IEICE TRANS. FUNDAMENTALS, VOL.E101–A, NO.11 NOVEMBER 2018

PAPER *Special Section on Smart Multimedia & Communication Systems*

Binary Sparse Representation Based on Arbitrary Quality Metrics and Its Applications

Takahiro OGAWA[†a)], **Sho TAKAHASHI**[††], **Naofumi WADA**[†††], **Akira TANAKA**[†], *and* **Miki HASEYAMA**[†], *Members*

SUMMARY Binary sparse representation based on arbitrary quality metrics and its applications are presented in this paper. The novelties of the proposed method are twofold. First, the proposed method newly derives sparse representation for which representation coefficients are binary values, and this enables selection of arbitrary image quality metrics. This new sparse representation can generate quality metric-independent subspaces with simplification of the calculation procedures. Second, visual saliency is used in the proposed method for pooling the quality values obtained for all of the parts within target images. This approach enables visually pleasant approximation of the target images more successfully. By introducing the above two novel approaches, successful image approximation considering human perception becomes feasible. Since the proposed method can provide lower-dimensional subspaces that are obtained by better image quality metrics, realization of several image reconstruction tasks can be expected. Experimental results showed high performance of the proposed method in terms of two image reconstruction tasks, image inpainting and super-resolution.

key words: *image approximation, binary sparse representation, image quality metrics, visual saliency*

1. Introduction

Image approximation in lower-dimensional subspaces can afford a number of fundamental applications such as image coding, super-resolution and restoration. Generally, the performance of these applications depends on image approximation performance in the given low-dimensional subspaces. Therefore, methods for accurate image approximation in such subspaces are desirable for satisfying demands in the applications.

In recent years, image representation based on multivariate analysis has been intensively studied. The most traditional method is principal component analysis (PCA), which can provide optimal approximation of target samples in the least-square criterion. Due to the rapid development of kernel methods, several image approximation methods based on kernel PCA (KPCA) [1], [2] have also been developed. The methods based on KPCA are suitable for approximating nonlinear visual features in images. Many new approximation methods based on sparse representation [3], [4], which can realize adaptive generation of subspaces, have also been proposed. Other methods using non-negative matrix factorization [5] and manifold learning techniques [6] have also been proposed.

In most studies, image representation was performed on the basis of minimization of approximation errors. The mean square error (MSE) is one of the representative and simplest metrics used for monitoring approximation errors, i.e., quality metrics. On the other hand, it has been reported that the MSE cannot reflect perceptual distortions [7], [8], and MSE-optimal approximation methods do not necessarily output images of high visual quality. For solving the above problems, many image quality metrics have been proposed [9]–[12], and they are widely used for several kinds of applications, e.g., image restoration [13], [14]. Some simple metrics such as the mean absolute error (MAE) and other distances can also improve image approximation. Although optimal quality metrics should be selected for target applications, there remains a big problem. When adopting a new metric, we have to derive a new approximation method. Furthermore, if a target quality metric is not differentiable, it becomes difficult to perform its derivation.

In this paper, binary sparse representation based on arbitrary quality metrics and its applications are presented. The proposed method performs sparse representation for which the image quality metric is arbitrary to enable generation of better low-dimensional subspaces. Specifically, binary sparse representation for which representation coefficients are binary values is newly introduced, and it is the main contribution of this paper. This approach can simplify "calculation of representation coefficients" to "a nearest neighbor search from candidate signal-atoms". It also simplifies "update of the dictionary" to "calculation of average vectors from given samples". Therefore, since the proposed method does not need differentiation of target quality metrics, arbitrary quality metrics can be adopted for the sparse representation. Furthermore, the proposed method uses a new approach that performs pooling of the quality values based on visual saliency, which can reflect human attention [15]–[17]. Then this enables more visually pleasant approximation of the target images. By introducing the above two novel approaches, successful image approximation becomes feasible. Consequently, since the proposed method spans better subspaces, improvement in the performance of several

Manuscript received January 18, 2018.
Manuscript revised May 27, 2018.

[†]The authors are with the Graduate School of Information Science and Technology, Hokkaido University, Sapporo-shi, 060-0814 Japan.
[††]The author is with the Faculty of Engineering, Hokkaido University, Sapporo-shi, 060-8628 Japan.
[†††]The author is with the Department of Information and Computer Science, Hokkaido University of Science, Sapporo-shi, 006-8585 Japan.
a) E-mail: ogawa@lmd.ist.hokudai.ac.jp

DOI: 10.1587/transfun.E101.A.1776

image reconstruction tasks can be expected.

This paper is organized as follows. In Sect. 2, a new image approximation method based on binary sparse representation is presented. In Sect. 3, we discuss the effectiveness of the proposed method and realization of its applications. In Sect. 4, results of experiments are shown for verifying the effectiveness of the proposed method. In this section, we also show results of inpainting and super-resolution (SR) as applications of the proposed method. Finally, concluding remarks are given in Sect. 5.

2. Binary Sparse Representation Based on Arbitrary Quality Metrics

Binary sparse representation based on arbitrary quality metrics is presented in this section. The proposed method simplifies the estimation of sparse representation coefficients and the dictionary by fixing the sparse representation coefficients to binary values. This enables selection of an arbitrary quality metric. Our cost function for the sparse representation is defined by pooling values of a given image quality metric obtained for all of the parts within a target image, with weighting factors for the pooling being determined from visual saliency. This enables the estimation of the dictionary that can successfully approximate visually important features within the target image.

