform and

$$\mathbf{w}_{(m,n)}^V=[W_1^V f(m,n)\cdots W_J^V f(m,n)] \tag{10}$$

for the vertical wavelet transform. The component $\mathbf{s}$ in Eq. (8) denotes the lowest frequency component as

$$\begin{aligned}\mathbf{s} \;=\; &[S_J f(0,0)\cdots S_J f(N_x-1,0)\\ &S_J f(0,1)\cdots S_J f(N_x-1,N_y-1)].\end{aligned} \tag{11}$$

Let assume that the inverse wavelet transform that is performed by reconstruction filter bank in Fig. 1 is now also represented as the matrix $\tilde{\mathbf{W}}$. If the pair of the course and inverse wavelet transform achieves perfect reconstruction, the inverse wavelet transform satisfies

$$\mathbf{f}=\tilde{\mathbf{W}}\mathbf{W}\mathbf{f}. \tag{12}$$

The wavelet maxima representation [5] are given from the maxima positions of the amplitude of the discrete dyadic wavelet transform. If the basic wavelet is defined as the first derivative of a smoothing function, the modulus maxima appear at image edge positions and characterize the edges by its multiscale nature [5]. Now, let assume that the set of the coordinates where modulus maxima appear on the j-th scale wavelet transform are defined as C_j^H for horizontal wavelet transform and C_j^V for vertical wavelet transform, respectively. The maxima representation $\mathbf{M}$ can be represented as the form:

$$\mathbf{M}=\mathbf{Z}\mathbf{W}\mathbf{f} \tag{13}$$

where $\mathbf{Z}$ is a diagonal matrix of which elements are zero or one. The number of rows and columns of $\mathbf{Z}$ are equal to the number of the dimension of $\mathbf{Wf}$. Only the diagonal elements of $\mathbf{Z}$ that are allocated to the maxima positions that included in C_j^H or C_j^V and the lowest frequency component $\mathbf{s}$ are one. The image recovery from the maxima representation is the problem that the image recovery from above incomplete set of the wavelet transform. Obviously, it is impossible to recover the original image by the inverse wavelet transform. In Ref. [5], POCS [9] is applied to recover the original image from the maxima representation. For

POCS algorithm, the constraints for the recovered image are described by the convex sets. After convergence of the iterative projections, we can get the recovered image that satisfies all constraints.

Two closed convex sets are defined by the wavelet maxima representation in the vector space of which dimension corresponds to the original wavelet transform $\mathbf{Wf}$ [5]. One convex set $\mathbf{V}$ is the set that consists of wavelet transforms of all possible images. The projection P_V onto the set $\mathbf{V}$ is realized by the pair of course and inverse wavelet transform as $\mathbf{W}\tilde{\mathbf{W}}$. Other convex set $\mathbf{\Gamma}$ is the set of all vectors that preserve the values of wavelet maxima and the lowest frequency component. The projection operator P_Γ onto Γ is the operator that replace the values of the wavelet transform with the known values of the wavelet transform which are preserved in the maxima representation. Since the both sets are convex, the iteration guarantees the convergence to the wavelet transform which exists in the intersection of two sets and is the closest to the initial wavelet transform [5], [9]. In Ref. [5], the condition of the smoothness of the wavelet transform is applied to the iteration to avoid irregular oscillations between two consecutive maxima. In Ref. [6], the wavelet maxima representation is extended to the wavelet extrema representation which includes modulus minima of the wavelet transform to achieve higher reconstruction precision.

The compact image coding is achieved in Ref. [5] by selection of the significant edges which are detected from the single scale. The iterative procedure of which iteration includes the pair of the course and inverse wavelet transform is required for decoding from the image coding based on the wavelet maxima. So, the computational cost for the image reconstruction is larger than other orthogonal transforms coding.

3. Image Decomposition Based on Wavelet Maxima

In this section, we derive an edge-based image representation $\mathbf{d}$ that gives the approximation $\mathbf{g}$ of the original image $\mathbf{f}$ by a single inverse wavelet transform as:

$$\mathbf{g} = \tilde{\mathbf{W}}\mathbf{d}. \tag{14}$$

Now, the elements of the coefficient vector $\mathbf{d}$ is defined as

$$\begin{aligned}\mathbf{d} = [&\mathbf{d}^H_{(0,0)} \cdots \mathbf{d}^H_{(N_x-1,0)}, \mathbf{d}^H_{(0,1)} \cdots \mathbf{d}^H_{(N_x-1,N_y-1)},\\ &\mathbf{d}^V_{(0,0)} \cdots \mathbf{d}^V_{(N_x-1,0)}, \mathbf{d}^V_{(0,1)} \cdots \mathbf{d}^V_{(N_x-1,N_y-1)}, \mathbf{u}]^T.\end{aligned} \tag{15}$$

where

$$\mathbf{d}^H_{(m,n)} = [D^H_1(m,n) \cdots D^H_J(m,n)] \tag{16}$$

for the horizontal direction wavelet basis and

$$\mathbf{d}^V_{(m,n)} = [D^V_1(m,n) \cdots D^V_J(m,n)] \tag{17}$$

for vertical direction. Both $\mathbf{d}^H_{(m,n)}$ and $\mathbf{d}^V_{(m,n)}$ are referred as the partial coefficient vectors. The component $\mathbf{u}$ denotes the coefficients which correspond to the lowest frequency component of the image as

$$\begin{aligned}\mathbf{u} = [&U(0,0) \cdots U(N_x-1,0),\\ &U(0,1) \cdots U(N_x-1,N_y-1)].\end{aligned} \tag{18}$$

To describe the characteristic of image edges, non-zero elements of $D^H_j(m,n)$ and $D^V_j(m,n)$ only exist on the maxima positions that included in $\{C^H_j\}_{1\le j\le J}$ and $\{C^V_j\}_{1\le j\le J}$. So, the coefficient vector $\mathbf{d}$ satisfies

$$\mathbf{d} = \mathbf{Z}\mathbf{d} \tag{19}$$

where $\mathbf{Z}$ is a diagonal matrix which defines sampling points of the wavelet transform for the wavelet maxima representation in (13). The number of the non-zero elements in $\mathbf{d}$ depends on the number of the maxima positions and is $N_x \times N_y + A$ where A denotes the number of entire wavelet maxima. The number of unknown elements in $\mathbf{d}$ is larger than the number of pixels in the original image. There is no unique solution of $\mathbf{d}$. The singular value decomposition [10] can obtain the least square vector $\mathbf{d}$ that minimizes the norm of the error between the original image $\mathbf{f}$ and the approximation $\mathbf{g}$. However, the singular value decomposition requires operations to the matrix of which size is $(N_xN_y + A) \times (N_xN_y)$. The number of elements in the matrix and the computational cost for the singular value decomposition are proportional to the square of the number of pixels in the original image. It is impractical to apply the singular value decomposition to the such huge matrix which consists of all wavelet bases that are allocated to the maxima positions.

