HOKKAIDO UNIVERSITY

Title	Adaptive missing texture reconstruction method based on kernel cross-modal factor analysis with a new evaluation criterion
Author(s)	Ogawa, Takahiro; Haseyama, Miki
Citation	Signal Processing, 103, 69-83 https://doi.org/10.1016/j.sigpro.2013.10.033
Issue Date	2014-10
Doc URL	http://hdl.handle.net/2115/57458
Type	article (author version)
File Information	Adaptive missing texture reconstruction method based on kernel cross-modal factor.pdf

Instructions for use

Adaptive Missing Texture Reconstruction Method Based on Kernel Cross-Modal Factor Analysis with a New Evaluation Criterion

Takahiro Ogawa[1], Miki Haseyama[2]

Graduate School of Information Science and Technology, Hokkaido University, JAPAN

Abstract

This paper presents an adaptive missing texture reconstruction method based on kernel cross-modal factor analysis (KCFA) with a new evaluation criterion. The proposed method estimates the latent relationship between two areas, which correspond to a missing area and its neighboring area, respectively, from known parts within the target image and realizes reconstruction of the missing textures. In order to obtain this relationship, KCFA is applied to each cluster containing similar known textures, and the optimal cluster is used for reconstructing each target missing area. Specifically, a new criterion obtained by monitoring errors caused in the latent space enables selection of the optimal cluster. Then each missing texture is adaptively estimated by the optimal cluster's latent relationship, which enables accurate reconstruction of similar textures. In our method, the above criterion is also used for estimating patch priority, which determines the reconstruction order of missing areas within the target image. Since patches, whose textures are ac-

---

[1]T. Ogawa is with Graduate School of Information Science and Technology, Hokkaido University, Sapporo, 060-0814 JAPAN, E-mail:ogawa@lmd.ist.hokudai.ac.jp

[2]M. Haseyama is with Graduate School of Information Science and Technology, Hokkaido University, Sapporo, 060-0814 JAPAN, E-mail:miki@ist.hokudai.ac.jp

curately modeled by our KCFA-based method, can be selected by using the new criterion, it becomes feasible to perform successful reconstruction of the missing areas. Experimental results show improvements of our KCFA-based reconstruction method over previously reported methods.

1. Introduction

Missing area reconstruction has been intensively studied in the field of image processing since it can afford a number of fundamental applications. Many methods that focus on the reconstruction of important visual features such as structures and textures within target images have been proposed. Most of the methods are broadly classified into two categories: missing structure reconstruction [1]–[9] and missing texture reconstruction [10]–[22]. In addition, there have been proposed reconstruction methods which adopt the combination use of the structure and texture reconstruction approaches [23]. The variational image inpainting methods which can successfully reconstruct structure components in images have been intensively studied in this research field. The variational image inpainting is performed based on the continuity of the geometrical structure of images. Most variational inpainting methods solve partial differential equations (PDEs). One of the pioneering work was proposed by Masnou et al. [1]. Furthermore, Bertalmio et al. proposed a representative image inpainting technique which is based on PDEs [2], and they have also realized several important achievements [3, 4]. In recent years, many improvement methods of the variational image inpainting have

2

been reported [7, 8, 9]. Although these variational image inpainting methods enable successful reconstruction of the structure components, images also include other different components, i.e., texture components, and alternative methods tend to output better results. The remainder of this paper focuses on the reconstruction of textures with discussion of its details.

It is well known that missing texture reconstruction is realized by several approaches such as exemplar-based methods [10]–[13] and multivariate analysis-based methods [14]–[22]. A method based on texture synthesis was first proposed by Efros et al. [10]. Based on their idea, an exemplar-based image inpainting method was proposed by Criminisi et al. [11, 12], and it became a representative method in this field. Recently, the exemplar-based approach has been improved by many researchers, and several state-of-the-art methods have also been proposed [13, 22]. Generally, the performance of exemplar-based methods tends to depend on the number of training examples. A sufficient number of training examples is necessary to accurately represent texture features within target images. Thus, if sufficient training examples cannot be provided, it becomes difficult to model the relationship between missing areas and other known areas in those methods.

Missing texture reconstruction methods using multivariate analysis have traditionally been proposed, and they are based on texture approximation using various methods such as principal component analysis (PCA), kernel PCA (KPCA) [14]–[17] and sparse representation. Several reconstruction methods based on sparse representation have recently been proposed [18]–[21], and a representative one was proposed by Mairal et al. [18]. Furthermore, sparse representation can be combined with the exemplar-based approach [22]. Since the conventional methods based on multivariate analysis represent textures by using their sub-

3

spaces, more successful approximation can be expected compared to exemplar-based methods. However, it should be noted that those methods generally assume that arbitrary local textures within the target image are similar to each other; that is, the target image contains almost one type of texture. Thus, if the target image consists of various textures, the missing textures should be adaptively reconstructed from only the same kinds of textures.

In this paper, we present a novel missing texture reconstruction method based on kernel cross-modal factor analysis (KCFA) [24, 25]. The main contributions of our method are threefold. First, the proposed method estimates the latent relationship between two areas, which respectively correspond to missing areas and their neighboring areas, from other known parts within the target image by using KCFA. Then this approach enables reconstruction of the missing areas based on the obtained KCFA-based relationship. Secondly, the proposed method performs clustering of known textures based on KCFA to realize reconstruction of missing areas by using optimal clusters. In this approach, a new criterion, which is obtained from errors caused in the KCFA-based latent space, is adopted to perform the clustering and select the optimal cluster for the target missing areas. Therefore, by monitoring this criterion, the reconstruction results by the optimal clusters can be adaptively obtained as the final results. Thirdly, the proposed method introduces a new priority estimation scheme based on the above criterion. The determination of the order for reconstructing missing areas, i.e., "priority estimation", is an important problem. By using the derived priority, missing areas, which can be successfully reconstructed by our KCFA-based method, are adaptively selected from the target image. Consequently, successful reconstruction of the missing textures can be expected by using the proposed method. It should

4

be noted that the proposed method shown in this paper is an improved version of [26]. The biggest difference between these two methods is the use of the clustering scheme. In the proposed method, we newly perform the clustering of known textures and the selection of the best matched cluster for the target missing areas, which are realized by monitoring the new evaluation criterion, in the KCFA-based reconstruction. Therefore, the proposed method is implemented in such a way that images including several kinds of textures can be reconstructed successfully.

This paper is organized as follows. First, in Section 2, we explain the concept of KCFA and its specific procedures. Next, a new missing texture reconstruction method based on KCFA is presented in Section 3. Section 4 shows experimental results in order to verify the performance of our method. Finally, concluding remarks are presented in Section 5.

2. Kernel Cross-Modal Factor Analysis

In this section, we present an overview of KCFA [25] as a preliminary. Suppose there is a pair of variables $\mathbf{x}_i \in \mathbf{R}^{n_x}$ and $\mathbf{y}_i \in \mathbf{R}^{n_y}$ ($i = 1, 2, \cdots, N$; N being the number of samples), KCFA tries to find two linear transformations that minimize the distance between two projections in the feature space.

First, $\mathbf{x}_i$ and $\mathbf{y}_i$ are respectively mapped into the feature space via nonlinear maps $\phi_\mathbf{x}$ and $\phi_\mathbf{y}$ [27] to obtain $\phi_\mathbf{x}(\mathbf{x}_i) \in \mathbf{R}^{\tilde{n}_x}$ and $\phi_\mathbf{y}(\mathbf{y}_i) \in \mathbf{R}^{\tilde{n}_y}$. Then given $\mathbf{\Xi}_\mathbf{x} = [\phi_\mathbf{x}(\mathbf{x}_1), \phi_\mathbf{x}(\mathbf{x}_2), \cdots, \phi_\mathbf{x}(\mathbf{x}_N)]$ and $\mathbf{\Xi}_\mathbf{y} = [\phi_\mathbf{y}(\mathbf{y}_1), \phi_\mathbf{y}(\mathbf{y}_2), \cdots, \phi_\mathbf{y}(\mathbf{y}_N)]$, KCFA estimates two orthonormal matrices $\hat{\mathbf{U}} \in \mathbf{R}^{\tilde{n}_x \times d}$ and $\hat{\mathbf{V}} \in \mathbf{R}^{\tilde{n}_y \times d}$ as follows:

$$\left\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\right\} = \min_{\mathbf{U}, \mathbf{V}} \left\|\mathbf{U}'\mathbf{\Xi}_\mathbf{x}\mathbf{H} - \mathbf{V}'\mathbf{\Xi}_\mathbf{y}\mathbf{H}\right\|_F^2, \tag{1}$$

where $\|\cdot\|_F$ represents the Frobenius norm. The matrix $\mathbf{H}$ is an $N \times N$ centering

matrix satisfying $\mathbf{H}' = \mathbf{H}$ and $\mathbf{H}^2 = \mathbf{H}$ and defined as follows:

$$\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{11}', \tag{2}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix, and $\mathbf{1} = [1, 1, \cdots, 1]'$ is an $N \times 1$ vector. In this paper, vector/matrix transpose is denoted by the superscript $'$.

Note that in the above equation,

$$
\begin{aligned}
\left\|\mathbf{U}'\mathbf{\Xi_x}\mathbf{H} - \mathbf{V}'\mathbf{\Xi_y}\mathbf{H}\right\|_F^2 &= \left\|\mathbf{H}\mathbf{\Xi_x}'\mathbf{U} - \mathbf{H}\mathbf{\Xi_y}'\mathbf{V}\right\|_F^2 \\
&= tr\left(\mathbf{H}\mathbf{\Xi_x}'\mathbf{\Xi_x}\mathbf{H}\right) + tr\left(\mathbf{H}\mathbf{\Xi_y}'\mathbf{\Xi_y}\mathbf{H}\right) \\
&\qquad -2tr\left(\mathbf{H}\mathbf{\Xi_x}'\mathbf{U}\mathbf{V}'\mathbf{\Xi_y}\mathbf{H}\right).
\end{aligned}
\tag{3}
$$

is satisfied, where $tr(\cdot)$ represents the trace of a matrix. It should be noted that $tr\left(\mathbf{H}\mathbf{\Xi_x}'\mathbf{\Xi_x}\mathbf{H}\right)$ and $tr\left(\mathbf{H}\mathbf{\Xi_y}'\mathbf{\Xi_y}\mathbf{H}\right)$ are constants, and Eq. (1) can be rewritten as follows:

$$\left\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\right\} = \max_{\mathbf{U},\mathbf{V}} tr\left(\mathbf{H}\mathbf{\Xi_x}'\mathbf{U}\mathbf{V}'\mathbf{\Xi_y}\mathbf{H}\right). \tag{4}$$

As shown in [24], the optimal matrices $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ can be obtained by performing the following singular value decomposition:

$$\mathbf{\Xi_x}\mathbf{H}\mathbf{H}\mathbf{\Xi_y}' = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{V}}', \tag{5}$$

where $\hat{\mathbf{\Lambda}}$ is an eigenvalue matrix. Unfortunately, since the columns of $\mathbf{\Xi_x}$ and $\mathbf{\Xi_y}$ are high dimensional or infinite dimensional, we cannot directly solve the above singular value decomposition. Therefore, by using "kernel trick" [25, 27], the calculation of the orthonormal matrices $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ becomes feasible. The details are shown below.