In 2.1, we first show the problem formulation of the proposed method. Then update of the dictionary with binary sparse presentation coefficients is explained in 2.2.

2.1 Problem Formulation

The problem formulation of our method is presented in this subsection. Given a target image, we clip small patches with the same intervals, and their intensity vectors are defined as $\mathbf{x}_i \in \mathcal{R}^M$ $(i = 1, 2, \cdots, N;\ N$ being the number of clipped patches), where M is the dimension of $\mathbf{x}_i$. Then the proposed method tries to solve the following problem:

$$\min_{\mathbf{D},\mathbf{A}} \sum_{i=1}^{N} w_i \,\mathrm{IQM}\left(\mathbf{x}_i, \mathbf{D}\mathbf{a}_i + \mu_i \mathbf{1}_M\right) \quad \text{subject to } ||\mathbf{a}_i||_0 \le T \text{ and } \mathbf{a}_i(k) = 1 \text{ or } 0, \quad (1)$$

where $\mathrm{IQM}(\cdot,\cdot)$ is an arbitrary image quality metric, with the assumption that $\mathrm{IQM}(\cdot,\cdot)$ represents dissimilarity between two given vectors. Then $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_K] \in \mathcal{R}^{M\times K}$ is the dictionary matrix including K signal-atoms, and $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N] \in \mathcal{R}^{K\times N}$ includes all of the representation coefficients $\mathbf{a}_i(k)$ $(i = 1, 2, \cdots, N; k = 1, 2, \cdots, K)$. The vector $\mathbf{1}_M = [1, 1, \cdots, 1]^\top \in \mathcal{R}^M$ is used for representing the direct current component, with its corresponding coefficient being μ_i and equivalent to the average value of $\mathbf{x}_i$, i.e., $\frac{1}{M}\mathbf{1}_M^\top \mathbf{x}_i$, where $^\top$ is a vector/matrix transpose. Furthermore, $||\cdot||_0$ represents the l_0-norm counting the number of non-zero elements.

In Eq. (1), w_i is a weighting factor that is determined for each patch based on visual saliency. For determining the quality of the target image, it is necessary to pool quality values calculated from small regions/pixels. Most methods perform pooling by using the same weight, i.e., the importance of each region becomes the same. However, since it is well known that the importance of each region is different, Ninassi et al., for example, introduced gaze information for pooling image quality values obtained for the whole image [18]. Furthermore, gaze information obtained when staring at images can be modelled as visual saliency [15]–[17], [19]–[21]. Therefore, in the proposed method, visual saliency is adopted for the pooling of quality values. The calculation of w_i based on visual saliency is shown below.

Given a saliency map, the proposed method clips patches including saliency values and calculates their vectors $\mathbf{m}_i \in \mathcal{R}^M$ $(i = 1, 2, \cdots, N)$ that correspond to $\mathbf{x}_i$. Then $v_i = \frac{1}{M}\mathbf{1}_M^\top \mathbf{m}_i$ is defined. The weighting factor w_i is defined by using its empirical distribution function of v_i as follows:

$$w_i = \frac{1}{N}\sum_{j=1}^{N} F_j\left(v_i\right), \quad (2)$$

where

$$F_j(x) = \begin{cases} 1 & \text{if } v_j^s \le x \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

In the above equation, we sort v_i $(i = 1, 2, \cdots, N)$ in ascending order and denote them by v_j^s $(j = 1, 2, \cdots, N)$. The transformation using the empirical distribution function in Eq. (2) corresponds to a histogram equalization operator of v_i $(i = 1, 2, \cdots, N)$. Thus, the weight w_i can be obtained by using arbitrary saliency maps since the histogram equalization provides almost the same distribution regardless of different saliency maps.

2.2 Estimation of Binary Sparse Representation Coefficients and Dictionary

The method for solving Eq. (1) is shown in this subsection. For obtaining the optimal results of $\mathbf{D}$ and $\mathbf{A}$ in Eq. (1), we iteratively update one side while fixing the other side in the same manner as the well-known KSVD algorithm [3]. Although the KSVD algorithm requires complex calculation procedures as reported in [22], [23], much simpler procedures are used in our method, the details of which are shown below.

2.2.1 Update of Binary Sparse Representation Coefficients

The proposed method fixes the dictionary matrix $\mathbf{D}$ and performs update of the representation coefficient vector $\mathbf{a}_i$ for each patch $\mathbf{x}_i$. Specifically, in a way similar to some matching pursuit algorithms [24], [25], we iteratively estimate the non-zero elements in $\mathbf{a}_i$. It should be noted that since all of the non-zero elements are one in the proposed method, we only have to perform selection the non-zero elements.