To reduce the computational costs for derivation of the coefficient vector $\mathbf{d}$, we propose an iterative algorithm based on a pair of inverse and course wavelet transform. The iterative algorithm produce the coefficient vector $\mathbf{d}$ that satisfies following three conditions:

condition 1. it satisfies $\mathbf{d} = \mathbf{Z}\mathbf{d}$
condition 2. it minimizes the error between the wavelet transform of the original image and the approximation of the original image $|\mathbf{Wf} - \mathbf{Wg}|$
condition 3. it minimizes the norm of the coefficient vector $|\mathbf{d}|$ under the conditions 1 and 2.

The norm of the coefficient vector $\mathbf{d}$ is not be bounded by conditions 1 and 2. The range of amplitude of elements in the coefficient vector have to be small for finite word-length representation to reduce word-length or quantization error. We hence introduce the condition 3 to minimize the variance of amplitude of elements in $\mathbf{d}$.

To get the coefficient vector $\mathbf{d}$, we define two convex sets X and Y in the vector space of which dimension

is equal to the wavelet transform. The convex set X is defined by the condition 1 and is the set of vectors that have non-zero elements only on the positions where the wavelet transform is sampled for the wavelet maxima representation. Any elements $\mathbf{x}$ in X satisfies

$$\mathbf{x} = \mathbf{Z}\mathbf{x}. \tag{20}$$

The other convex set Y is the set of vectors that obtain the original image $\mathbf{f}$ by the inverse wavelet transform. Any elements $\mathbf{y}$ in Y satisfies

$$\mathbf{f} = \tilde{\mathbf{W}}\mathbf{y}. \tag{21}$$

The projection operator onto the convex set X is now defined as P_X and can be implemented as:

$$P_X(\mathbf{v}) = \mathbf{Z}\mathbf{v} \tag{22}$$

for any vector $\mathbf{v}$. Next, the projection operator onto the convex set Y is defined as P_Y and is represented as:

$$P_Y(\mathbf{v}) = \mathbf{v} + \mathbf{p}(\mathbf{v}) \tag{23}$$

for any vector $\mathbf{v}$. $\mathbf{p}(\mathbf{v})$ is the difference between the vector and its projected point onto Y. $\mathbf{p}(\mathbf{v})$ has to satisfy

$$\mathbf{f} = \tilde{\mathbf{W}}(\mathbf{v} + \mathbf{p}(\mathbf{v})). \tag{24}$$

since the inverse wavelet transform of every elements in Y corresponds to the original image. If any $\mathbf{p}(\mathbf{v})$ satisfy Eq. (24), then $\mathbf{W}\tilde{\mathbf{W}}\mathbf{p}(\mathbf{v})$ can also satisfy Eq. (24) instead of $\mathbf{p}(\mathbf{v})$. To achieve the projection onto the convex set Y, $\mathbf{p}(\mathbf{v})$ has to be defined as the vector of which norm is minimum while satisfying the above Eq. (24). Since a pair of the inverse and course wavelet transform $\mathbf{W}\tilde{\mathbf{W}}$ is the projector onto the set of the wavelet transforms [5] and is nonexpansive, the vector $\mathbf{p}(\mathbf{v})$ of which norm is minimum has to satisfy

$$|p(\mathbf{v})| = |\mathbf{W}\tilde{\mathbf{W}}p(\mathbf{v})|. \tag{25}$$

The vector that satisfies both conditions (24) and (25) is only the wavelet transform of the difference between the inverse wavelet transform of $\mathbf{v}$ and the original image as

$$\mathbf{p}(\mathbf{v}) = \mathbf{W}(\mathbf{f} - \tilde{\mathbf{W}}\mathbf{v}). \tag{26}$$

The vector $\mathbf{p}(\mathbf{v})$ is equal to the error between the wavelet transform of the original image and the wavelet transform $\tilde{\mathbf{W}}\mathbf{v}$.

Next, the non-expansive map T is defined by a pair of projectors as

$$T = P_X P_Y. \tag{27}$$

If there exist the intersection of two convex sets X and Y, the iterative procedure from any initial coefficient vector $\mathbf{d}^{(0)}$

$$\mathbf{d}^{(i+1)} = T(\mathbf{d}^{(i)}) \tag{28}$$

converges to one of the fixed points that are included in the intersection of X and Y [9]. The inverse wavelet transforms of every elements in the intersection obtain the original image. However, the existence of the intersection depends on the sampling points of the wavelet maxima. If the intersection does not exist, then the iteration converges to one of fixed points of T that is closest points to Y on X [11]. Any fixed point of T on X minimizes the error between $\mathbf{W}\mathbf{f}$ and $\mathbf{W}\tilde{\mathbf{W}}\mathbf{d}$. We can hence get the coefficient vector that satisfies the conditions 1 and 2. However, there is no guarantee that the given fixed point satisfies the condition 3. So, we define the cost function $\theta(\mathbf{d})$

$$\theta(\mathbf{d}) = \frac{1}{2}|\mathbf{d}|^2 \tag{29}$$

for the iteration. The iterative procedure that obtain the fixed point of T which minimizes the convex function is investigated in Ref. [12]. The cost function $\theta(\mathbf{d})$ is imposed on the iteration in Eq. (28). The iteration in Eq. (28) is modified to

$$\mathbf{d}^{(i+1)} = (1 - \lambda_{i+1})T(\mathbf{d}^{(i)}). \tag{30}$$

In this paper, the sequence λ_i is defined as $1/i$ that is found in [12]. $\mathbf{d}^{(i)}$ converges to the fixed point that minimizes the cost function θ and is unique since the cost function is a convex function [12]. We can hence get unique coefficient vector $\mathbf{d}$ that satisfies conditions 1, 2 and 3.