Given two Gram matrices

$$\mathbf{K_x} = \mathbf{\Xi_x}'\mathbf{\Xi_x}, \tag{6}$$

$$\mathbf{K_y} = \mathbf{\Xi_y}'\mathbf{\Xi_y}, \tag{7}$$

we focus on the following eigenvalue decomposition problem:

$$\left(\mathbf{HK_xHHK_yH}\right)\mathbf{E} = \mathbf{LE}, \tag{8}$$

where $\mathbf{E}$ and $\mathbf{L}$ are respectively eigenvector and eigenvalue matrices. From Eqs. (6) and (7),

$$\left(\mathbf{H\Xi_x}'\mathbf{\Xi_xHH\Xi_y}'\mathbf{\Xi_yH}\right)\mathbf{E} = \mathbf{LE}. \tag{9}$$

Furthermore, by multiplying both sides by $\mathbf{\Xi_yH}$,

$$\mathbf{\Xi_yH}\left(\mathbf{H\Xi_x}'\mathbf{\Xi_xHH\Xi_y}'\mathbf{\Xi_yH}\right)\mathbf{E} = \mathbf{\Xi_yHEL}. \tag{10}$$

Then

$$\left(\mathbf{\Xi_xHH\Xi_y}'\right)'\left(\mathbf{\Xi_xHH\Xi_y}'\right)\mathbf{\Xi_yHE} = \mathbf{\Xi_yHEL}. \tag{11}$$

The above equation means that the eigenvector matrix of $\left(\mathbf{\Xi_xHH\Xi_y}'\right)'\left(\mathbf{\Xi_xHH\Xi_y}'\right)$ corresponds to $\mathbf{\Xi_yHE}$ but is not the same since the norm of its columns is not one. Therefore, we use a diagonal matrix $\mathbf{D}$ satisfying

$$\begin{aligned}\left(\mathbf{\Xi_yHED}\right)'\left(\mathbf{\Xi_yHED}\right) &= \mathbf{D}\left(\mathbf{E}'\mathbf{HK_yHE}\right)\mathbf{D} \\ &= \mathbf{I}. \tag{12}\end{aligned}$$

Then the eigenvalue matrix of $\left(\mathbf{\Xi_xHH\Xi_y}'\right)'\left(\mathbf{\Xi_xHH\Xi_y}'\right)$ is denoted as $\mathbf{P}^2$. Thus, Eq. (11) can be rewritten as follows:

$$\left(\mathbf{\Xi_xHH\Xi_y}'\right)'\left(\mathbf{\Xi_xHH\Xi_y}'\right)\mathbf{\Xi_yHED} = \mathbf{\Xi_yHEDP}^2. \tag{13}$$

Note that from Eq. (5),

$$\left(\mathbf{\Xi_xHH\Xi_y}'\right)'\left(\mathbf{\Xi_xHH\Xi_y}'\right)\hat{\mathbf{V}} = \hat{\mathbf{V}}\hat{\mathbf{\Lambda}}^2 \tag{14}$$

7

is satisfied. Therefore, from Eqs. (13) and (14),

$$\hat{\mathbf{V}} = \boldsymbol{\Xi}_\mathbf{y}\mathbf{HED}, \tag{15}$$

$$\hat{\boldsymbol{\Lambda}} = \mathbf{P}. \tag{16}$$

Furthermore, from Eq. (5), (15) and (16),

$$\hat{\mathbf{U}} = \boldsymbol{\Xi}_\mathbf{x}\mathbf{HK}_\mathbf{y}\mathbf{HEDP}^{-1}. \tag{17}$$

In this way, KCFA can estimate the optimal orthonormal matrices $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ which optimize Eq. (1).

3. KCFA-Based Missing Texture Reconstruction

In this section, a KCFA-based missing texture reconstruction method is presented. In Fig. 1, an outline of the proposed method is shown. From a target image, we clip a patch f ($w \times h$ pixels) including missing areas and reconstruct its missing textures from textures in the other known areas. For the following explanations, two areas for which the intensities are known and unknown within the target patch f, are denoted as $\bar{\Omega}$ and Ω, respectively. Furthermore, we respectively denote vectors for which elements are intensities within $\bar{\Omega}$ and Ω as $\mathbf{x}(\in \mathbf{R}^{N^{\bar{\Omega}}})$ and $\mathbf{y}(\in \mathbf{R}^{N^{\Omega}})$, where $N^{\bar{\Omega}}$ and N^{Ω} are the numbers of pixels within $\bar{\Omega}$ and Ω, respectively. In the proposed method, we estimate the relationship between two areas corresponding to $\bar{\Omega}$ and Ω from known parts within the target image. This means that the proposed method calculates the orthonormal matrices as shown in the previous section in order to realize the reconstruction of the missing area Ω within the target patch f.

Note that in the target image, there are many known patches for which the textures are different from that of the target patch f. Such patches should not affect the reconstruction of the target patch f. Therefore, the proposed method applies KCFA to each cluster of patches containing the same kind of texture, and the optimal cluster is adaptively utilized for the reconstruction of the target patch f. In order to realize this scheme, clustering of the known patches within the target image must be performed before reconstruction of the target patch f. Thus, we firstly show the KCFA-based clustering algorithm of the known patches within the target image in 3.1. The adaptive reconstruction algorithm of the target patch f based on KCFA is shown in 3.2. Furthermore, the priority estimation algorithm, which is necessary to determine the order of reconstruction in the target image, is presented in 3.3.

3.1. Texture Clustering Algorithm

In this subsection, KCFA-based clustering of known patches within the target image as preprocessing for reconstruction of the target patch f is described. First, we clip known patches f_i ($i = 1, 2, \cdots, N$) whose size is $w \times h$ pixels from the target image, where N is the number of clipped known patches. Next, the vectors $\mathbf{x}_i (\in \mathbf{R}^{N^{\Omega}})$ and $\mathbf{y}_i (\in \mathbf{R}^{N^{\Omega}})$, which correspond to $\mathbf{x}$ and $\mathbf{y}$, respectively, are calculated for each known patch f_i. Furthermore, the proposed method maps $\mathbf{x}_i$ and $\mathbf{y}_i$ into the feature space via nonlinear maps $\phi_{\mathbf{x}}$ and $\phi_{\mathbf{y}}$ to obtain $\phi_{\mathbf{x}}(\mathbf{x}_i)$ and $\phi_{\mathbf{y}}(\mathbf{y}_i)$.

From $\phi_{\mathbf{x}}(\mathbf{x}_i)$ and $\phi_{\mathbf{y}}(\mathbf{y}_i)$ ($i = 1, 2, \cdots, N$), the proposed method performs their clustering that minimizes the following criterion:

$$J = \sum_{k=1}^{K} J^k, \tag{18}$$

$$J^k = \left\| \hat{\mathbf{U}}^{k'} \mathbf{\Xi}_{\mathbf{x}}^k \mathbf{H}^k - \hat{\mathbf{V}}^{k'} \mathbf{\Xi}_{\mathbf{y}}^k \mathbf{H}^k \right\|_F^2, \tag{19}$$

where K is the number of clusters. Furthermore,

$$\Xi_{\mathbf{x}}^k = [\phi_{\mathbf{x}}(\mathbf{x}_1^k), \phi_{\mathbf{x}}(\mathbf{x}_2^k), \cdots, \phi_{\mathbf{x}}(\mathbf{x}_{N^k}^k)], \tag{20}$$

$$\Xi_{\mathbf{y}}^k = [\phi_{\mathbf{y}}(\mathbf{y}_1^k), \phi_{\mathbf{y}}(\mathbf{y}_2^k), \cdots, \phi_{\mathbf{y}}(\mathbf{y}_{N^k}^k)], \tag{21}$$

and the vectors $\mathbf{x}_j^k$ and $\mathbf{y}_j^k$ ($j = 1, 2, \cdots, N^k$) represent $\mathbf{x}_i$ and $\mathbf{y}_i$ ($i = 1, 2, \cdots, N$) assigned to cluster k. The value N^k is the number of elements belonging to cluster k. The matrix $\mathbf{H}^k$ is the $N^k \times N^k$ centering matrix defined in the same way as Eq. (2). Then $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ are orthonormal matrices obtained by applying KCFA to $\phi_{\mathbf{x}}(\mathbf{x}_j^k)$ and $\phi_{\mathbf{y}}(\mathbf{y}_j^k)$ ($j = 1, 2, \cdots, N^k$). The details of their calculation are shown below.

By using KCFA [25], the proposed method calculates the two orthonormal matrices $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ minimizing the following criterion:

$$\left\{\hat{\mathbf{U}}^k, \hat{\mathbf{V}}^k\right\} = \min_{\mathbf{U}^k, \mathbf{V}^k} \left\|\mathbf{U}^{k'}\Xi_{\mathbf{x}}^k\mathbf{H}^k - \mathbf{V}^{k'}\Xi_{\mathbf{y}}^k\mathbf{H}^k\right\|_F^2, \tag{22}$$

where its overview is shown in Fig. 2. Furthermore, as shown in the previous section, the above problem can be rewritten as

$$\left\{\hat{\mathbf{U}}^k, \hat{\mathbf{V}}^k\right\} = \max_{\mathbf{U}^k, \mathbf{V}^k} tr\left(\mathbf{H}^k\Xi_{\mathbf{x}}^{k'}\mathbf{U}^k\mathbf{V}^{k'}\Xi_{\mathbf{y}}^k\mathbf{H}^k\right). \tag{23}$$

Then $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ satisfy the following singular value decomposition:

$$\Xi_{\mathbf{x}}^k\mathbf{H}^k\mathbf{H}^k\Xi_{\mathbf{y}}^{k'} = \hat{\mathbf{U}}^k\hat{\mathbf{\Lambda}}^k\hat{\mathbf{V}}^{k'}. \tag{24}$$