The proposed method selects tth $(t = 1, 2, \cdots, T)$ optimal non-zero elements for a patch $\mathbf{x}_i$ as

$$\min_{k=\{1,2,\cdots,K|k\notin I_i(t-1)\}} \mathrm{IQM}\left(\mathbf{x}_i\ ,\ \mathbf{d}_k + \sum_{k'\in I_i(t-1)} \mathbf{d}_{k'} + \mu_i \mathbf{1}_M\right), \tag{4}$$

where $I_i(t-1)$ provides a set of indices of signal-atoms previously selected in the $t-1$ iterations for $\mathbf{x}_i$. The above problem is the selection of the signal-atom $\mathbf{d}_k$ optimizing the quality metric, i.e., the dissimilarity between $\mathbf{x}_i$ and $\mathbf{d}_k + \sum_{k'\in I_i(t-1)} \mathbf{d}_{k'} + \mu_i \mathbf{1}_M$. Note that since $\sum_{k'\in I_i(t-1)} \mathbf{d}_{k'} + \mu_i \mathbf{1}_M$ is the approximation result obtained by the $t-1$ iterations, its elements are constants in the tth iteration. As shown in Eq. (4), since the non-zero elements are all one, we do not have to calculate the representation coefficients. Therefore, the approximation result of $\mathbf{x}_i$ becomes only the simple sum of the selected signal-atoms.

Several matching pursuit algorithms generally select the optimal signal-atoms and then calculate the optimal representation coefficients in each iteration. Note that in the procedures for searching for the optimal signal-atoms, representation coefficients of candidate signal-atoms also have to be calculated. Furthermore, in the calculation of the optimal representation coefficients, update of the non-zero elements for the previously selected signal-atoms is necessary. On the other hand, by constraining the non-zero elements to one, our method can drastically simplify the update procedure. In addition, since this procedure corresponds to the nearest neighbor search, our method does not involve any differentiation of the quality metric. Therefore, arbitrary quality metrics can be adopted in the above algorithm.

2.2.2 Update of Dictionary

The proposed method fixes the representation coefficient matrix $\mathbf{A}$ and performs update of the dictionary matrix $\mathbf{D}$. Generally, update of the dictionary is performed for each signal-atom $\mathbf{d}_k$ $(k = 1, 2, \cdots, K)$. For optimizing the signal-atom $\mathbf{d}_k$, it is necessary to take the derivative of the cost function shown in Eq. (1) for $\mathbf{d}_k$. However, several image quality metrics may not be differentiable functions of $\mathbf{d}_k$. On the other hand, in [13], Rehman et al. performed update of the dictionary based on the KSVD algorithm [3] since estimation of the sparse representation coefficients is more important than update of the dictionary. The cost function for updating the dictionary in the KSVD algorithm is based on the MSE, but when concerning the tradeoff between the performance and its complexity, their choice is reasonable. Therefore, by adopting their approach, the proposed method updates the dictionary on the basis of the KSVD algorithm. It should be noted that in the binary sparse representation, update of the KSVD algorithm becomes much simpler. Its details are shown below.

Based on the KSVD algorithm, each signal-atom $\mathbf{d}_k$ is updated with its corresponding non-zero representation coefficients in such a way that the following problem is optimized:

$$\min_{\mathbf{d}_k} \left\| \left\{ \mathbf{d}_k \mathbf{a}_k^{R\top} - \left(\mathbf{X}_k^R - \mathbf{D}_{\bar{k}}^R \mathbf{A}_{\bar{k}}^R \right) \right\} \mathbf{W}_k^{\frac{1}{2}} \right\|_F^2, \tag{5}$$

where $\mathbf{a}_k^R \in \mathcal{R}^{N_k}$ is a vector including only non-zero sparse representation coefficients, N_k being the number of patches for which sparse representation coefficients corresponding to $\mathbf{d}_k$ are not zero. Furthermore, $\mathbf{X}_k^R \in \mathcal{R}^{M\times N_k}$ is a matrix including the above patches, and $\mathbf{D}_{\bar{k}}^R \in \mathcal{R}^{M\times K-1}$ and $\mathbf{A}_{\bar{k}}^R \in \mathcal{R}^{K-1\times N_k}$ are a dictionary matrix in which $\mathbf{d}_k$ is removed and a sparse representation coefficient matrix in which the kth row is removed, respectively. The matrix $\mathbf{W}_k^{\frac{1}{2}} \in \mathcal{R}^{N_k\times N_k}$ is a diagonal matrix including the square root of weight factors in Eq. (1) corresponding to the above patches. Note that the weight matrix is newly introduced since the proposed method adopts the weight factors in Eq. (1).

It should be noted that in the KSVD algorithm in [3], $\mathbf{W}_k^{\frac{1}{2}}$ is the identity matrix, and singular value decomposition of $\mathbf{X}_k^R - \mathbf{D}_{\bar{k}}^R \mathbf{A}_{\bar{k}}^R$ is performed for simultaneously updating $\mathbf{d}_k$ and $\mathbf{a}_k^R$. On the other hand, all of the non-zero elements are one, i.e., $\mathbf{a}_k^R = \mathbf{1}_{N_k}$, in the proposed method. Then since $\mathbf{a}_k^R$ is fixed, we do not have to perform singular value decomposition. Then the update of $\mathbf{d}_k$ can be simplified as calculation of the weighted average of the columns of $\mathbf{X}_k^R - \mathbf{D}_{\bar{k}}^R \mathbf{A}_{\bar{k}}^R$, and it is written by

$$\mathbf{d}_k \leftarrow \frac{1}{\mathbf{1}_{N_k}^\top \mathbf{W}_k \mathbf{1}_{N_k}} \left(\mathbf{X}_k^R - \mathbf{D}_{\bar{k}}^R \mathbf{A}_{\bar{k}}^R \right) \mathbf{W}_k \mathbf{1}_{N_k}, \tag{6}$$

where $\frac{\mathbf{W}_k \mathbf{1}_{N_k}}{\mathbf{1}_{N_k}^\top \mathbf{W}_k \mathbf{1}_{N_k}}$ corresponds to the operator calculating the weighted average. In this way, the simplified update algorithm of the dictionary is realized.

