Image Inpainting Based on Energy Minimization

Norihiko Kawai, Tomokazu Sato and Naokazu Yokoya

Graduate School of Information Science, Nara Institute of Science and Technology,
Takayama 8916-5, Ikoma, Nara, 630-0192, Japan

ABSTRACT

Image inpainting techniques have been widely used to remove undesired visual objects in images such as damaged portions of photographs and people who have accidentally entered into pictures. Conventionally, the missing parts of an image are completed by optimizing the objective function which is defined based on the sum of SSD (sum of squared differences). However, the naive SSD-based objective function is not robust against intensity change in an image. Thus, unnatural intensity change often appears in the missing parts. In addition, when an image has continuously changing texture patterns, the completed texture in a resultant image sometimes blurs due to inappropriate pattern matching. In this paper, in order to improve the image quality of the completed texture, the conventional objective function is newly extended by considering intensity changes and spatial locality to prevent unnatural intensity changes and blurs in a resultant image. By minimizing the extended energy function, the missing regions can be completed without unnatural intensity changes and blurs. In experiments, the effectiveness of the proposed method is successfully demonstrated by applying our method to various images and comparing the results with those obtained by the conventional method.

Keywords: image inpainting, image completion, energy minimization

1. INTRODUCTION

With the spread of Internet, it has become usual for people to upload photographs and videos to their personal homepage and blog. For this purpose, in case of analog pictures taken in the past are digitalized with a scanner, it might be difficult to use those digitized pictures due to physical damages such as scratches and stains in analog pictures. In addition, even if there are no physical damages on digitized pictures and videos, it might be hard to use them when undesired objects such as people who have accidentally entered into pictures are taken in the pictures or videos. For such problems, image inpainting techniques have been widely investigated to remove undesired visual objects in images.

In simple image inpainting techniques[1,2] , missing regions are interpolated considering the continuity of pixel intensity. These methods fill missing regions by propagating information from the boundaries of missing regions while preserving edges. They are effective for small image gaps like scratches in a photograph. However, the resultant image easily becomes unclear when the missing region is large. In order to solve this problem, the methods which generate the image by synthesizing textures in the missing region have been investigated. Until now, several texture synthesis methods have been proposed; that is, the method that interpolates the missing region by generating eigen-space[3] , the method that completes the missing region by synthesizing the texture from the rest of the image successively[4,5] and the method that optimizes the texture in the missing region by using the objective function representing the plausibility of the missing region[6] . In the following, these methods are described in some more detail.

Amano et al.[3] use a subspace constructed by image vectors to interpolate the missing region. In this method, eigen-space is constructed by cutting many windows from the rest of the image as learning samples and the missing region is filled by combining eigen vectors. This method can complete the missing region successfully without image blurs. However, it is difficult to apply the method to the image with large missing regions due to the methodological limitation such that a cut window has to include both the missing region and the data region.

Further author information:
N.K,T.S,N.Y: E-mail: {norihi-k, tomoka-s, yokoya}@is.naist.jp

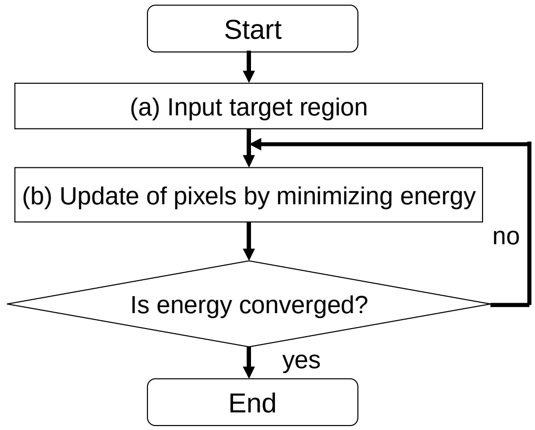


Figure 1. Procedure of the proposed method.

Criminisi et al.[4] complete the missing region by texture synthesis. In this approach, appropriate textures similar to those in the boundary of the missing region are searched for from the data region and by pasting the most similar texture from the boundary to the center of missing region, the missing region is successively completed. However, this method has a problem. Although the duplication of similar textures preserves the local texture continuity, discontinuous textures are easily synthesized in the completed image due to the way that the missing region is completed successively from the boundary. In order to avoid this problem, Sun at el.[5] have proposed a technique that fills the missing region by synthesizing textures preferentially along the lines that are specified manually by users considering the continuity of edges. However, when an image has complex textures, it is difficult for the user to specify the appropriate lines for image completion.