After the iteration, we get the coefficient vector $\mathbf{d}$ that satisfies

$$\mathbf{d} = T(\mathbf{d}) = \mathbf{Z}\mathbf{d} + \mathbf{Z}\mathbf{W}\mathbf{f} - \mathbf{Z}\mathbf{W}\tilde{\mathbf{W}}\mathbf{d}. \tag{31}$$

By substituting $\mathbf{Z}\mathbf{d} = \mathbf{d}$ and $\mathbf{g} = \tilde{\mathbf{W}}\mathbf{d}$ to Eq. (31), we get the relation between the wavelet maxima representation in (13) and the approximation $\mathbf{g}$ as

$$\mathbf{M} = \mathbf{Z}\mathbf{W}\mathbf{g}. \tag{32}$$

This relation indicates that the coefficient vector $\mathbf{d}$ obtains the approximation $\mathbf{g}$ that preserves the values of the wavelet transform at maxima positions and the lowest frequency component of the original image.

To compute iteration in Eq. (30), the coefficient vector and the wavelet transform of the original image have to be stored during the iteration. Other requirement for storage is only for the inverse and course wavelet transform. The entire size of storage for implementation of the iteration is proportional to the number of pixels of the original image. Since the computational cost for the course and inverse wavelet transform is proportional to the number of pixels, the entire computational cost is also proportional to the number of pixels in the original image and the number of the iterations. In image coding application based on the coefficient

vector **d**, the iterations are done in the encoding process. The approximation of the original image can be reconstructed by a single inverse wavelet transform as Eq. (14).

Since the inverse wavelet transform is defined as the linear combination of the synthesis wavelet bases, the approximation result (14) can be also represented as:

$$\begin{aligned} g(m,n) &= q(m,n) \\ &+ \sum_{j=1}^{J} \sum_{(m_c,n_c)\in C_j^H} \psi_j^H(m-m_c, n-n_c) D_j^H(m_c,n_c) \\ &+ \sum_{j=1}^{J} \sum_{(m_c,n_c)\in C_j^V} \psi_j^V(m-m_c, n-n_c) D_j^V(m_c,n_c) \end{aligned} \tag{33}$$

where $q(m,n)$ is the lowest frequency component as:

$$q(m,n) = \phi * U(m,n) \tag{34}$$

where * denotes the two-dimensional convolution. The two-dimensional function ψ_j^H and ψ_j^V denote the j-th scale horizontal and vertical synthesis wavelet bases which are obtained by the filter bank structure in Fig. 1. ϕ denotes the J-th scale scaling function of the wavelet transform. The approximation of the original image can be reconstructed by the addition between the low-passed image $q(m,n)$ and the synthesis wavelet bases.

For image coding application by the wavelet maxima representation, sampling points of the wavelet transform are defined on the wavelet transform at a single scale [5]. In this case, the coefficients of the wavelet transform at all scales are sampled at the same position. If the coefficients vector **d** is derived from two sets of maxima positions C^V and C^H which are detected from a single scale, Eq. (33) is expressed to

$$\begin{aligned} g(m,n) &= q(m,n) \\ &+ \sum_{(m_c,n_c)\in C^H} h_{(m_c,n_c)}^H(m-m_c, n-n_c) \\ &+ \sum_{(m_c,n_c)\in C^V} h_{(m_c,n_c)}^V(m-m_c, n-n_c) \end{aligned} \tag{35}$$

where the synthesis functions are defined as

$$h_{(m_c,n_c)}^H(m,n) = \sum_{j=1}^{J} \psi_j^H(m,n) D_j^H(m_c,n_c) \tag{36}$$

for horizontal edges and

$$h_{(m_c,n_c)}^V(m,n) = \sum_{j=1}^{J} \psi_j^V(m,n) D_j^V(m_c,n_c) \tag{37}$$

for vertical edges. Let us define the edge positions as $C = C^H \cap C^V$, the image approximation in the formula (1) is now achieved. Figure 2 shows the illustration of the image reconstruction by the proposed representation. Approximated image can be reconstructed

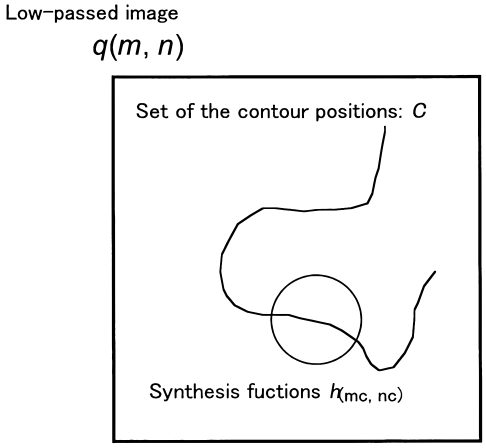


Fig. 2 Illustration of edge-based image synthesis model.

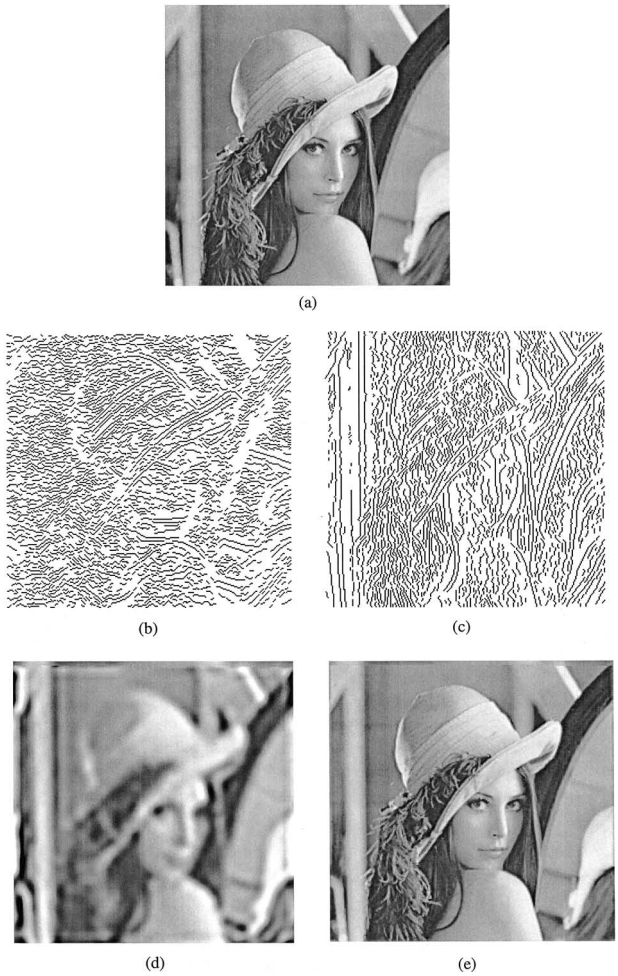


Fig. 3 (a) original image, (b) horizontal edges and (c) vertical edges for decompostition, (d) lowest frequency component and (e) approximated image.