3. Discussion of the Effectiveness and Applications

The discussion of the effectiveness of the proposed method (see 3.1) and its applications (see 3.2) are discussed in this section.

3.1 Effectiveness of Binary Sparse Representation

The most effective point of our binary sparse representation is the use of arbitrary quality metrics. By replacing the calculation of non-zero representation coefficients with only the selection of optimal signal-atoms, we do not have to perform optimization involving the differentiation of adopted image quality metrics. Generally, since we have to take derivatives of the quality metrics with respect to the sparse representation coefficients for the optimization, many quality metrics may not be adopted. Therefore, this is the biggest barrier for realizing sparse representation using better quality metrics. On the other hand, since the non-zero coefficients are all one in our method, we have to determine only which signal-atoms are used. Therefore, the proposed method can adopt any quality metrics without concern about whether their derivatives with respect to the sparse representation

coefficients can be obtained or not.

It should be noted that although the binary sparse representation realizes the use of arbitrary quality metrics, its representation performance becomes worse compared to that of general sparse representation since the representation coefficients are restricted to binary values. Thus, for reducing this problem, it is necessary to set the number of signal-atoms K and the number of non-zero representation coefficients T to larger values. Nevertheless, update of the sparse representation coefficients becomes iteration of the simple nearest neighbor search as shown in Eq. (4). Furthermore, update of the signal-atoms in the dictionary is only calculation of the weighted average as shown in Eq. (5). Therefore, the two important update procedures in the sparse representation can be easily simplified.

In the proposed method, pooling of quality values based on visual saliency is used. In most image representation tasks, bases spanning subspaces are calculated in such a way that the pooled quality values become optimal. For example, when using the MSE as the quality metric, bases minimizing the sum of the MSEs calculated for all of the small patches within target images are generally calculated. Generally, PCA, KPCA and sparse representation provide orthonormal bases and an overcomplete dictionary minimizing the sum of MSEs for all small patches, where the weights of these patches are the same. However, as described above, since the importance of each small patch is different, pooling of the quality values should be performed on the basis of visual saliency. The effectiveness of using visual saliency for pooling the quality values has been shown in several reports [12].

3.2 Applications of Binary Sparse Representation

As shown in the previous section, the proposed method can perform image approximation based on binary sparse representation, i.e., approximation in lower-dimensional subspaces becomes feasible. Therefore, application of the proposed method to several image reconstruction tasks is expected. Specifically, given an original patch and its corresponding corrupted patch in a target image as $\mathbf{x}$ and $\mathbf{y} \in \mathcal{R}^M$, respectively, their relationship can be simply written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{7}$$

where $\mathbf{H} \in \mathcal{R}^{M \times M}$ is a corruption matrix, and $\mathbf{n} \in \mathcal{R}^M$ represents noise, with $\mathbf{n}$ becoming a zero vector in our applications shown in the following section. If the target reconstruction tasks are inpainting and SR, $\mathbf{H}$ becomes a diagonal binary matrix and a matrix representing a low-pass filter, respectively. Given a dictionary $\mathbf{D}$, the estimation result $\hat{\mathbf{x}}$ is obtained as

$$\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{a}} + \hat{\mu}\mathbf{1}_D, \tag{8}$$

where $\hat{\mathbf{a}}$ and $\hat{\mu}$ are obtained as

$$\{\hat{\mathbf{a}}, \hat{\mu}\} = \arg\min_{\mathbf{a},\, \mu} \mathrm{IQM}\left(\mathbf{y}, \mathbf{H}\{\mathbf{D}\mathbf{a} + \mu\mathbf{1}_D\}\right)$$
$$\text{subject to } ||\mathbf{a}||_0 \le T \text{ and } \mathbf{a}(k) = 1 \text{ or } 0. \tag{9}$$

By using the above simple procedures, the proposed method can realize several fundamental applications. As described above, binary sparse representation enables the use of arbitrary image quality metrics. Furthermore, the obtained dictionary enables visually pleasant approximation of images by using visual saliency. Then, based on the lower-dimensional subspaces, successful reconstruction can be expected.

4. Results of Experiments

In this section, we show results of experiments for verifying the effectiveness of the proposed method. First, in 4.1, we verify the performance of image representation by the proposed method. Specifically, we select several quality metrics and saliency maps and verify the image representation performance for determining the optimal combination. Furthermore, in 4.2 and 4.3, we apply the proposed method to image inpainting and SR, respectively. By comparing recent methods, we show the effectiveness of the proposed method in terms of applicability as well as reconstruction performance.

4.1 Image Representation Performance Evaluation

In this experiment, we selected 16 images, which are shown in Fig. 1, from the LIVE Image Quality Assessment Database [9], [26], [27]. For each image, we clipped patches of 8×8 pixels in size and generated 192-dimensional vectors, i.e., $D = 192$, since each pixel included RGB color values. Furthermore, the clipping interval of each patch was set to 8 pixels in width and 8 pixels in height. In the binary sparse representation, K and T were set to 800 and 5, respectively. In this experiment, we used over-complete DCTs for the initial dictionary, and this condition was also adopted in the experiments shown in 4.2 and 4.3. It should be noted that we selected the above conditions in such a way that the difference in the representation performance became clearer.