Wexeler et al.[6] have proposed the method that completes the missing region by optimizing the objective function. The function represents the plausibility of the image and is defined based on the sum of SSD (sum of squared differences). This method can be applied for not only a static image but also a video sequence and the method can generate the optimal image independent of texture filling order. However, the naive SSD-based objective function is not robust against intensity change in an image. Thus, unnatural intensity change often appears in the missing region. In addition, when an image contains continuously changing texture patterns, the completed texture in a resultant image sometimes blurs due to inappropriate pattern matching.

In this paper, we propose a new energy function which is defined by expanding the conventional objective function proposed by Wexeler[6] considering intensity changes and spatial locality to prevent unnatural intensity changes and blurs in a completed image. By minimizing the newly defined energy function, the missing region can be successfully completed without unnatural intensity changes and blurs.

2. IMAGE INPAINTING BY MINIMIZING A ENERGY FUNCTION

Figure 1 illustrates the procedure of the proposed method. First, a user manually selects regions to be repaired such as physically damaged regions and undesired object regions in an image (a). Next, selected regions are completed by minimizing the energy function (b). In the following sections, first, we describe the conventional objective function which is defined based on pattern similarity SSD in Section 2.1. The proposed energy function is then defined in Section 2.2 and the inpainting procedure by minimizing the energy function is described in Section 2.3.

2.1. Energy function based on pattern similarity

We describe the SSD-based objective function for image inpainting originally proposed by Wexeler et al.[6] . Here, although the original objective function is defined as a probability density function, we re-define this objective function as an equivalent energy function.

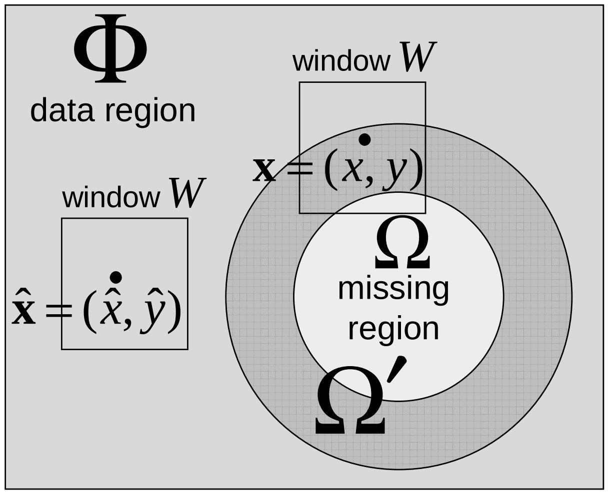


Figure 2. Illustration of missing and data regions in an image.

As illustrated in Figure 2, first, an image is divided into the region Ω' including the missing region Ω selected by a user and the data region Φ, which is the rest of the image. The plausibility in the missing region Ω is defined by using image patterns in the data region Φ. Here, Ω' is the expanded area of the missing region Ω in which there is a central pixel of a square window W of size N_W(where N_W is constant) overlapping the region Ω. The energy function which represents the plausibility in the missing region is defined as the weighed sum of SSD between the pixels around the pixel $\mathbf{x}$ in the region Ω' and those around the pixel $\hat{\mathbf{x}}_{org}$ in the region Φ as follows:

$$E_{org} = \sum_{\mathbf{x}\in\Omega'} w_{\mathbf{x}} SSD(\mathbf{x}, \hat{\mathbf{x}}_{org}), \tag{1}$$

where $\hat{\mathbf{x}}_{org}$ in the data region Φ denotes the pixel around which the pattern is most similar to that around $\mathbf{x}$ in the region Ω', and $SSD(\mathbf{x}, \hat{\mathbf{x}}_{org})$ is defined as follows.

$$SSD(\mathbf{x}, \hat{\mathbf{x}}_{org}) = \sum_{\mathbf{p}\in W} \{I(\mathbf{x}+\mathbf{p}) - I(\hat{\mathbf{x}}_{org}+\mathbf{p})\}^2. \tag{2}$$