Table 1 Filter coefficients for a quadratic spline wavelet [5].

	−3	−2	−1	0	1	2	3
H		0.125	0.375	0.375	0.125		
G			−2.0	2.0			
K	0.0078125	0.00546875	0.171875	−0.171875	−0.00546875	−0.0078125	
L	0.0078125	0.046875	0.1171875	0.65625	0.1171875	0.046875	0.0078125

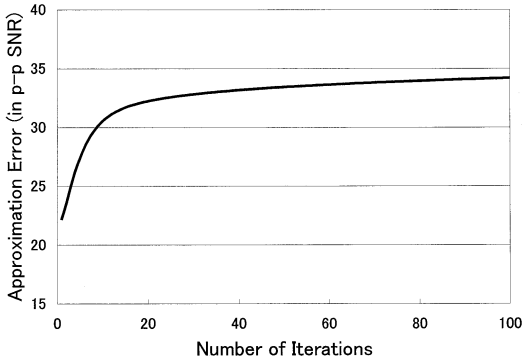


Fig. 4 Relationship between the approximation error and the number of iterations.

by additions between the lowest frequency component $q(m,n)$ and the synthesis functions $h^V_{(m_c,n_c)}(m,n)$ or $h^H_{(m_c,n_c)}(m,n)$.

Figure 3 shows the example of the image approximation by the proposed method. Table 1 shows the filter bank coefficients for the wavelet transform. Figure 3(a) shows the original image. Figures 3(b) and (c) show the positions of the edge that are detected from the wavelet maxima at scale $j = 2$. C^H_j and C^V_j are defined as Figs. 3(b) and (c) for all scales j. 12,613 and 12,350 edge positions are detected in horizontal and vertical direction wavelet, respectively. Figure 3(d) shows the lowest frequency component $\mathbf{u}$ in the coefficient vector given after 100 iterations. The approximation that is reconstructed from the coefficient vector $\mathbf{d}$ by the inverse wavelet transform is shown in Fig. 3(e). The reconstruction precision is 34.2 dB in peak-to-peak SNR. Figure 4 shows the relationship between the approximation error in p-p SNR and the number of iterations.

4. Synthesis Function Design by the Vector Quantization

In Sect. 3, the partial coefficient vectors $\{\mathbf{d}^H_{(m,n)}\}_{(m,n)\in C^H}$, $\{\mathbf{d}^H_{(m,n)}\}_{(m,n)\in C^V}$ are obtained for the image decomposition. In this section, we apply the vector quantization to the coefficient vector to limit the number of the synthesis functions. VQ is now applied for all set of the partial coefficient vectors. Since the each wavelet basis is not orthogonal to the others, it is difficult to minimize the approximation error for the original image by VQ. VQ is hence applied to minimize the error defined as

$$E = \sum_{(m,n)\in C^H} \|\mathbf{d}^H_{(m,n)} - Q[\mathbf{d}^H_{(m,n)}]\| + \sum_{(m,n)\in C^V} \|\mathbf{d}^V_{(m,n)} - Q[\mathbf{d}^V_{(m,n)}]\| \quad (38)$$

where $Q[\cdot]$ denotes the quantization. After VQ, the synthesis functions that are defined by the representative vectors and the wavelet bases approximate the partial coefficient vectors. The representative vectors that are included in the VQ codebook are defined as:

$$\mathbf{y}^H_i = [y^H_{i,1}, y^H_{i,2} \dots y^H_{i,J}]^T \quad (39)$$

for horizontal edges and

$$\mathbf{y}^V_i = [y^V_{i,1}, y^V_{i,2} \dots y^V_{i,J}]^T \quad (40)$$

for vertical edges. i indicates the index of the representative vector and is limited to the number of the code-book. By the substituting Eqs. (39) and (40) to Eq. (36) and Eq. (37), the synthesis functions of which number is limited to the number of the code-book is given. Each synthesis function is approximated by one of the representative synthesis functions

$$\xi^H_i(m,n) = \sum_{j=1}^{J} \psi^H_j(m,n) y^H_{i,j} \quad (41)$$

or

$$\xi^V_i(m,n) = \sum_{j=1}^{J} \psi^V_j(m,n) y^V_{i,j}. \quad (42)$$

Figure 5 shows the several image approximation examples after the vector quantization. We suppose that the low-bit rate image compression by the proposed approximation model and select only significant edges. The edges are detected from the maxima of the wavelet transform at $j = 2$. Moreover, we suppose that the small modulus maxima are insignificant and set the threshold to maxima. Only the maxima which is larger than the threshold are recorded to image approximation. The threshold value is set as 1/8 of maximum modulus of the wavelet transform at $j = 2$. Figures 5(a) and (b) show the vertical and horizontal maxima positions that are obtained after the thresholding operation