In the proposed method, we have to provide an image quality metric and a saliency map. We used three kinds of quality metrics, MSE, MAE and l_p-distance with p set to 1.5 ($l_{1.5}$-dist), where MSE and MAE correspond to the l_p-distance with p values of 2 and 1, respectively. Although many image quality metrics have been proposed, they generally try to measure the quality of the whole image. On the other hand, since we calculated the quality of small patches, we used simpler and general metrics in this experiment. Next, two kinds of saliency maps obtained by representative and benchmarking methods (Itti and Hou) [16], [17] were used. For comparison, we adopted a condition in which a saliency map was not used (No saliency), i.e., all of the weights w_i in Eq. (2) were the same value.

Thirteen subjects participated in this experiment, and each subject performed rating for the approximated images with rating scores ranging from 1 (worst) to 5 (best). The results of the subjective evaluation are shown in Table1. From

Fig. 1 Sixteen test images used for evaluating image representation performance of binary sparse representation.

Table 1 Results of subjective evaluation. Average and standard deviation correspond to those of the evaluation scores rated by 13 subjects for 16 test images shown in Fig. 1.

	MSE + No saliency	MSE + Itti [16]	MSE + Hou [17]	$l_{1.5}$-dist + No saliency	$l_{1.5}$-dist + Itti [16]	$l_{1.5}$-dist + Hou [17]	MAE + No saliency	MAE + Itti [16]	MAE + Hou [17]
Average	2.51	2.68	2.78	2.26	2.26	3.09	2.58	2.75	3.32
Standard deviation	0.925	1.09	0.969	1.16	1.19	1.04	1.23	1.31	0.992

(a) (b) (c)

(d) (e) (f)

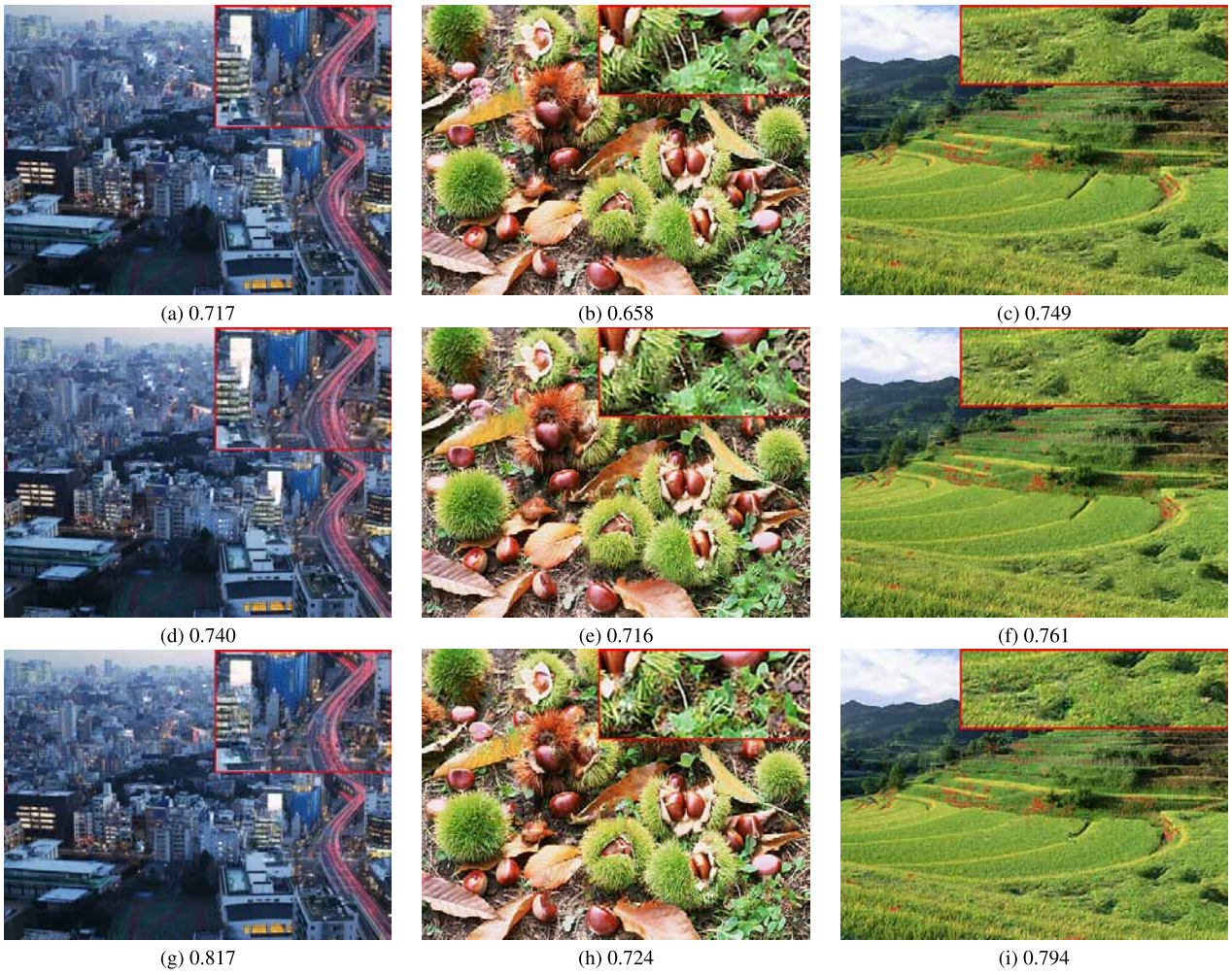
Fig. 2 Test images used for verifying inpainting performance: (a)–(c) original images with sizes of 480×360 pixels, (d)–(f) corrupted images obtained by adding missing areas to (a)–(c).