The optimal solution of the above equation can be obtained as shown in the previous section. Specifically,

$$\hat{\mathbf{V}}^k = \Xi_{\mathbf{y}}^k\mathbf{T}_{\mathbf{V}}^k, \tag{25}$$

where

$$\mathbf{T}_\mathbf{V}^k = \mathbf{H}^k\mathbf{E}^k\mathbf{D}^k. \tag{26}$$

Note that $\mathbf{E}^k$ is the eigenvector matrix of the following eigenvalue problem:

$$\left(\mathbf{H}^k\mathbf{K}_\mathbf{x}^k\mathbf{H}^k\mathbf{H}^k\mathbf{K}_\mathbf{y}^k\mathbf{H}^k\right)\mathbf{E}^k = \mathbf{L}^k\mathbf{E}^k, \tag{27}$$

where $\mathbf{L}^k$ is the eigenvalue matrix, and $\mathbf{K}_\mathbf{x}^k$ and $\mathbf{K}_\mathbf{y}^k$ are the Gram matrices respectively defined as

$$\mathbf{K}_\mathbf{x}^k = \mathbf{\Xi}_\mathbf{x}^{k\prime}\mathbf{\Xi}_\mathbf{x}^k, \tag{28}$$

$$\mathbf{K}_\mathbf{y}^k = \mathbf{\Xi}_\mathbf{y}^{k\prime}\mathbf{\Xi}_\mathbf{y}^k. \tag{29}$$

In addition, $\mathbf{D}^k$ in Eq. (26) is a diagonal matrix satisfying

$$\begin{aligned}
\left(\mathbf{\Xi}_\mathbf{y}^k\mathbf{H}^k\mathbf{E}^k\mathbf{D}^k\right)^\prime\left(\mathbf{\Xi}_\mathbf{y}^k\mathbf{H}^k\mathbf{E}^k\mathbf{D}^k\right) &= \mathbf{D}^k\left(\mathbf{E}^{k\prime}\mathbf{H}^k\mathbf{K}_\mathbf{y}^k\mathbf{H}^k\mathbf{E}^k\right)\mathbf{D}^k \\
&= \mathbf{I}^k,
\end{aligned} \tag{30}$$

where $\mathbf{I}^k$ is the $N^k \times N^k$ identity matrix. Then, from Eqs. (24) and (25),

$$\hat{\mathbf{U}}^k = \mathbf{\Xi}_\mathbf{x}^k\mathbf{T}_\mathbf{U}^k, \tag{31}$$

where

$$\mathbf{T}_\mathbf{U}^k = \mathbf{H}^k\mathbf{K}_\mathbf{y}^k\mathbf{T}_\mathbf{V}^k\hat{\mathbf{\Lambda}}^{k^{-1}}. \tag{32}$$

In the above equations, the eigenvalue matrix $\hat{\mathbf{\Lambda}}^k$ satisfies

$$\hat{\mathbf{\Lambda}}^k = \mathbf{P}^k, \tag{33}$$

11

and the eigenvalue matrix $\mathbf{P}^k$ can be obtained by the following eigenvalue problem:

$$\left(\boldsymbol{\Xi}_{\mathbf{x}}^k \mathbf{H}^k \mathbf{H}^k \boldsymbol{\Xi}_{\mathbf{y}}^{k\prime}\right)' \left(\boldsymbol{\Xi}_{\mathbf{x}}^k \mathbf{H}^k \mathbf{H}^k \boldsymbol{\Xi}_{\mathbf{y}}^{k\prime}\right) \boldsymbol{\Xi}_{\mathbf{y}}^k \mathbf{H}^k \mathbf{E}^k \mathbf{D}^k = \boldsymbol{\Xi}_{\mathbf{y}}^k \mathbf{H}^k \mathbf{E}^k \mathbf{D}^k \mathbf{P}^{k2}, \tag{34}$$

where $\boldsymbol{\Xi}_{\mathbf{y}}^k \mathbf{H}^k \mathbf{E}^k \mathbf{D}^k$ becomes the eigenvector matrix.

By using Eqs. (25) and (31), Eq. (19) is rewritten as follows:

$$
\begin{aligned}
J^k &= \left\| \mathbf{T}_{\mathbf{U}}^{k\prime} \boldsymbol{\Xi}_{\mathbf{x}}^{k\prime} \boldsymbol{\Xi}_{\mathbf{x}}^k \mathbf{H}^k - \mathbf{T}_{\mathbf{V}}^{k\prime} \boldsymbol{\Xi}_{\mathbf{y}}^{k\prime} \boldsymbol{\Xi}_{\mathbf{y}}^k \mathbf{H}^k \right\|_F^2 \\
&= \left\| \mathbf{T}_{\mathbf{U}}^{k\prime} \mathbf{K}_{\mathbf{x}}^k \mathbf{H}^k - \mathbf{T}_{\mathbf{V}}^{k\prime} \mathbf{K}_{\mathbf{y}}^k \mathbf{H}^k \right\|_F^2.
\end{aligned}
\tag{35}
$$

Then, by minimizing Eq. (18) which is the sum of the above equation for all clusters, clustering of the known patches f_i ($i = 1, 2, \cdots, N$) becomes feasible. Specifically, for each patch, we assign it to the cluster minimizing the errors in the latent space and recalculate $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ from the elements belonging to cluster k ($k = 1, 2, \cdots, K$). By iterating these procedures, the final results of clustering and $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ can be obtained.

The KCFA can provide the latent relationship that minimizes the errors between the pairs of multi-variates. As shown in the above explanation, the proposed method calculates the orthonormal projection matrices $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ for each cluster k and performs the assignment of known patches f_i ($i = 1, 2, \cdots, N$). Note that from Eq. (22), the differences between the projection results of $\phi_{\mathbf{x}}(\mathbf{x}_j^k)$ and $\phi_{\mathbf{y}}(\mathbf{y}_j^k)$ ($j = 1, 2, \cdots, N^k$) become minimum in the same cluster. Therefore, clustering that can provide the optimal relationship for each cluster becomes feasible in the proposed KCFA-based method. In [26], the clustering of known textures is not adopted, and only one pair of $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ is derived. On the other hand, the proposed method newly adopts the clustering for obtaining $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ for each cluster k, i.e., each kind of texture. Therefore, if the optimal pair of $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$

can be used for the reconstruction of the target patch f, i.e., if the selection of the optimal cluster becomes feasible, adaptive reconstruction of each texture can be expected.

3.2. Texture Reconstruction Algorithm

In this section, we present an algorithm for reconstructing the missing area Ω within the target patch f from the clustering results obtained in the previous subsection. For each cluster k, we can obtain the orthonormal matrices $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$. Based on the obtained matrices, the proposed method performs reconstruction of the missing area Ω within the target patch f. Specifically, as shown in the lower part of Fig. 2, the estimation result $\phi_{\mathbf{y}}(\hat{\mathbf{y}}^k)$ of the unknown vector $\phi_{\mathbf{y}}(\mathbf{y})$ by the kth cluster is obtained as follows:

$$\phi_{\mathbf{y}}(\hat{\mathbf{y}}^k) = \hat{\mathbf{V}}^k \hat{\mathbf{U}}^{k'} \left(\phi_{\mathbf{x}}(\mathbf{x}) - \bar{\phi}_{\mathbf{x}}^k \right) + \bar{\phi}_{\mathbf{y}}^k, \tag{36}$$

where

$$\bar{\phi}_{\mathbf{x}}^k = \frac{1}{N^k} \mathbf{\Xi}_{\mathbf{x}}^k \mathbf{1}^k, \tag{37}$$

$$\bar{\phi}_{\mathbf{y}}^k = \frac{1}{N^k} \mathbf{\Xi}_{\mathbf{y}}^k \mathbf{1}^k, \tag{38}$$

and $\mathbf{1}^k = [1, 1, \cdots, 1]'$ is an $N^k \times 1$ vector. Furthermore, from Eqs. (25) and (31), Eq. (36) can be rewritten as follows:

$$\begin{aligned}
\phi_{\mathbf{y}}(\hat{\mathbf{y}}^k) &= \mathbf{\Xi}_{\mathbf{y}}^k \mathbf{T}_{\mathbf{V}}^k \mathbf{T}_{\mathbf{U}}^{k'} \mathbf{\Xi}_{\mathbf{x}}^{k'} \left(\phi_{\mathbf{x}}(\mathbf{x}) - \frac{1}{N^k} \mathbf{\Xi}_{\mathbf{x}}^k \mathbf{1}^k \right) + \frac{1}{N^k} \mathbf{\Xi}_{\mathbf{y}}^k \mathbf{1}^k \\
&= \mathbf{\Xi}_{\mathbf{y}}^k \left\{ \mathbf{T}^k \mathbf{\Xi}_{\mathbf{x}}^{k'} \left(\phi_{\mathbf{x}}(\mathbf{x}) - \frac{1}{N^k} \mathbf{\Xi}_{\mathbf{x}}^k \mathbf{1}^k \right) + \frac{1}{N^k} \mathbf{1}^k \right\} \\
&= \mathbf{\Xi}_{\mathbf{y}}^k \left\{ \mathbf{T}^k \mathbf{\Xi}_{\mathbf{x}}^{k'} \phi_{\mathbf{x}}(\mathbf{x}) - \frac{1}{N^k} \left(\mathbf{T}^k \mathbf{K}_{\mathbf{x}}^k - \mathbf{I}^k \right) \mathbf{1}^k \right\}, \tag{39}
\end{aligned}$$

13

where

$$\mathbf{T}^k = \mathbf{T}_{\mathbf{V}}^k \mathbf{T}_{\mathbf{U}}^{k\,\prime}. \tag{40}$$

Then, from Eq. (39), we can obtain the estimation result $\hat{\mathbf{y}}^k$ of the intensities within the missing area Ω. Note that the calculation of $\hat{\mathbf{y}}^k$ in the input space from $\phi_{\mathbf{y}}(\hat{\mathbf{y}}^k)$ in the feature space is possible when $\phi_{\mathbf{y}}$ is invertible. However, for many cases, e.g. the Gaussian kernel, the feature space obtained by the nonlinear map becomes high dimensional or infinite dimensional. In such cases, we have to solve the pre-image problem for estimating the corresponding vector in the input space by using some previously reported methods [28].