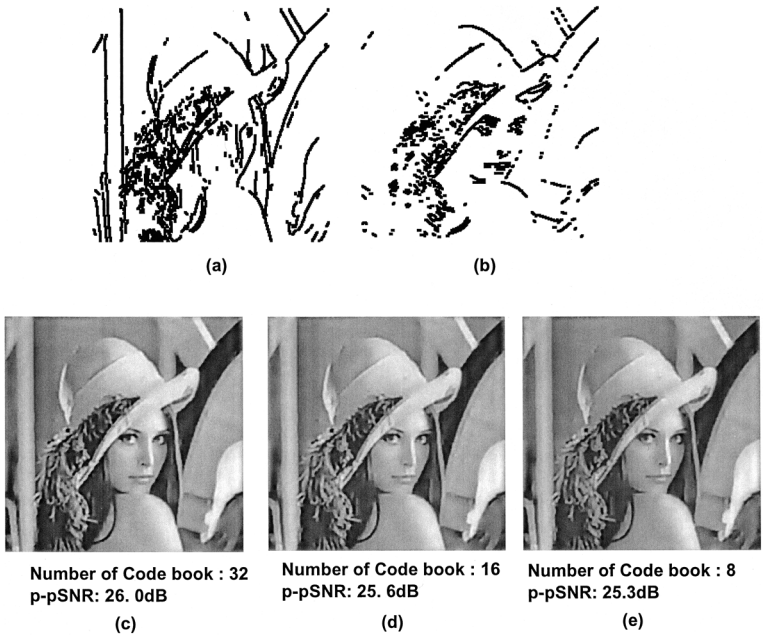


Fig. 5 (a) Detected edges from Fig. 3(a), the reconstructed image with (b) 32 (c) 16 and (d) 8 synthesis functions.

from the original image in Fig. 2(a). In the synthesis function design, we apply the LBG algorithm [14] to the partial coefficient vectors. The reconstruction result with 32, 16 and 8 synthesis functions are shown in Figs. 5(c) to (e). The smoothness or sharpness of the edges are lost by the reduction of the number of the synthesis functions. Although, 16 or 8 synthesis functions are enough to represent outlines of the original image. Next section, the entropy coding is applied to the edge geometry and the indices of the synthesis functions that are allocated to the edge positions to achieve the image compression and coding.

5. Application to Image Coding

In Sects. 3 and 4, the image approximation by the form of (1) has been described. The entire image is now represented as a linear combination of the scaling functions and the synthesis functions of which number is limited by the codebook of VQ. In this section, we apply the edge-based image synthesis model to the sketch-based image coding.

In the coding experiments, the edge-based image synthesis models are derived from three images Barbara, Lena and CG in Fig. 6. The maximum scale of the wavelet transform is set as $J = 3$. During the image encoding process, the image edges are detected from the maxima positions of the wavelet transform at scale $j = 2$ of an original image. To achieve high compression ratio, we set the thresholds on modulus maxima of the wavelet transform. Only the wavelet maxima of which amplitude are larger than 1/8 of the maximum amplitude of the wavelet transform at $j = 2$ is detected as the edges. Moreover, we suppose that short edges are insignificant for the human perception, we set a threshold on the length of the edges. The edges of which length are shorter than four pixels are eliminated in the image coding experiments. The edge positions that are detected from the original images are also shown in

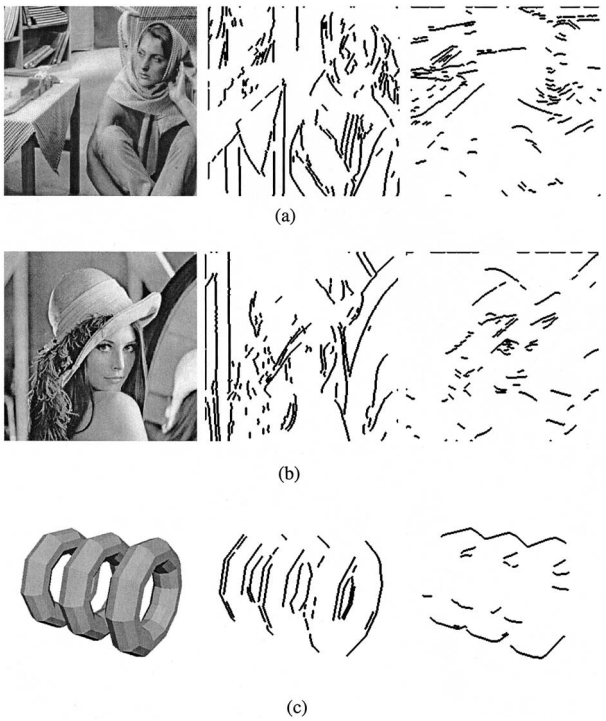


Fig. 6 Images for coding experiments and detected edges. (a) Barbara, (b) Lena and (c) CG.

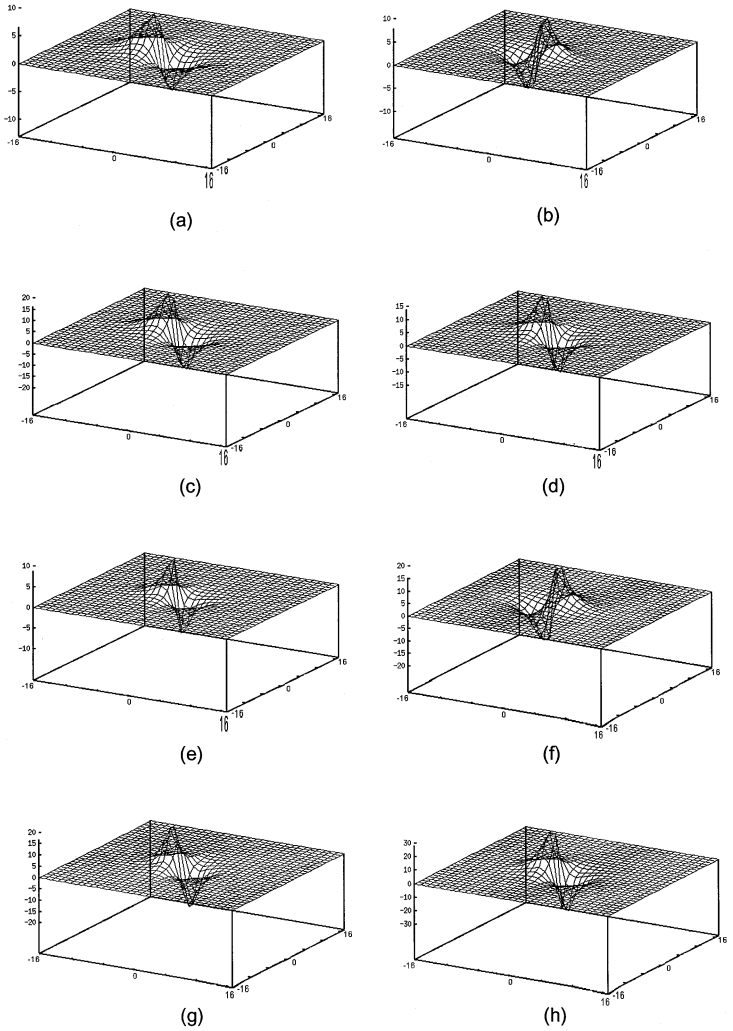


Fig. 7 Eight synthesis functions for the vertical edges in Fig. 6(a).