the obtained results, it can be seen that the average evaluation score becomes highest when adopting the combination of MAE and Hou [17]. This best combination was statistically superior to the other combinations by Welch's t-test with $p < 0.01$ given a significance level $\alpha = 0.01$. Therefore, in the following subsections, we adopted the combination of MAE and Hou [17] for the proposed method.

4.2 Performance Verification of Image Inpainting

In this subsection, we show results of image inpainting obtained by the proposed method. For images shown in Figs. 2(a)–(c), we added missing areas to obtain their corrupted images as shown in Figs. 2(d)–(f), where the positions of missing pixels were known. From the other known regions within the target image, we clipped patches and performed construction of the dictionary in our method, where we set K to 1000 and T to 10 and used a patch size of 8×8 pixels, with a clipping interval half the size of patches. Then for patches including missing pixels, which were selected by patch priority based on [28], the proposed method performed their recovery to obtain the inpainting results. Note that we calculated the average value $\hat{\mu}$ in Eq. (9) from the known parts within the target patch.

In the following, we explain the details of the experimental procedures for inpainting in our method. From a target image including missing areas, we clipped training patches from only known areas to obtain $\mathbf{x}_i \in \mathcal{R}^{192}$, where



(a) 0.717 (b) 0.658 (c) 0.749

(d) 0.740 (e) 0.716 (f) 0.761

(g) 0.817 (h) 0.724 (i) 0.794

Fig. 3 Results of inpainting obtained by using the proposed method and the comparative methods: (a)–(c) results obtained by the method in [28], (d)–(f) results obtained by the method in [29], (g)–(i) results obtained by the proposed method. The values shown in each caption represent the SSIM index [9] calculated from only the recovered regions.

these vectors include RGB pixel values. Furthermore, the weights w_i were obtained from the target image, where the weights became the same for the elements corresponding to the same pixels. By using the training patches $\mathbf{x}_i$ and the weights w_i, the dictionary $\mathbf{D} \in \mathcal{R}^{192\times1000}$ was obtained. From a clipped patch including missing pixels selected by the patch priority, we defined its pixel value vector $\mathbf{y} \in \mathcal{R}^{192}$. Then its reconstruction result $\hat{\mathbf{x}} \in \mathcal{R}^{192}$ in Eq. (8) was obtained by solving Eq. (9). Note that $\mathbf{H}$ became a binary diagonal matrix. Its diagonal elements were one if the corresponding pixels were known. Otherwise, the diagonal elements were zero. By the above procedures, the inpainting was realized.

Experimental results are shown in Fig. 3. For comparison, we used the methods in [28] and [29] as recent and state-of-the-art methods. From the obtained results, it can be seen that the proposed method successfully recovers missing areas by using binary sparse representation in which MAE and Hou [17] are used as the quality metric and saliency map, respectively. Even though the sparse representation coefficients are binary values, the proposed method enables accurate approximation of the target images and can successfully realize the representation of textures.

4.3 Performance Verification of Image Super-Resolution

In this subsection, we show results of SR obtained by the proposed method. In this experiment, we prepared three original high-resolution images shown in Figs. 4(a)–(c) and then performed their downsampling using the well-known Lanczos filter to obtain their quarter-sized images as shown in Figs. 4(d)–(f). These low-resolution images are enlarged to the original sizes in this figure. From the target low-resolution image, we clipped patches for constructing the dictionary and then recovered the high-resolution image patch-by-patch from its upsampled images based on Eq. (9). Note that we could calculate the average value $\hat{\mu}$ in Eq. (9) from the low-resolution images. In our method, we set K to 1000 and T to 10 and used a patch size of 8×8 pixels. The high-

Fig. 4 Test images used for verifying SR performance: (a)–(c) original high-resolution images with sizes of 640 × 480 pixels, (d)–(f) corresponding low-resolution images for which resolution is a quarter of the original size. In this figure, the low-resolution images in (d)–(f) are enlarged to the same size as that of their original images.

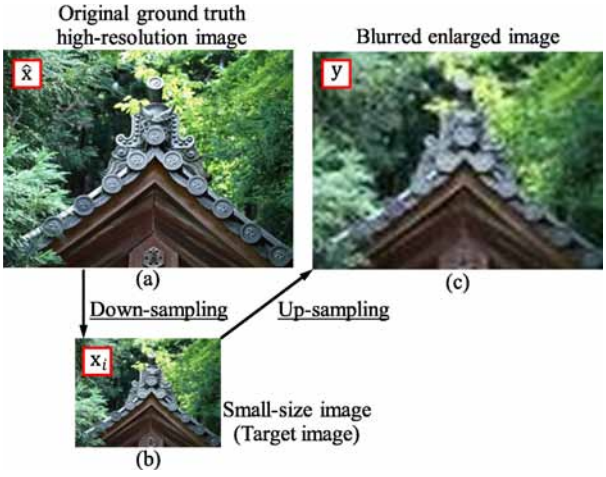


Fig. 5 Relationship between (a) original ground truth high-resolution image, (b) small-size image (target image) and (c) blurred enlarged image.

resolution images were recovered by sliding patches whose sliding interval was 2 pixels in height and 2 pixels in width. Note that since each pixel had multiple results, we output their average values as the final output.