By calculating Eq. (39), the missing intensities in Ω can be estimated from cluster k. In the proposed method, the orthonormal matrices $\hat{\mathbf{U}}^k$ and $\hat{\mathbf{V}}^k$ minimize the errors in the latent space of cluster k. Therefore, if we can assign the target patch f containing the missing area Ω to the optimal cluster k^{opt}, the proposed method accurately estimates the unknown vector $\mathbf{y}$ from the known vector $\phi_{\mathbf{x}}(\mathbf{x})$ in Eq. (39). In order to achieve this assignment, the proposed method focuses on the criterion J^k in Eq. (35). This criterion represents how easily we can approximate $\phi_{\mathbf{y}}(\mathbf{y}_j^k)$ by $\phi_{\mathbf{x}}(\mathbf{x}_j^k)$ in the latent space. Therefore, similar to this criterion, we use the following criterion:

$$
\begin{aligned}
J_f^k &= \left\| \hat{\mathbf{U}}^{k\prime} \left(\phi_{\mathbf{x}}(\mathbf{x}) - \bar{\boldsymbol{\phi}}_{\mathbf{x}}^k \right) - \hat{\mathbf{V}}^{k\prime} \left(\phi_{\mathbf{y}}(\hat{\mathbf{y}}^k) - \bar{\boldsymbol{\phi}}_{\mathbf{y}}^k \right) \right\|^2 \\
&= \left\| \mathbf{T}_{\mathbf{U}}^{k\,\prime} \boldsymbol{\Xi}_{\mathbf{x}}^{k\prime} \left(\phi_{\mathbf{x}}(\mathbf{x}) - \frac{1}{N^k} \boldsymbol{\Xi}_{\mathbf{x}}^k \mathbf{1}^k \right) - \mathbf{T}_{\mathbf{V}}^{k\,\prime} \boldsymbol{\Xi}_{\mathbf{y}}^{k\prime} \left(\phi_{\mathbf{y}}(\hat{\mathbf{y}}^k) - \frac{1}{N^k} \boldsymbol{\Xi}_{\mathbf{y}}^k \mathbf{1}^k \right) \right\|^2 \\
&= \left\| \mathbf{T}_{\mathbf{U}}^{k\,\prime} \left(\boldsymbol{\Xi}_{\mathbf{x}}^{k\prime} \phi_{\mathbf{x}}(\mathbf{x}) - \frac{1}{N^k} \mathbf{K}_{\mathbf{x}}^k \mathbf{1}^k \right) - \mathbf{T}_{\mathbf{V}}^{k\,\prime} \left(\boldsymbol{\Xi}_{\mathbf{y}}^{k\prime} \phi_{\mathbf{y}}(\hat{\mathbf{y}}^k) - \frac{1}{N^k} \mathbf{K}_{\mathbf{y}}^k \mathbf{1}^k \right) \right\|^2
\end{aligned}
\tag{41}
$$

for assignment of the target patch f to the optimal cluster k^{opt}. Selection of the optimal cluster k^{opt} minimizing Eq. (41) for the target patch f then becomes feasible.

14

Furthermore, the proposed method regards the result $\hat{\mathbf{y}}^{k^{\mathrm{opt}}}$ obtained by the optimal cluster k^{opt} as the final output. Consequently, by performing the non-conventional approach, which adaptively selects the optimal cluster for the target missing area, we can reconstruct the missing textures in the target patch f accurately. As shown in the previous subsection, the proposed method newly introduces the clustering scheme into our previous work [26] for adaptively reconstructing images including several kinds of textures. Then, for the reconstruction of the target patch f, the selection of the optimal cluster k^{opt} becomes necessary. Therefore, in the proposed method, we use the criterion J_f^k in Eq. (41), which is derived on the basis of Eq. (35), for realizing this selection.

3.3. Priority Estimation Algorithm

As shown in the previous subsection, we can reconstruct the missing area Ω within the target patch f. Therefore, the proposed method clips patches including missing areas and performs their reconstruction to estimate all missing intensities. It should be noted that in order to realize this scheme, we have to determine the order in which patches along the fill-front of missing areas are filled. We call this order "patch priority". In the proposed method, patch priorities are determined on the basis of an improved version of the method proposed by Criminisi et al. [12]. Specifically, given a patch $f_{\mathbf{p}}$ centered at pixel $\mathbf{p}$ that is in the fill-front of the missing areas within the target image, its priority $P(\mathbf{p})$ is defined as follows:

$$P(\mathbf{p}) = C(\mathbf{p}) \cdot D(\mathbf{p}), \tag{42}$$

where $C(\mathbf{p})$ and $D(\mathbf{p})$ are called confidence term and data term, respectively, and they are defined as follows:

$$C(\mathbf{p}) \quad = \quad \frac{\sum_{\mathbf{q} \in f_{\mathbf{p}} \cap (I - \Theta)} C(\mathbf{q})}{|f_{\mathbf{p}}|}, \tag{43}$$

15

$$D(\mathbf{p}) \quad = \quad \frac{|\nabla I_{\mathbf{p}}^{\perp} \cdot \mathbf{n_p}|}{\alpha}. \tag{44}$$

In the above equations, I and Θ are the whole areas of the target image and whole missing areas, respectively. Furthermore, $|f_{\mathbf{p}}|$ $(= w \times h)$ represents the number of pixels included within the target patch $f_{\mathbf{p}}$. Then α is a normalization factor (e.g., $\alpha = 255$ for a typical grayscale image), $\nabla I_{\mathbf{p}}^{\perp}$ is an isophote at pixel $\mathbf{p}$, and $\mathbf{n_p}$ is a unit vector orthogonal to the fill-front at pixel $\mathbf{p}$. Note that $C(\mathbf{p})$ is initially set as $C(\mathbf{p}) = 0$ $\forall \mathbf{p} \in \Theta$ and $C(\mathbf{p}) = 1$ $\forall \mathbf{p} \in (I - \Theta)$.

In the proposed method, we assign new values to the confidence term for the reconstructed areas, i.e., we perform the renewal of the confidence term within the reconstructed area of the target patch after its reconstruction procedure. Specifically, we focus on the criterion $J_f^{k^{\text{opt}}}$ derived in Eq. (41) for the optimal cluster k^{opt} and denote it as $\xi(\mathbf{p})$. The confidence term of pixel $\mathbf{q}$ in the reconstructed area $\Omega_{\mathbf{p}}$ of the target patch $f_{\mathbf{p}}$ centered at $\mathbf{p}$ is calculated by $C(\mathbf{q}) = \exp\left(-\frac{\xi(\mathbf{p})}{\zeta}\right)$, where ζ is the average of the values obtained from known patches in the same way as Eq. (41), i.e., $\zeta = \frac{J}{N}$ from Eqs. (18) and (19). Note that we denote the reconstructed area and the known area within $f_{\mathbf{p}}$ as $\Omega_{\mathbf{p}}$ and $\bar{\Omega}_{\mathbf{p}}$, respectively. Therefore, the renewal of $C(\mathbf{q})$ $(\mathbf{q} \in f_{\mathbf{p}})$ is specifically performed as follows:

$$C(\mathbf{q}) \leftarrow \begin{cases} C(\mathbf{q}) & \text{if } \mathbf{q} \in \bar{\Omega}_{\mathbf{p}} \\ \exp\left(-\frac{\xi(\mathbf{p})}{\zeta}\right) & \text{if } \mathbf{q} \in \Omega_{\mathbf{p}} \end{cases}. \tag{45}$$

This means if the target pixel $\mathbf{q}$ is the pixel whose intensity is known, i.e., $\mathbf{q} \in \bar{\Omega}_{\mathbf{p}}$, the confidence terms do not change. Since the intensities within $\bar{\Omega}_{\mathbf{p}}$ do not change, the values of $C(\mathbf{q})$ do not also change. Thus, if $\mathbf{q}$ is the original pixel whose intensity is original, $C(\mathbf{q})$ is one. Furthermore, if $\mathbf{q}$ is the previously reconstructed pixel whose intensity was estimated by other patches, $C(\mathbf{q})$ does not change from

16

the value provided in the previous reconstruction process. On the other hand, if the target pixel $\mathbf{q}$ is the pixel whose intensity is estimated in this reconstruction process, i.e., $\mathbf{q} \in \Omega_{\mathbf{p}}$, the confidence term is calculated based on the criterion in Eq. (41), which corresponds to the error caused in the latent space. Therefore, we can regard $C(\mathbf{q})$ derived from this criterion as the reconstruction performance, i.e., the confidence, of the target area $\Omega_{\mathbf{p}}$. Therefore, it is reasonable to introduce $C(\mathbf{q})$ derived from the criterion in Eq. (41) into the KCFA-based reconstruction method. In this way, we can reconstruct all of the missing areas within the target image according to the patch priorities in Eq. (42).

4. Experimental Results

In this section, the performance of the proposed method is verified from results of experiments in order to confirm its effectiveness. We prepared three test images including missing areas and performed their reconstruction by using the proposed method and several conventional methods. The reconstruction performance of our method is verified from the obtained results, and the effectiveness of the proposed KCFA-based method is also discussed.

In the experiment, a corrupted image that contains text regions as missing areas was generated for each test image (Figs. 3(a), 5(a) and 7 (a)) as shown in Figs. 3(c), 5(c) and 7 (c). Note that these images were used in the previously reported papers [17, 21], and we used the same test images in this paper. Then we applied the proposed method and the conventional methods to these corrupted images and obtained their reconstruction results. For comparison, we used the methods in [14], [17], [12], [13] and [22] as conventional methods. Note that the methods in [14] and [17] generate subspaces of patches for reconstructing

17

missing areas based on PCA and KPCA, respectively. In the method in [17], a clustering scheme is also introduced into the reconstruction, and we therefore used this method as a recent state-of-the-art method. The conventional method in [12] is a representative method in the field of exemplar-based image inpainting. The methods in [13] and [22] can be regarded as its improved versions, and we thus regard them as state-of-the-art approaches.

In the experiment, the patch size was fixed to 17 (about two-times larger than the patch size usually used in exemplar-based methods), and the number of training examples thus became smaller. Note that since the methods in [14] and [17] do not adopt patch priority estimation, the size was set to 35. Since this comparison scheme was adopted in several papers, we also used such difficult conditions in order to make the difference in the performance of the proposed method and the conventional methods clearer. Furthermore, the number of clusters in our method was simply determined as $K = 4$. The kernel function of $\phi_{\mathbf{x}}(\cdot)$ was the Gaussian kernel whose parameter was twice of the variance of $\left\| \mathbf{x}_i - \mathbf{x}_j \right\|$ $(i, j = 1, 2, \cdots, N)$. Note that if we use the nonlinear map to the high-dimensional feature space for $\phi_{\mathbf{y}}(\cdot)$, the pre-image estimation is necessary to obtain $\hat{\mathbf{y}}^k$ in 3.2. Thus, in order to avoid errors of this estimation problem, we simply used the linear kernel.

From the obtained results shown in Figs. 3, 5 and 7, we can see that our method can achieve reconstruction of missing areas successfully. For better subjective evaluation, we show their zoomed portions in Figs. 4, 6 and 8. As shown in the previous section, the proposed method can estimate the latent relationship between unknown areas and their neighboring areas based on KCFA. Therefore, optimal reconstruction derived from the other known areas within the target image can be performed. Next, by introducing the new criterion into the clustering-based

reconstruction algorithm, the proposed method can output the results obtained by the optimal clusters as the final results. Furthermore, in the proposed method, the above criterion is also used in the determination of patch priority, and this also improves the missing area reconstruction.