Fig. 6.

Every image is coded by its own different codebook. The number of representative vectors is set as eight. The eight synthesis functions for the vertical edges in Fig. 6(a) are shown in Fig. 7. In encoding the edge geometry, we employ the chain coding [16]. Since the image which consists 256×256 samples are coded, the coordinate of the start position of a edge curve is recorded in 16 bits. The chain codes for horizontal edges or vertical edges include four types of symbols (up, down, straight, stop) or (left, right, straight, stop). The examples of the chain code are shown in Fig. 8. The chain codes are recorded after run-length Huffman coding. The indices of the synthesis functions are coded along edges. Since the pixel intensity changes slowly along edges, just a single vector index that is dominant among consecutive three edge positions. Sequences of vector indexes are coding by run-length coding. The lowest frequency component $q(m, n)$ of the model is down-sampled by factor 8 and is quantized to 6 bits.

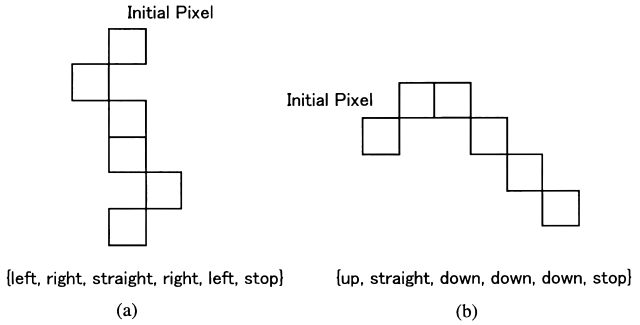


Fig. 8 Examples of edges and chain codes. (a) an example of a vertical edge and (b) an example of horizontal edge.

After quantization, the lowest frequency component is coded by predictive coding and Huffman coding. The decoding from the proposed model is realized by entropy decoder and additions of eight synthesis functions to the decoded lowest frequency component.

All coding results are compared with transform

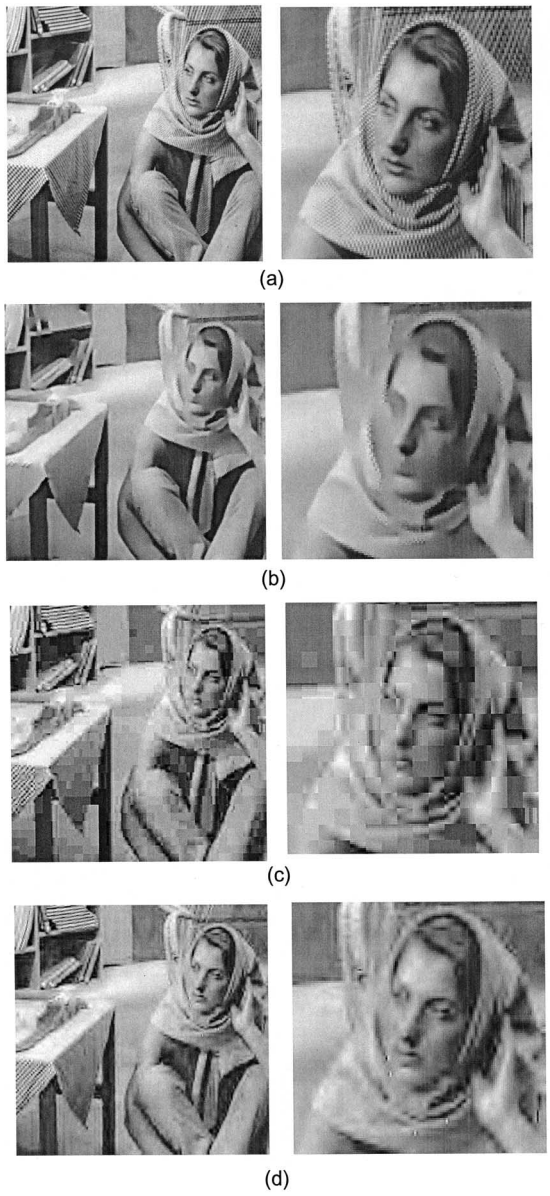


Fig. 9 (a) Original image, (b) proposed method, (c) JPEG and (d) EZW (Left: Entire decoded images. Right: Partial display of decoded images).

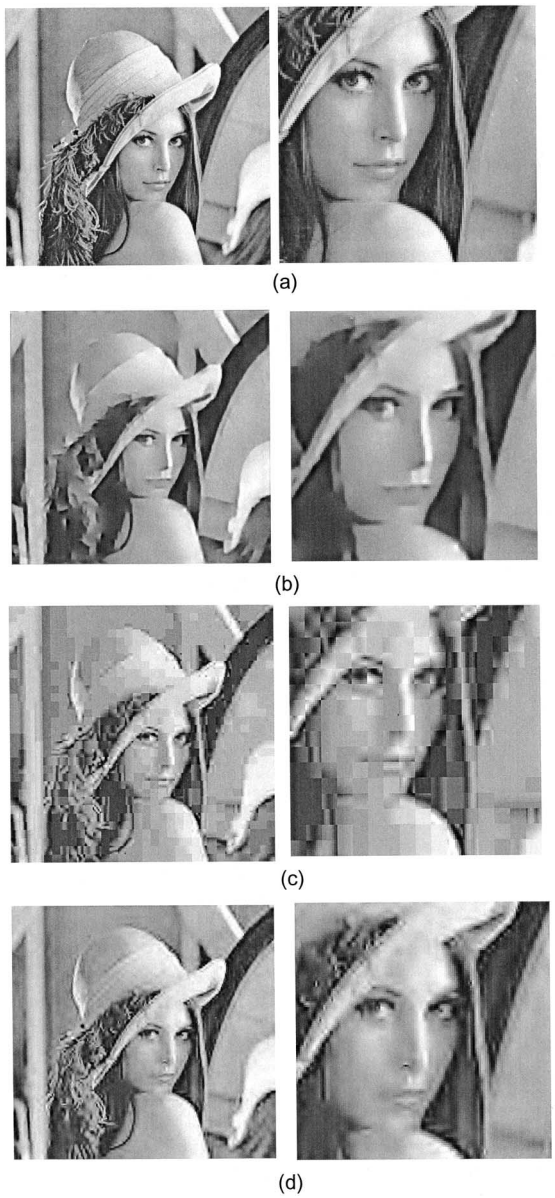


Fig. 10 (a) Original image, (b) proposed method, (c) JPEG and (d) EZW (Left: Entire decoded images. Right: Partial display of decoded images).