In the following, we explain the details of the experimental procedures for SR in our method. Given an original ground truth high-resolution image (Fig. 5(a)), we downsampled this image to obtain its small-size image (Fig. 5(b)). We denote this low-resolution image as a target image. From this target image (Fig. 5(b)), we clipped patches as training high-resolution patches. Then, from the training high-resolution patches, we obtained intensity vectors $\mathbf{x}_i \in \mathcal{R}^{64}$, where the SR was performed for only the luminance components in our method. Furthermore, the weights w_i were obtained from the small-size image (Fig. 5(b)). From $\mathbf{x}_i$ and w_i, the dictionary $\mathbf{D} \in \mathcal{R}^{64\times1000}$ could be obtained. In our method, by applying up-sampling to the target image (Fig. 5(b)), we obtained a blurred enlarged image (Fig. 5(c)). From this image, we clipped a patch and defined its intensity vector $\mathbf{y}$. The intensity vector $\hat{\mathbf{x}}$ in Eq. (8) of the corresponding high-resolution patch in the original high-resolution image (Fig. 5(a)) was estimated based on Eq. (9), where $\mathbf{H}$ was the operator of the low-pass filter (Lanczos filter). In this way, our method realized SR for the target image (Fig. 5(b)) to estimate its high-resolution image (Fig. 5(a)). It should be noted that for chroma components, we simply used those of the blurred enlarged image (Fig. 5(b)) which could be directly obtained from the target image (Fig. 5(b)).

In Fig. 6, the results of SR obtained by the state-of-the-art methods [30]–[32] and the proposed method are shown. Compared to the results of inpainting shown in the previous subsection, the difference between the proposed method and other comparative methods does not seem to be significant, but the proposed method tends to preserve the sharpness in the recovery results of the obtained high-resolution images.

5. Conclusions

A binary sparse representation method based on arbitrary image quality metrics was presented in this paper. The proposed method newly derives binary sparse representation for

Fig. 6 Results of SR obtained by using the proposed method and the comparative methods: (a)–(c) results obtained by the method in [30], (d)–(f) results obtained by the method in [31], (g)–(i) results obtained by the method in [32], (j)–(l) results obtained by the proposed method. The values shown in each caption represent the SSIM index [9] calculated from the obtained results.

which sparse representation coefficients are binary values. This approach realizes the use of arbitrary image quality metrics, and it is the main contribution of this paper. Furthermore, visual saliency is used in the proposed method for pooling the quality values to construct a dictionary that can successfully approximate visually important parts. By using these two novel approaches, the proposed method can realize successful approximation of patches within images in lower-dimensional subspaces. Therefore, high performance in several image reconstruction tasks can be also realized. In experiments, the effectiveness and applicability of the proposed method were shown by applying it to image inpainting and SR.

Finally, we refer to the computation cost of our method. The average computation time for the dictionary learning and the reconstruction in inpainting were 24.6 sec and 209.8 sec, respectively. Furthermore, those in SR were 12.2 sec and 100.9 sec, respectively. The experiments were performed

on a personal computer using Intel(R) Core(TM) i7 980 CPU 3.33 GHz with 4.0 Ggytes RAM. The implementation was performed by using Matlab. Although one of the advantage of our method is simplicity, the implementation of our method was not optimized for the fast processing. The computation time will become faster by improving this implementation. Furthermore, as stated above, calculation of the sparse representation coefficients becomes a nearest neighbor search in our method. In recent years, many fast searching algorithms have been proposed in the field of data mining and multimedia data retrieval. Therefore, by introducing these algorithms into the proposed method, a faster version could be implemented. This will be addressed in our future work.

Acknowledgements

This work was partly supported by JSPS KAKENHI Grant Numbers JP17H01744, JP18K11367.

References

[1] 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. Neural Netw., vol.10, no.5, pp.1000–1017, 1999.

[2] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," Advances in Neural Information Processing Systems, vol.11, pp.536–542, 1999.

[3] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process., vol.54, no.11, pp.4311–4322, 2006.

[4] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process., vol.15, no.12, pp.3736–3745, 2006.

[5] D.D. Lee and H.S. Seung, "Learning the parts of objects with nonnegative matrix factorization," Nature, vol.401, no.6755, pp.788–791, 1999.

[6] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol.290, no.5500, pp.2323–2326, 2000.

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

[8] Z. Wang and A.C. Bovik, Modern Image Quality Assessment, Morgan & Claypool Publishers, March 2006.

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

[10] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process., vol.20, no.8, pp.2378–2386, 2011.

[11] W. Xue, L. Zhang, X. Mou, and A.C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," IEEE Trans. Image Process., vol.23, no.2, pp.684–695, 2014.

[12] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," IEEE Trans. Image Process., vol.23, no.10, pp.4270–4281, 2014.