Next, we show results of quantitative evaluation for the proposed method and the conventional methods. Results of SSIM index [29] calculated from the reconstruction results in Figs. 3, 5 and 7 are shown in Table 1. It is well known that the MSE (PSNR) and its variants cannot successfully reflect visual image quality [30]–[32]. Thus, since the SSIM index is one of the representative criteria measuring visual image quality, we adopted the SSIM index as the criterion for quantitative evaluation in this experiment. The definition of the SSIM index is shown below. The SSIM index represents the similarity between two signal vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ ($\in \mathbf{R}^n$), and its specific definition is as follows:

$$\mathrm{SSIM}(\mathbf{y}_1, \mathbf{y}_2) = [l(\mathbf{y}_1, \mathbf{y}_2)]^{\alpha} \cdot [c(\mathbf{y}_1, \mathbf{y}_2)]^{\beta} \cdot [s(\mathbf{y}_1, \mathbf{y}_2)]^{\gamma}, \tag{46}$$

where the terms $l(\mathbf{y}_1, \mathbf{y}_2)$ and $c(\mathbf{y}_1, \mathbf{y}_2)$ respectively compare the mean and variance of the two signal vectors. Furthermore, $s(\mathbf{y}_1, \mathbf{y}_2)$ measures their structural correlation. These three terms, $l(\mathbf{y}_1, \mathbf{y}_2)$, $c(\mathbf{y}_1, \mathbf{y}_2)$ and $s(\mathbf{y}_1, \mathbf{y}_2)$, are obtained as

$$l(\mathbf{y}_1, \mathbf{y}_2) = \frac{2\mu_{\mathbf{y}_1}\mu_{\mathbf{y}_2} + C_1}{\mu_{\mathbf{y}_1}^2 + \mu_{\mathbf{y}_2}^2 + C_1}, \tag{47}$$

$$c(\mathbf{y}_1, \mathbf{y}_2) = \frac{2\sigma_{\mathbf{y}_1}\sigma_{\mathbf{y}_2} + C_2}{\sigma_{\mathbf{y}_1}^2 + \sigma_{\mathbf{y}_2}^2 + C_2}, \tag{48}$$

$$s(\mathbf{y}_1, \mathbf{y}_2) = \frac{\sigma_{\mathbf{y}_1,\mathbf{y}_2} + C_3}{\sigma_{\mathbf{y}_1}\sigma_{\mathbf{y}_2} + C_3}. \tag{49}$$

In the above equations, $\mu_{\mathbf{y}_1}$ and $\mu_{\mathbf{y}_2}$ are the means of $\mathbf{y}_1$ and $\mathbf{y}_2$, $\sigma_{\mathbf{y}_1}^2$ and $\sigma_{\mathbf{y}_2}^2$ are the variances of $\mathbf{y}_1$ and $\mathbf{y}_2$, and $\sigma_{\mathbf{y}_1,\mathbf{y}_2}$ is the cross covariance between $\mathbf{y}_1$ and $\mathbf{y}_2$. The

19

constants C_1, C_2 and C_3 are necessary for avoiding instability when the denominators are very close to zero. The parameters $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ determine the relative importance of the three components in Eq. (46). As shown in [29], those parameters are set as $\alpha = \beta = \gamma = 1$ and $C_3 = \frac{C_2}{2}$, and formulation of the SSIM index is simplified by

$$\text{SSIM}(\mathbf{y}_1, \mathbf{y}_2) = \frac{\left(2\mu_{\mathbf{y}_1}\mu_{\mathbf{y}_2} + C_1\right)\left(2\sigma_{\mathbf{y}_1,\mathbf{y}_2} + C_2\right)}{\left(\mu_{\mathbf{y}_1}^2 + \mu_{\mathbf{y}_2}^2 + C_1\right)\left(\sigma_{\mathbf{y}_1}^2 + \sigma_{\mathbf{y}_2}^2 + C_2\right)}. \tag{50}$$

As shown in Eqs. (46), the SSIM index is based on human visual system (HVS). Specifically, the first term defined in Eq. (47) is consistent with Weber's law, which states that the HVS is sensitive to the relative luminance change, and not to the absolute luminance change. The second term defined in Eq. (48) is derived based on the contrast masking characteristic that the contrast change is less sensitive when there is a high base contrast than there is a low base contrast. Then, in the third term defined in Eq. (49), the structure comparison is conducted after luminance subtraction and contrast normalization. If we ignore C_3, it is equivalent to calculating the correlation coefficient. In this way, the SSIM index can measure the similarity between two signal vectors according to the HVS. This means the SSIM index can provide the similarity between the original image and the reconstructed image with considering the sensitivity of their difference to the HVS. The details of mathematical properties of the SSIM index are also shown in [33].

In addition to the three test images shown in Figs. 3, 5 and 7, we added six test images shown in Figs. 9–14 to the results of the quantitative evaluation. In these figures, we only show results of the proposed method due to the limitation of pages. Note that the SSIM values are calculated from only the reconstructed areas. As shown in Table 1, we can confirm that the proposed method mostly achieves the

20

improvement. Even though the use of the SSIM index is effective, it is difficult to perfectly determine the order of reconstruction performance that is the same as the subjective evaluation. This means that ranking of the reconstruction performance that perfectly reflects subjective evaluation is difficult, and further improvement is necessary in future work.

In the above experiment, we used the difficult condition, i.e., larger size patches, in order to make the difference of the reconstruction performance between our method and the conventional methods clearer. Next, we show other different experimental results obtained by using conditions which were adopted in each paper. This means the conditions of the conventional methods were determined according to their papers. In the new experiments, we prepared six test images shown in Fig. 15 and randomly added missing blocks of size 8×8 pixels with changing the ratio of the missing pixels. Figure 16 shows the relationship between the ratio of the missing pixels and the SSIM index calculated from the reconstructed image. In the proposed method, the patch size was simply set to 15×15 pixels. From these results, we can see the proposed method tends to output better results than those of the conventional methods.

Finally, we discuss the limitations of the proposed method. In this paper, we have focused on the reconstruction of missing textures, and the proposed method is optimized for realizing the accurate texture reconstruction. Therefore, when applying the proposed method to an image including strong structure components, undesired degradations are observed. We show an example obtained by reconstructing an image including structure components based on our method in Fig. 17. From the obtained results, we can see successful reconstruction of some structure components , i.e., missing edges is difficult in the proposed method. In this

figure, we also show an example reconstructed by our previously reported method which focuses on accurate reconstruction of missing structure components [6]. This method adopts a probabilistic model, GMRF model, but the main concept is quite similar to the variational image inpainting. Compared with the result obtained by the proposed method, the missing edges can be accurately recovered by this method. Therefore, it is necessary to adopt variational image inpainting methods [1]–[9] for assisting accurate reconstruction of images including structure components.

We also discuss the computation cost of the proposed method. In the experiments shown in Fig. 16, we measured the computation times of the proposed method and the conventional method in [12]. Since our method and the method in [12] adopt the similar patch priority estimation scheme, we first compare these two methods. Then the computation time of our method was about 2.1-times longer than that of the method in [12], averagely. The experiments were performed on a personal computer using Intel(R) Core(TM) i7 950 CPU 3.06 GHz with 8.0 Ggytes RAM. The implementation was performed by using Matlab. Note that in [34], Kwok et al. reported inpainting that was about 15–50-times faster than that of the method in [12], where the method in [34] adopts a similar reconstruction approach to that of [13] used as the comparison in this paper. The computation cost of the proposed method is thus larger than those of the previously reported methods since the proposed method has to perform the clustering of known patches and the selection of the best matched cluster for each target patch. It is therefore necessary to improve the speed of computation by introducing some alternative approaches into our reconstruction method. This topic will be investigated in subsequent studies.

22

5. Conclusions

In this paper, we have presented a novel missing texture reconstruction method based on KCFA. The proposed method enables derivation of the latent relationship by calculating the KCFA-based optimal orthonormal projection matrices for reconstructing missing areas from their neighboring areas. Furthermore, we used the new criterion monitoring errors caused in the latent space to select the optimal cluster for reconstruction and determine the patch priority. Improvements by the proposed method were confirmed from experimental results.

Acknowledgement

References

[1] S. Masnou and J. Morel, "Level lines based disocclusion," *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 3, pp. 259–263, 1998.

[2] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester, "Image inpainting," *Proc. ACM SIGGRAPH Conference on Computer Graphics*, pp. 417–424, 2000.

[3] M. Bertalmio, A. Bertozzi and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I-355–I362, 2001.

[4] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, "Filling-in by joint interpolation of vector fields and gray levels," *IEEE Trans. on Image Processing*, vol.10, no.8, pp.1200–1211, 2001.

[5] T. F. Chan and J. Shen, "Nontexture inpainting by curvature-driven diffusions," *Journal of Visual Communication and Image Representation,* vol.12, no.4, pp.436–449, 2001.

[6] T. Ogawa, M. Haseyama and H. Kitajima, "Restoration method of missing areas in still images using GMRF model," *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 5, pp. 4931–4934, 2005.

[7] A. Rares, M. J. T. Reinders, and J. Biemond, "Edge-based image restoration," *IEEE Trans. on Image Processing,* vol.14, no.10, pp. 1454–1468, 2005.

[8] D. Z. M.-F. Auclair-Fortier, "A global approach for solving evolutive heat transfer for image denoising and inpainting," *IEEE Trans. on Image Processing*, vol. 15, no.9, pp. 2558–2574, 2006

[9] M. Bertalmio, "Strong-continuation, contrast-invariant inpainting with a third-order optimal PDE," *IEEE Trans. on Image Processing*, vol. 15, no. 7, pp. 1934–1938, 2006.

[10] A. A. Efros and T. K. Leung, "Texture synthesis by nonparametric sampling," *Proc. IEEE International Conference on Computer Vision,* Corfu, Greece, pp.1033–1038, 1999.

[11] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplar-based inpainting," *Proc. IEEE Computer Vision and Pattern Recognition*, 2003.

[12] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[13] T. H. Kwok, H. Sheung, and C. C. L. Wang, "Fast query for exemplar-based image completion," *IEEE Trans. on Image Processing*, vol. 19, no. 12, pp. 3106–3115, 2010.

[14] T. Amano and Y. Sato, "Image interpolation using BPLP method on the eigenspace," *Systems and Computers in Japan*, vol.38, no.1, pp.87–96, 2007.