coding methods. Base-line JPEG and EZW [15] coders are employed to compare with the proposed method. JPEG and EZW are based on DCT and the biorthogonal wavelet transform, respectively. The wavelet filter-bank of EZW coder is selected as 9-tap QMF. Figures 9, 10 and 11 show the coding result of 256×256 images by the proposed model. The results of JPEG and EZW at the same bit-rates are also shown. The bit-rate and peak-to-peak SNR of the coded images are shown in Table 2, where SNR before coefficient quantization are also shown. The data amount for each components that comprise the lowest frequency component, code for the edge positions, the code-book of VQ and the vector index allocated to each edge position are given in Table 3. Almost half of the entire data is spent for the edge positions. The data amount for the edge geometry increases in proportion to the number of the edge positions. In Table 4, the number of the edge positions for each original image is shown. The average bit rate is about 2.5 bits to record the position of an edge pixel.

In the coding result of Fig. 9, fine textures appear over the entire original image. SNR before quantization is the lowest among three test images. The decoded image of the proposed method loses all of fine textures and details. SNR after coding is lower than the transform coding methods. However, the outlines of objects are still reserved well. The original image shown in Fig. 11 consists of planer regions. In cod-

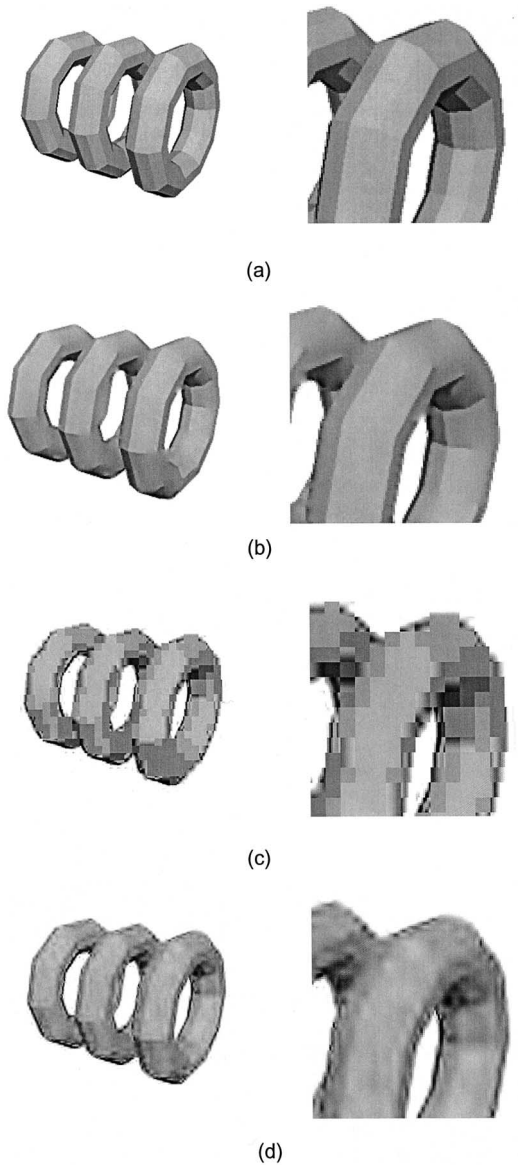


Fig. 11 (a) Original image, (b) proposed method, (c) JPEG and (d) EZW (Left: Entire decoded images. Right: Partial display of decoded images).

Table 2 Coding results for test images.

Image	Coder	Size in bytes	bits/pixel	PSNR in dB (before quantization)	PSNR in dB (after quantization)
Barbara	Proposed	2,556	0.312	25.97	25.10
	Baseline JPEG	2,377	0.290		25.19
	EZW	2,556	0.293		27.27
Lena	Proposed	2,087	0.255	26.60	25.70
	Baseline JPEG	2,172	0.265		24.95
	EZW	2,087	0.255		27.75
CG	Proposed	1,128	0.137	30.90	28.01
	Baseline JPEG	1,567	0.192		24.45
	EZW	1,128	0.137		27.58

Table 3 Data amount of coding images of each component.

Image	Lowest frequency component (bytes)	Edge positions (bytes)	Code-book for VQ (bytes)	Vector index allocated to each contour position (bytes)
Barbara	490	1,447	48	571
Lena	499	1,124	48	416
CG	302	595	48	183

Table 4 Edge geometry for each image.

Image	Number of edge positions	Number of initial positions for edge curves	Average bit rate for a edge pixel to record its position (Bits/Contour pixel)
Barbara	4,743	272	2.44
Lena	3,659	136	2.45
CG	1,883	54	2.57

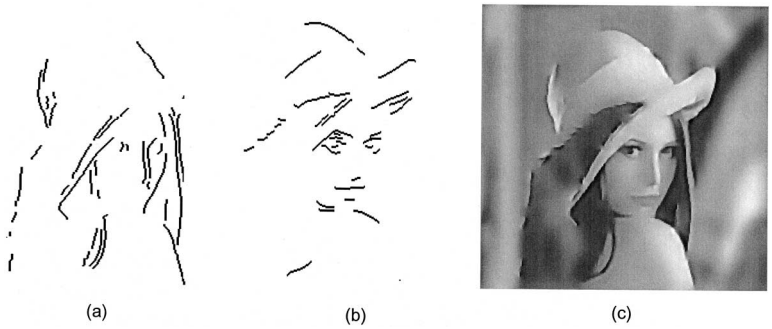


Fig. 12 (a), (b) Extracted edges from Figs. 10(a) and (c) the decoded image.

ing results by the transform coders, distortions appear around the position where the intensity changes. By contrast, these distortions are not found in the image by the proposed model. Since the zero-crossing points of synthesis functions only exist on edge positions, illegal oscillations are not created around edges. SNR is higher than other conventional transform methods for the image that does not include fine texture and small edges.