[13] A. Rehman, M. Rostami, Z. Wang, D. Brunet, and E.R. Vrscay, "SSIM-inspired image restoration using sparse representation," EURASIP J. Adv. Signal Process., vol.2012, 16, 2012.

[14] T. Ogawa and M. Haseyama, "Image inpainting based on sparse representations with a perceptual metric," EURASIP J. Adv. Signal Process., vol.2013, 179, 2013.

[15] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," Human Neurobiology, vol.4, no.4, pp.219–227, 1985.

[16] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.11, pp.1254–1259, 1998.

[17] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp.1–8, 2007.

[18] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," Proc. IEEE International Conference on Image Processing (ICIP), pp.II-169–II-172, 2007.

[19] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency," Advances in Neural Information Processing Systems 19, pp.545–552, MIT Press, 2007.

[20] S. Goferman, L.Z.-Manor and A. Tal, "Context aware saliency detection" IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.10, pp.1915–1926, 2012.

[21] L. Zhang, Z. Gu, and H. Li, "SDSP: A novel saliency detection method by combining simple priors," Proc. IEEE International Conference on Image Processing (ICIP), pp.171–175, 2013.

[22] S.K. Sahoo and A. Makur, "Dictionary training for sparse representation as generalization of k-means clustering," IEEE Signal Process. Lett., vol.20, no.6, pp.587–590, 2013.

[23] S.K. Sahoo and A. Makur, "Sparse sequential generalization of k-means for dictionary training on noisy signals," Signal Process., vol.129, pp.62–66, 2016.

[24] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Process., vol.41, no.12, pp.3397–3415, 1993.

[25] J.A. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. Inf. Theory, vol.50, no.10, pp.2231–2242, 2004.

[26] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "LIVE Image Quality Assessment Database Release 2," http://live.ece.utexas.edu/research/quality

[27] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process., vol.15, no.11, pp.3440–3451, 2006.

[28] C. Guillemot, M. Turkan, O.L. Meur, and M. Ebdelli, "Object removal and loss concealment using neighbor embedding methods," Signal Processing: Image Communication, vol.28, no.10, pp.1405–1419, 2013.

[29] Z. Li, H. He, H.-M. Tai, Z. Yin, and F. Chen, "Color-direction patch-sparsity-based image inpainting using multidirection features," IEEE Trans. Image Process., vol.24, no.3, pp.1138–1152, 2015.

[30] K. Zhang, D. Tao, X. Gao, X. Li, and Z. Xiong, "Learning multiple linear mappings for efficient single image super-resolution," IEEE Trans. Image Process., vol.24, no.3, pp.846–861, 2015.

[31] F. Zhou, T. Yuan, W. Yang, and Q. Liao, "Single-image super-resolution based on compact KPCA coding and kernel regression," IEEE Signal Process. Lett., vol.22, no.3, pp.336–340, 2015.

[32] J. Jiang, X. Ma, Z. Cai, and R. Hu, "Sparse support regression for image super-resolution," IEEE Photonics J., vol.7, no.5, pp.1–11, 2015.

Takahiro Ogawa received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2003, 2005 and 2007, respectively. He joined Graduate School of Information Science and Technology, Hokkaido University in 2008. He is currently an associate professor in the Graduate School of Information Science and Technology, Hokkaido University. His research interests are multimedia signal processing and its applications. He has been an Associate Editor of ITE Transactions on Media Technology and Applications. He is a member of the IEEE, ACM, EURASIP, IEICE, and ITE.

Sho Takahashi received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2008, 2010 and 2013, respectively. He joined the Graduate School of Information Science and Technology, Hokkaido University, as an Assistant Professor in 2013. He is currently an Associate Professor in the Education and Research Center for Mathematical and Data Science, Hokkaido University. His research interests include semantic analysis and visualization in videos. He is a member of the IEEE, IEICE and Institute of Image Information and Television Engineers (ITE).

Naofumi Wada received his B.S. and M.S. degrees in Engineering and Ph.D. degree in Electronics and Information Engineering from Hokkaido University, Japan in 2002, 2004 and 2015, respectively. He worked at the R&D Center of Toshiba Corporation from 2004 to 2009 and Samsung R&D Institute Japan from 2009 to 2016. He is currently a Lecturer in the Dept. of Information and Computer Science, Faculty of Engineering, Hokkaido University of Science. His research interests include image and video processing, video codecs, and computer vision. He is a member of the IEICE, IPSJ and ITE.

Akira Tanaka received the D.E. degree from Hokkaido University, Sapporo, Japan, in 2000. He is with the Graduate School of Information Science and Technology, Hokkaido University. His research interests include image processing, acoustic signal processing, and machine learning.

Miki Haseyama received her B.S., M.S. and Ph.D. degrees in Electronics from Hokkaido University, Japan in 1986, 1988 and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University as an associate professor in 1994. She was a visiting associate professor of Washington University, USA from 1995 to 1996. She is currently a professor in the Graduate School of Information Science and Technology and a Director of Education and Research Center for Mathematical and Data Science, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She has been a Vice-President of the Institute of Image Information and Television Engineers, Japan (ITE), an Editor-in-Chief of ITE Transactions on Media Technology and Applications, a Director, International Coordination and Publicity of The Institute of Electronics, Information and Communication Engineers (IEICE). She is a member of the IEEE, IEICE, ITE and ASJ.