[15] K. I. Kim, M. O. Franz , B. Schölkopf, "Iterative kernel principal component analysis for image modeling," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.9, pp. 1351–1366, 2005.

[16] T. Ogawa, M. Haseyama, "POCS-based texture reconstruction method using clustering scheme by kernel PCA," *IEICE Trans. Fundamentals*, vol. E90-A, no. 8, pp. 1519–1527, Aug. 2007.

[17] T. Ogawa, M. Haseyama, "Missing intensity interpolation using a kernel PCA-Based POCS algorithm and its applications," *IEEE Trans. on Image Processing*, vol. 20, no. 2, pp. 417–432, 2011.

[18] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. on Image Processing*, vol.17, no.1, pp. 53–69, 2008.

[19] B. Wohlberg, "Inpainting with sparse linear combinations of exemplars," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 689–692, 2009.

[20] B. Shen, W. Hu, Y. Zhang, and Y.-J. Zhang, "Image inpainting via sparse representation," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 697–700, 2009.

[21] T. Ogawa and M. Haseyama, "Missing image data reconstruction based on adaptive inverse projection via sparse representation," *IEEE Trans. on Multimedia*, vol. 13, no. 5, pp. 974-992, 2011.

[22] Z. Xu and J. Sun, "Image Inpainting by Patch Propagation Using Patch Sparsity," *IEEE Trans. on Image Processing*, vol. 19, no. 5, pp. 1153–1165, 2010.

[23] M. Bertalmio, L. Vese, G. Sapiro and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. on Image Processing*, vol. 12, no. 8, pp. 882–889, 2003.

[24] D. Li, N. Dimitrova, M. Li, and I. K. Sethi, "Multimedia content processing through cross-modal association," *Proc. of the ACM International Conference on Multimedia*, pp.604–611, 2003.

[25] Y. Wang and A. V. L. Guan, "Kernel cross-modal factor analysis for multi-modal information fusion," *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2384–2387, 2011.

[26] T. Ogawa and M. Haseyama, "KCFA-based missing area restoration including new priority estimation," *Proc. of the IEEE International Conference on Image Processing 2013 (ICIP 2013)*. (to appear)

[27] B. Schölkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch,

and A. J. Smola, "Input space versus feature space in kernel-based methods," *IEEE Trans. on Neural Networks*, vol. 10, no. 5, pp. 1000–1017, 1999.

[28] J. T. Y. Kwok, and I. W. H. Tsang, "The pre-image problem in kernel methods," *IEEE Trans. on Neural Networks*, vol. 15, no. 6, pp. 1517–1525, 2004.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[30] B. Girod, "What's wrong with mean-squared error?," in *Digital Images and Human Vision*, A.B. Watson, Ed. Cambridge, MA: MIT Press, pp.207–220, 1993.

[31] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. on Image Processing*, vol. 14, no. 12, pp.2117–2128, 2005.

[32] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Morgan & Claypool Publishers*, Mar. 2006.

[33] D. Brunet, E. R. Vrscay and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Trans. on Image Processing*, vol. 21, no. 4, pp. 1488–1499, 2012.