Finally, an example of object extraction in the code domain is presented. The edges that indicate a human face are extracted from the coded data rather than a decoded image. The extracted edges are shown in Figs. 12(a) and (b). The edge selection itself is performed manually. The decoded image is shown in Fig. 12(c). The data for synthesizing the image Fig. 12(c) amounts to 1,567 bytes. In the proposed coding method, the coded data correspond to the edge geometry of an original image. The image coding by the proposed model will be feasible for such a code domain image processing.

6. Conclusions

We have proposed a new image synthesis model based on the wavelet transform. An image is approximated by the synthesis functions that are translated to the edge positions. We have also proposed a synthesis function design by an iterative procedure using a pair of the course and inverse wavelet transforms. We apply the proposed model to the sketch-based image coding. By the proposed model, only additions of the synthesis functions are required for the image synthesis. In coding experiments, the image coding by using eight synthesis functions is demonstrated. Image edges are better preserved than transform coding methods. By this property, the proposed method can be apply to image coding for edge-based image processing such as image measurement, acquisition and many computer vision applications. For the derivation of the coefficient vector, we have imposed three conditions as convex sets and a convex function. Other conditions will be introduced for the specific application. The application specific conditions are future subject.

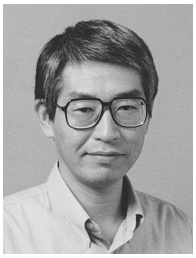
In image compression experiments, the compression results in our study are limited in extremely high compression ratios. Textures and fine structure of images are almost removed in the encoding process. By encoding the residual that represents removed components, a layered image coding will be implemented. If the synthesis functions are designed as the integer coefficient functions, the decoding process can be implemented by only integer additions and the coding error between the decoded image and the original image will be also an integer. By encoding this integer-valued error, the proposed coding method would be applied to lossless compression. The application of the proposed coding method is also future subject.

References

[1] S. Carlsson, "Sketch based coding of gray level images," Signal Processing, vol.15, pp.57–83, 1988.
[2] M. Kunt, M. Benard, and R. Leonardi, "Recent results in high-compression image coding," IEEE Trans. Circuits & Syst., vol.34, no.11, pp.1306–1336, Nov. 1987.
[3] X. Ran and N. Farvardin, "A perceptually motivated three-component image model—Part II: Applications to image compression," IEEE Trans. Image Processing, vol.4, no.4, pp.430–447, April 1995.
[4] Y. Itoh, "An Edge-oriented progressive image coding," IEEE Trans. Circuits & Syst. for Video Technol., vol.6, no.2, pp.135–142, April 1996.
[5] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," IEEE Trans. Pattern Anal. & Mach. Intell., vol.14, no.7, pp.710–732, July 1992.
[6] Z. Cvetković and M. Vetterli, "Discrete-time wavelet extrema representation design and consistent reconstruction," IEEE Trans. Signal Processing, vol.43, no.3, pp.681–693, March 1995.
[7] L.H. Croft and J.A. Robinson, "Subband image coding using watershed and watercourse lines of the wavelet transform," IEEE Trans. Image Processing, vol.3, no.6, pp.759–772, Nov. 1994
[8] M. Nakashizuka and H. Kikuchi, "Edge-based image synthesis model and its application to image coding," Proc. IEEE Int. Symp. on Circuits and Systems, vol.IV, pp.25–29, Orlando, May 1999.
[9] D. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," IEEE Trans. Med. Imaging, vol.1, pp.81–101, Oct. 1982.
[10] G.H. Golub and C.F.V. Loan, Matrix Computations, The Johns Hopkins University Press, 1996.
[11] D. Youla and V. Velasco, "Extensions of a result on the synthesis of signals in the presence of inconsistent constraints," IEEE Trans. Circuits & Syst., vol.33, no.4, pp.465–468, April 1986.
[12] F. Deutsch and I. Yamada, "Minimizing certain convex functions over the intersection of the fixed point sets of nonexpansive mappings," Numer. Funct. Anal. and Optim., vol.19, no.1 & 2, pp.33–56, 1998.
[13] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, 1992.
[14] A. Gersho and R.M. Gray, Vector quantization and signal compression, Kluwer Academic Publishers, 1992.
[15] J.M. Shapiro, "Embeded image coding using zerotrees of wavelet coefficients," IEEE Trans. Signal Processing, vol.41, no.12, pp.3445–3462, Dec. 1993.
[16] H. Freeman, "On the encoding of arbitrary geometric configurations," IRE Trans. Electron. Comp., vol.EC-10, pp.260–268, Oct. 1963.

Makoto Nakashizuka was born in Niigata, Japan, on January 31, 1966. He received B.E., M.E., and Ph.D. degrees from Niigata University, Niigata, Japan, in 1988, 1990, 1993. He was with Department of Information Engineering, Niigata University as a Research Associate from 1993 to 1997. He is now an Associate Professor at Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology. His research interests include time-scale/frequency analysis and its applications to signal, bio-medical, vision and image processing. Dr. Nakashizuka is a member of IEEE, Japan Society of Medical Electronics and Biological Engineering and Institute of Image Information and Television Engineers.

Hidetoshi Okazaki was born in Kouchi, Japan, on April 23, 1962. From 1983 to 1984, he worked at Olympus Optical Ltd. From 1985 to 1996, he worked at EDEC Ltd. He received B.E. and M.E. degrees from Tokyo University of Agriculture and Technology in 1997 and 1999, respectively. He is now Ph.D. student of Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology. His research interests include image processing and multiresolution analysis.



Hisakazu Kikuchi was born in Niigata, Japan, on March 16, 1952. He received B.E. and M.E. degrees from Niigata University, Niigata, in 1974 and 1976, respectively, and Dr.Eng. degree in electrical and electronic engineering from Tokyo institute of Technology, Tokyo, in 1988. From 1976 to 1979 he worked at Information Processing Systems Laboratories, Fujitsu Ltd. Since 1979 he has been with Niigata University, where he is Professor in electrical engineering. During a year of 1992 to 1993, he was a visiting scientist at University of California, Los Angeles sponsored by the Ministry of Education, Science and Culture. His research interests include digital signal processing, image processing, wavelets, and mobile communication. Dr. Kikuchi is a member of IEEE, Institute of Image Information and Television Engineers and Japan SIAM.