[34] T. H. Kwok and C. C. L. Wang, "Interactive image inpainting using DCT based exemplar matching," *Advances in Visual Computing, Lecture Notes in Computer Science*, vol. 5876/2009, pp.709–718, 2009.
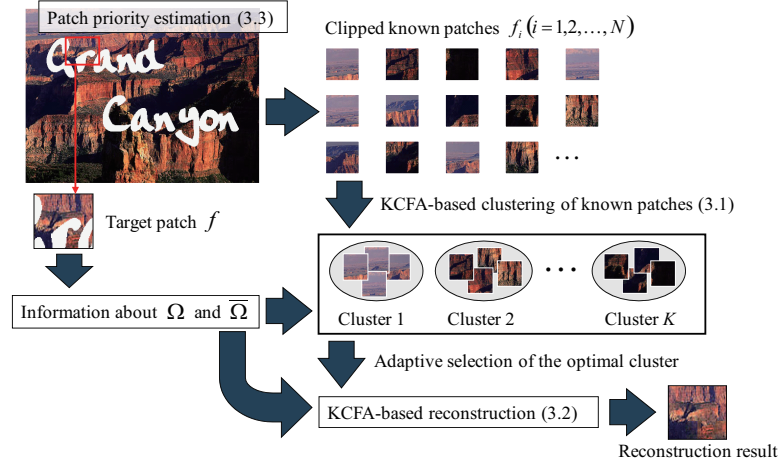
Figure 1: Outline of the proposed method. Our reconstruction method is composed of two algorithms, "KCFA-based clustering of known patches" (3.1) and "KCFA-based reconstruction including adaptive selection of the optimal clusters" (3.2). Furthermore, patch priority estimation is introduced into the proposed KCFA-based reconstruction method (3.3).
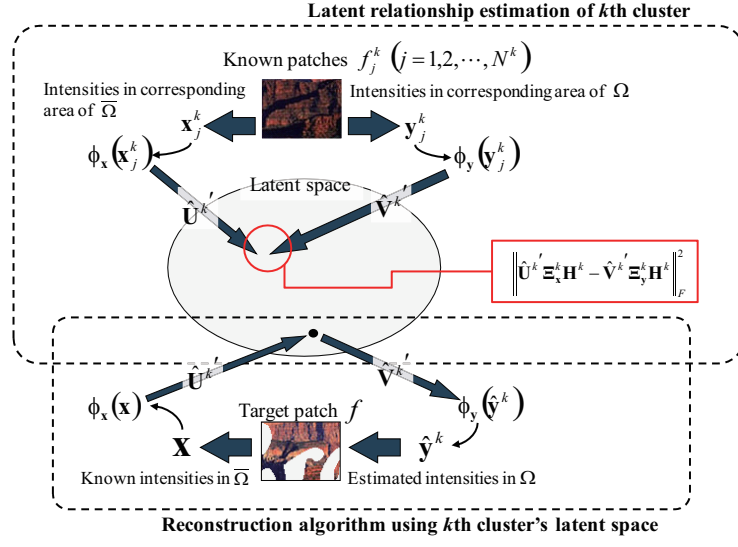
Figure 2: Illustrations of "latent relationship estimation of the kth cluster" and "reconstruction algorithm using the kth cluster's latent space".
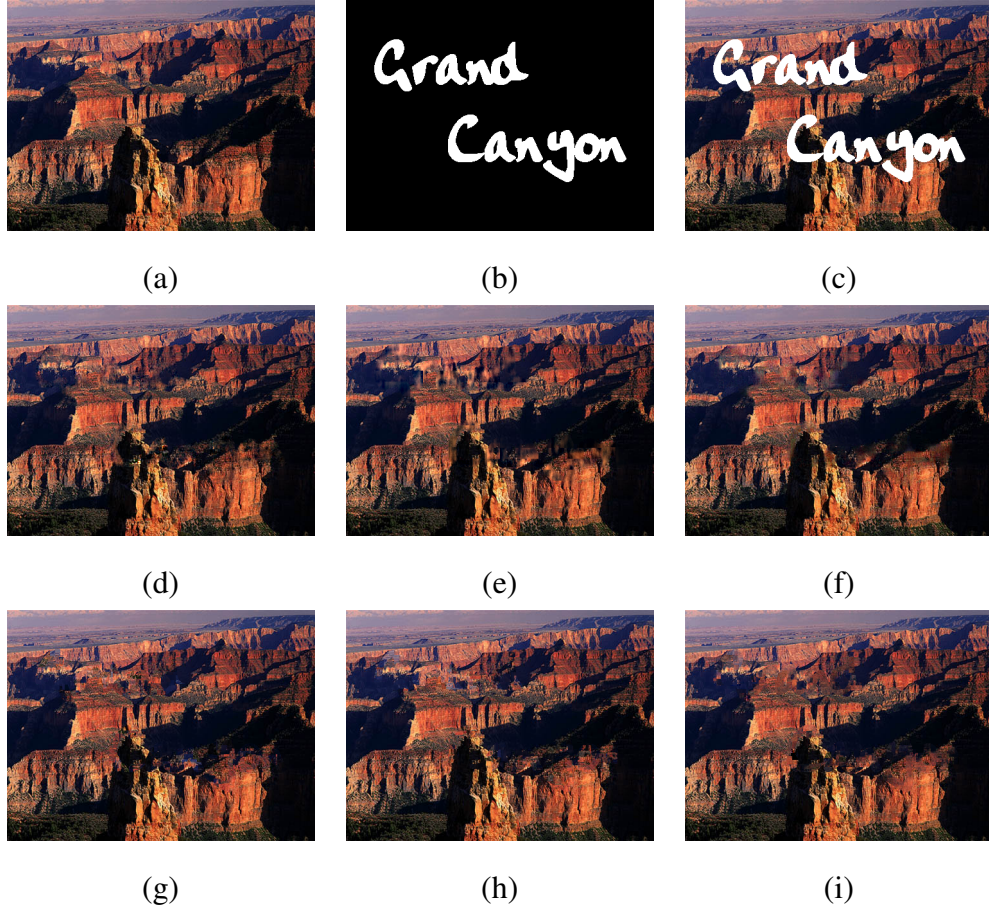
(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)
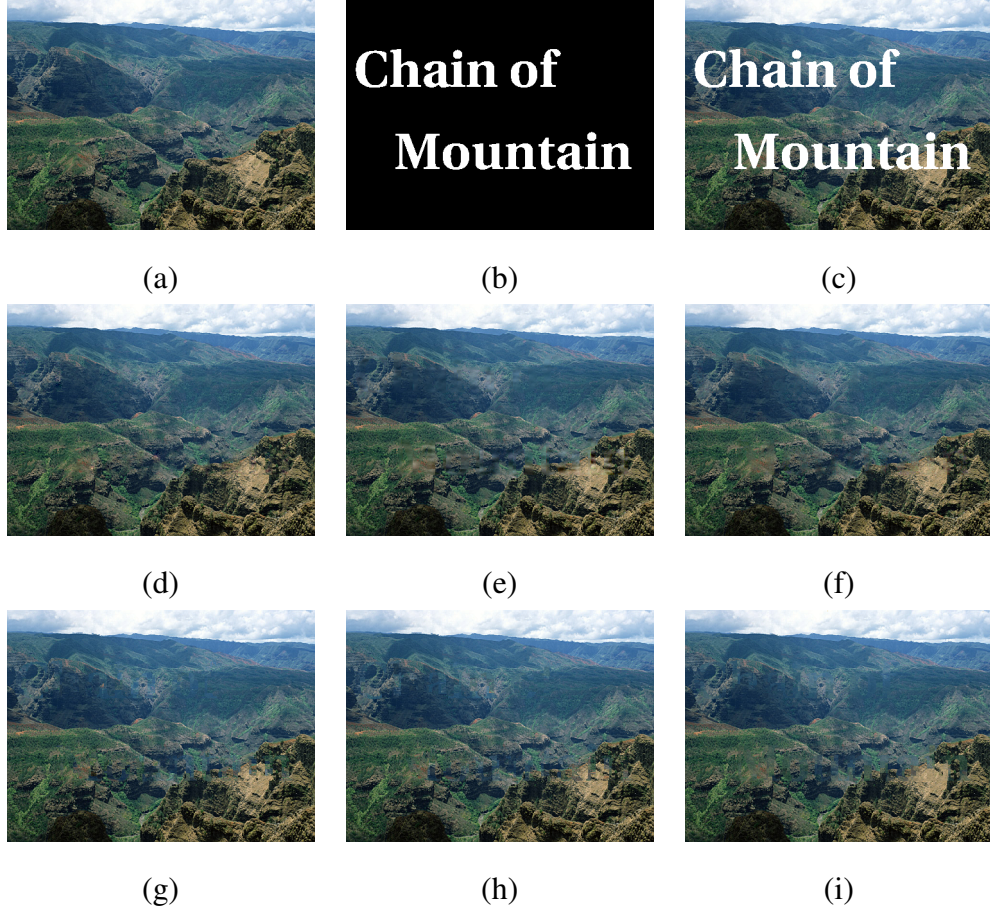
Figure 3: (a) Original image (480 × 359 pixels), (b) Flag image of missing areas, (c) Corrupted image including text regions "Chain of Mountain" (8.9% loss), (d) Reconstructed image by the proposed method, (e) Reconstructed image by the conventional method [14], (f) Reconstructed image by the conventional method [17], (g) Reconstructed image by the conventional method [12], (h) Reconstructed image by the conventional method [13], (i) Reconstructed image by the conventional method [22].
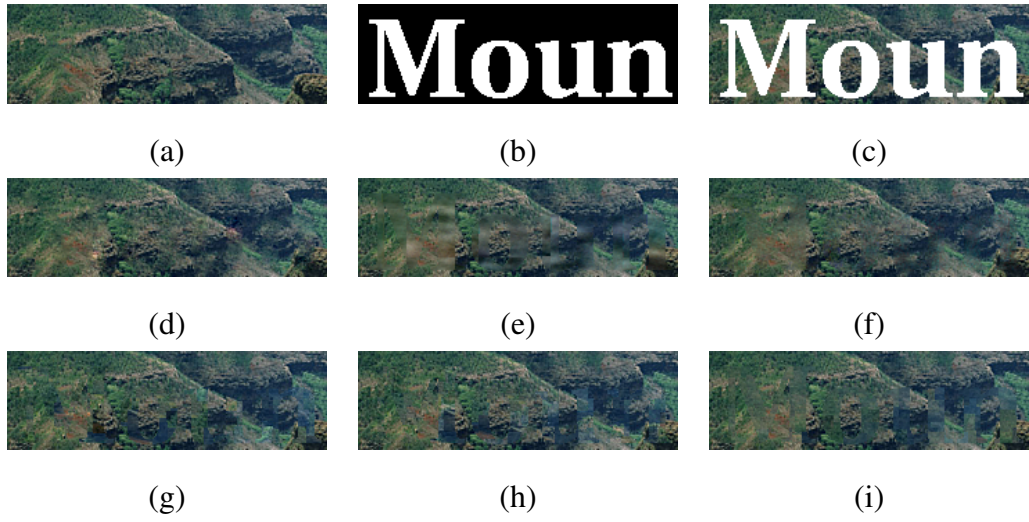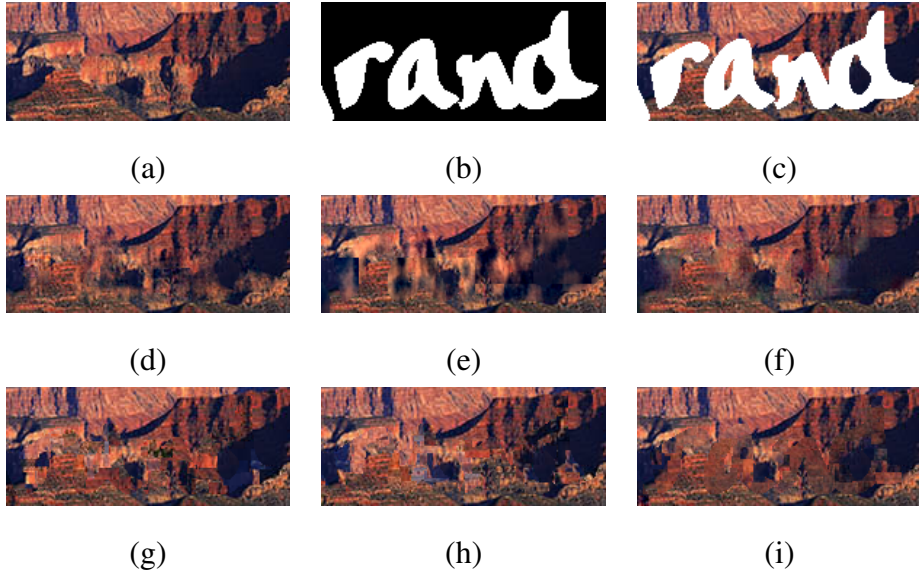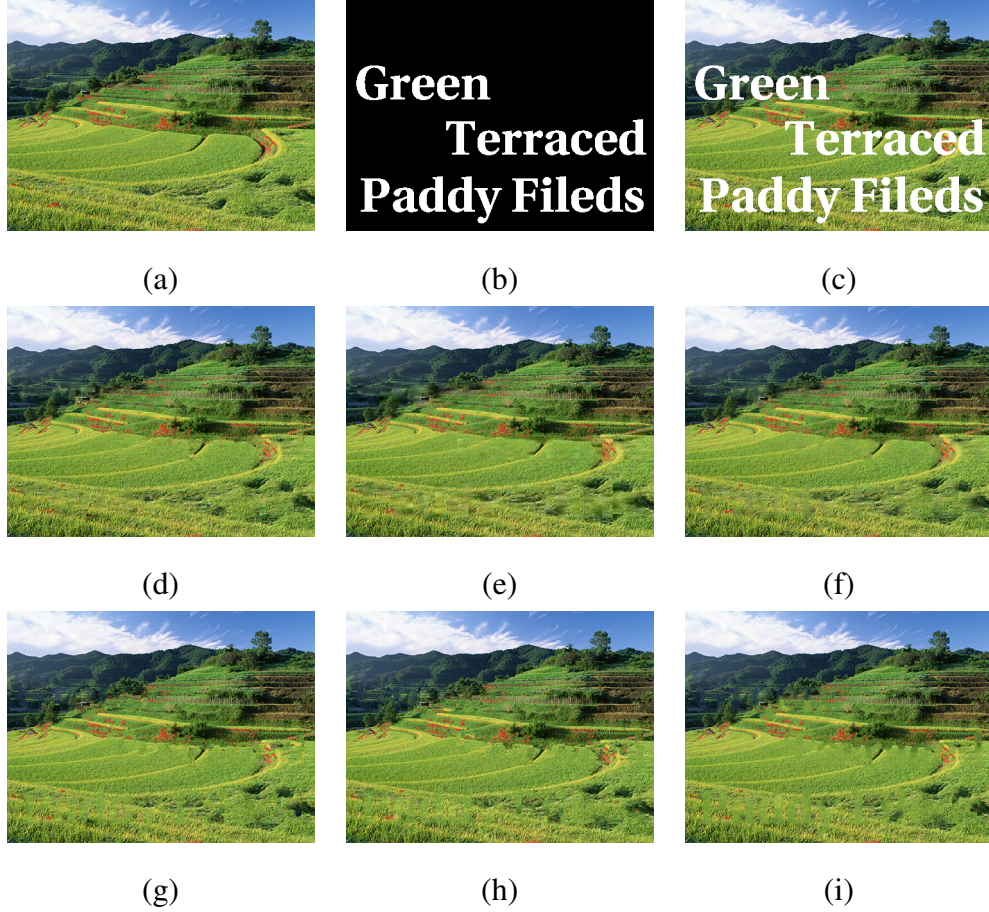
Figure 4: (a)–(i) Zoomed portions of Figs. 3(a)–(i).

Figure 5: (a) Original image (480×360 pixels), (b) Flag image of missing areas, (c) Corrupted image including text regions "Grand Canyon" (8.9% loss), (d) Reconstructed image by the proposed method, (e) Reconstructed image by the conventional method [14], (f) Reconstructed image by the conventional method [17], (g) Reconstructed image by the conventional method [12], (h) Reconstructed image by the conventional method [13], (i) Reconstructed image by the conventional method [22].

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

Figure 6: (a)–(i) Zoomed portions of Figs. 5(a)–(i).

33

Figure 7: (a) Original image (480×360 pixels), (b) Flag image of missing areas, (c) Corrupted image including text regions "Green Terraced Paddy Fileds" (11.9% loss), (d) Reconstructed image by the proposed method, (e) Reconstructed image by the conventional method [14], (f) Reconstructed image by the conventional method [17], (g) Reconstructed image by the conventional method [12], (h) Reconstructed image by the conventional method [13], (i) Reconstructed image by the conventional method [22].

(a)　　　　　　　　(b)　　　　　　　　(c)

(d)　　　　　　　　(e)　　　　　　　　(f)

(g)　　　　　　　　(h)　　　　　　　　(i)
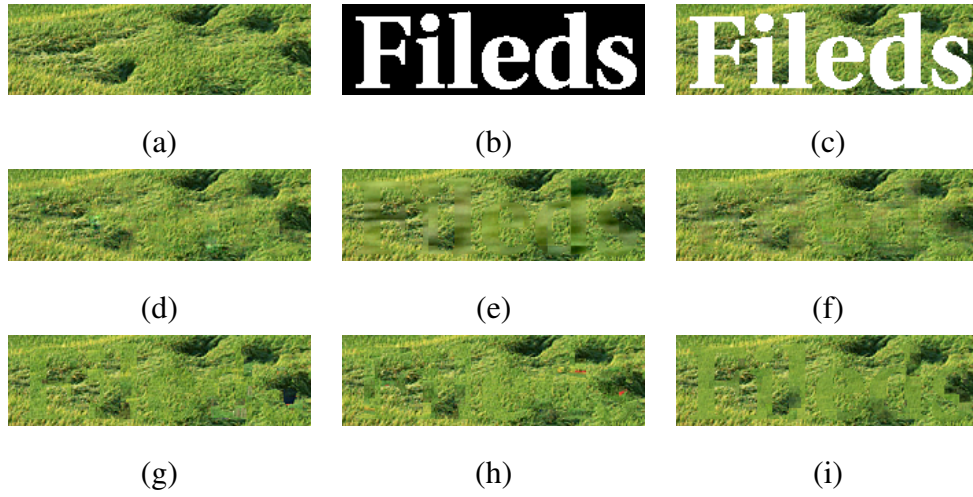
Figure 8: (a)–(i) Zoomed portions of Figs. 7(a)–(i).

Figure 9: (a) Original image (480×360 pixels), (b) Corrupted image including text regions "Buildings in Tokyo" (10.7% loss), (c) Reconstructed image by the proposed method.

Figure 10: (a) Original image (480 × 640 pixels), (b) Corrupted image including text regions "Market of Agricultural Products" (6.2% loss), (c) Reconstructed image by the proposed method. These images are shown by rotating 90 degrees to the right.

Figure 11: (a) Original image (640×480 pixels), (b) Corrupted image including text regions "Blue Sky and Beautiful Cityspace" (5.9% loss), (c) Reconstructed image by the proposed method.
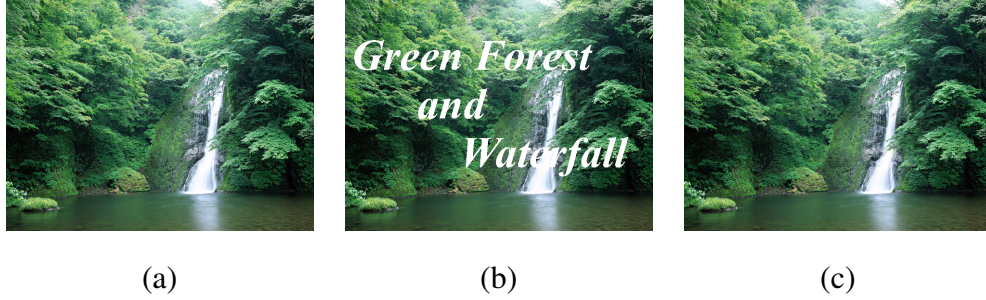
Figure 12: (a) Original image (640 × 480 pixels), (b) Corrupted image including text regions "Green Forest and Waterfall" (6.7% loss), (c) Reconstructed image by the proposed method.
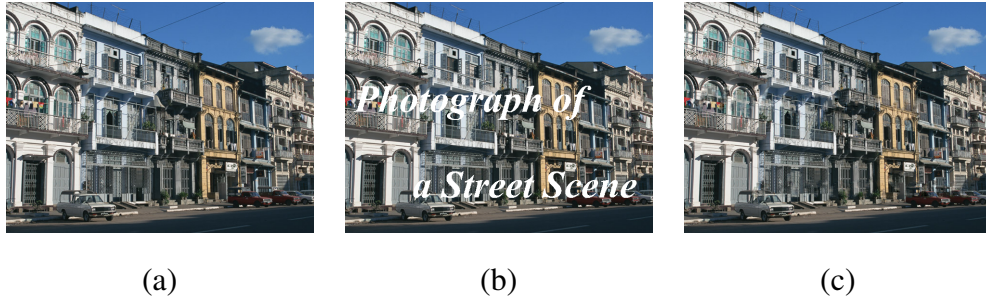

Figure 13: (a) Original image (640 × 480 pixels), (b) Corrupted image including text regions "Photograph of a Street Scene" (5.5% loss), (c) Reconstructed image by the proposed method.
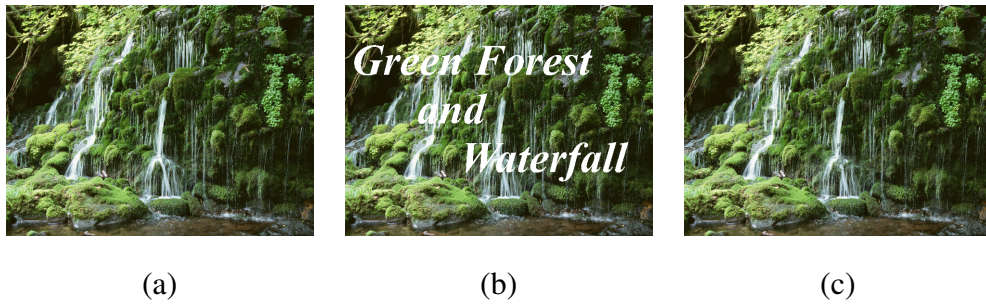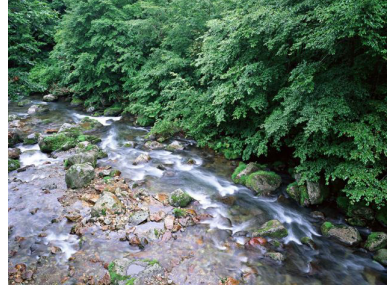

Figure 14: (a) Original image (640 × 480 pixels), (b) Corrupted image including text regions "Green Forest and Waterfall" (6.7% loss), (c) Reconstructed image by the proposed method.

Figure 15: (a)–(f) Test images used for examining the relationship between the ratio of missing pixels and the SSIM index of the reconstruction results obtained by each method. The size of the test images shown in (a)–(f) are 480×360 pixels, 640×480 pixels, 480×640 pixels, 640×480 pixels, 640×480 pixels and 480×360 pixels, respectively. The image in (c) is shown by rotating 90 degrees to the right.
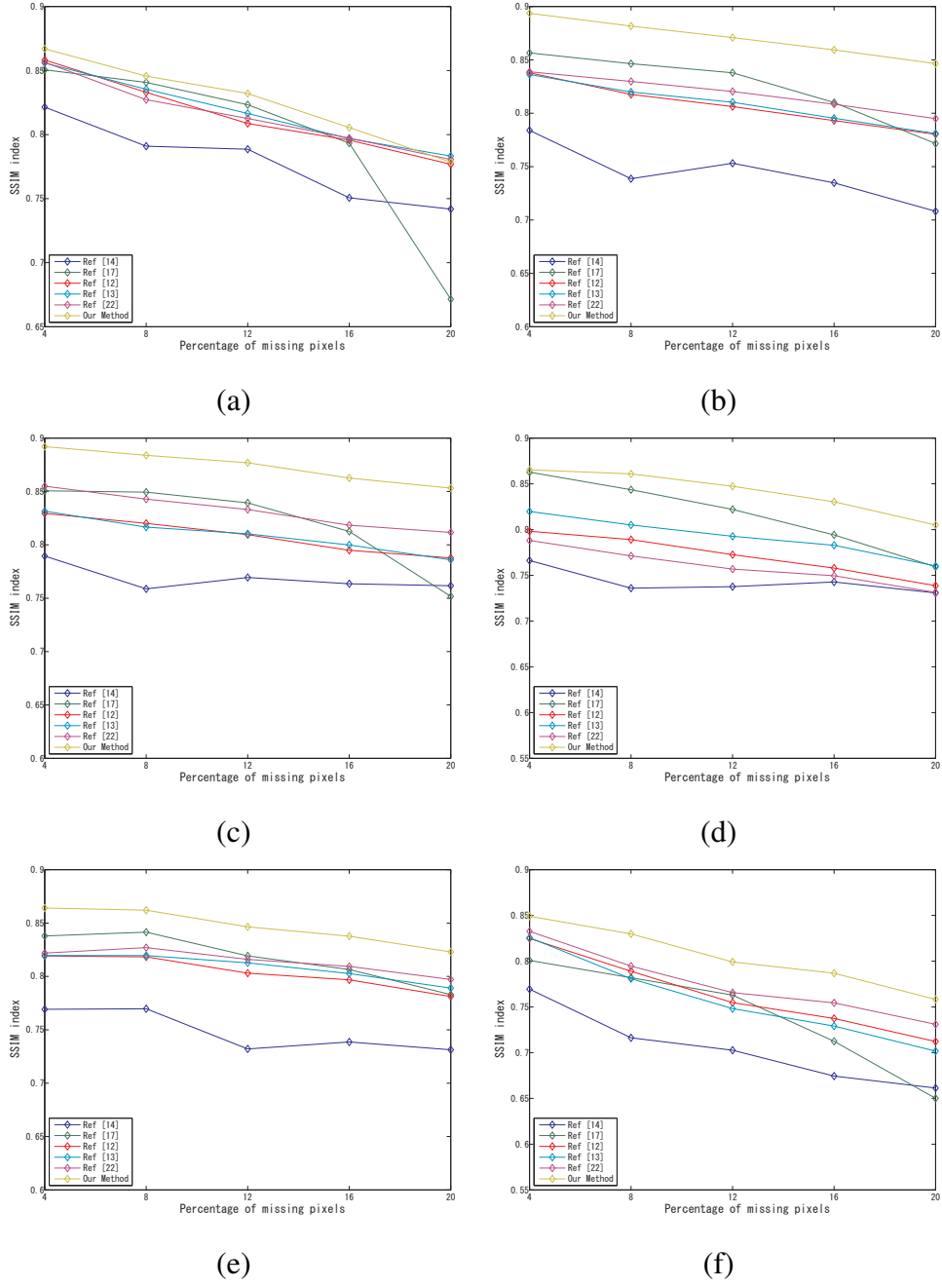
Figure 16: Relationship between the ratio of missing pixels and the SSIM index of the reconstruction results obtained by each method: (a)–(f) respectively show the results obtained from the test images shown in Figs. 15(a)–(f).

39

Figure 17: Reconstruction example of an image including structure components: (a) Original image (480×640 pixels), (b) Corrupted image including missing regions, (c) Reconstructed image by the proposed method, (d) Reconstructed image by our previously reported method [6].

Table 1: Performance comparison (SSIM) of the proposed method and the conventional methods.

Image	Ref [14]	Ref [17]	Ref [12]	Ref [13]	Ref [22]	Our Method
Figure 3	0.6382	0.7311	0.7318	0.7246	0.7635	**0.7949**
Figure 5	0.5239	**0.6475**	0.5077	0.5277	0.5314	0.5886
Figure 7	0.6416	0.6836	0.6822	0.6773	0.7188	**0.7478**
Figure 9	0.5986	0.6893	0.6563	0.6708	0.6949	**0.7416**
Figure 10	0.6217	0.7548	0.6992	0.6980	0.7060	**0.7837**
Figure 11	0.7154	0.7806	0.7155	0.7108	0.7351	**0.8192**
Figure 12	0.6362	0.7208	0.7298	0.7196	0.7331	**0.7931**
Figure 13	0.6710	0.7914	0.6852	0.6799	0.6799	**0.8126**
Figure 14	0.6223	0.7748	0.7366	0.7288	0.7505	**0.8325**