Here, $I(\mathbf{x})$ represents the intensity of pixel $\mathbf{x}$. The pixel $\hat{\mathbf{x}}_{org}$ for minimizing E_{org} is decided as follows.

$$\hat{\mathbf{x}}_{org} = f_{org}(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{x}'\in\Phi} SSD(\mathbf{x}, \mathbf{x}'). \tag{3}$$

Note that, the weight $w_{\mathbf{x}}$ is set as 1 if $\mathbf{x}$ is inside of the region $\Omega' \cap \overline{\Omega}$ because pixel values in this region are fixed, otherwise $w_{\mathbf{x}}$ is set as c^{-d} (d is the distance from the boundary of Ω, c is constant) because pixel values around the boundary have more confidence than those in the center of the missing region.

In the Wexeler's work[6] , the missing region is completed by calculating the pixel value $I(\mathbf{x})$ in the missing region and the position of the pixel $\hat{\mathbf{x}}_{org}$ that minimize the energy function E_{org}.

2.2. Extension of energy function considering intensity change and spatial locality

In this study, we extend the original energy function E_{org} defined in Eq. (1) by considering intensity change and spatial locality of texture pattern. Concretely, we introduce a modification coefficient to allow linear intensity change of image pattern. For considering spatial pattern locality, the term for distance between the pixel in the missing region and the corresponding pixel in the data region is also added to the original energy function. The extended energy function is defined as follows.

$$E = \sum_{\mathbf{x}\in\Omega'} w_{\mathbf{x}} \left\{SSD'(\mathbf{x}, \hat{\mathbf{x}}) + w_{dis} AD(\mathbf{x}, \hat{\mathbf{x}})\right\}, \tag{4}$$

where $SSD'(\mathbf{x}, \hat{\mathbf{x}})$ represents the pattern similarity considering intensity change, and $AD(\mathbf{x}, \hat{\mathbf{x}})$ means the Euclid distance between the pixel $\mathbf{x}$ and the pixel $\hat{\mathbf{x}}$. Each term is defined as follows:

$$SSD'(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{\mathbf{p} \in W} \left\{ I(\mathbf{x} + \mathbf{p}) - \alpha_{\mathbf{x}\hat{\mathbf{x}}} I(\hat{\mathbf{x}} + \mathbf{p}) \right\}^2, \tag{5}$$

$$AD(\mathbf{x}, \hat{\mathbf{x}}) = \parallel \mathbf{x} - \hat{\mathbf{x}} \parallel, \tag{6}$$

where $\hat{\mathbf{x}}$ is determined as the pixel position that minimizes the extended energy function E as follows:

$$\hat{\mathbf{x}} = f(\mathbf{x}) = \underset{\mathbf{x}' \in \Phi}{\operatorname{argmin}} (SSD'(\mathbf{x}, \mathbf{x}') + w_{dis} AD(\mathbf{x}, \mathbf{x}')). \tag{7}$$

In this paper, we employ the ratio of average pixel values around the pixels $\mathbf{x}$ and $\hat{\mathbf{x}}$ as a modification coefficient $\alpha_{\mathbf{x}\hat{\mathbf{x}}}$. However, an unnatural image is easily generated if large intensity change is also approximated by linear transformation. Therefore, we limit the range of the value $\alpha_{\mathbf{x}\hat{\mathbf{x}}}$ where intensity change can be approximated linearly as given in Eq. (8).

$$\alpha_{\mathbf{x}\hat{\mathbf{x}}} = \begin{cases} 1 - D & (\beta_{\mathbf{x}\hat{\mathbf{x}}} < 1 - D) \\ \beta_{\mathbf{x}\hat{\mathbf{x}}} & (1 - D \leq \beta_{\mathbf{x}\hat{\mathbf{x}}} \leq 1 + D) \\ 1 + D & (\beta_{\mathbf{x}\hat{\mathbf{x}}} > 1 + D), \end{cases} \tag{8}$$

where D is a constant $(0 < D < 1)$ and $\beta_{\mathbf{x}\hat{\mathbf{x}}}$ is defined as follows:

$$\beta_{\mathbf{x}\hat{\mathbf{x}}} = \frac{\sqrt{\sum_{\mathbf{q} \in W} I(\mathbf{x} + \mathbf{q})^2}}{\sqrt{\sum_{\mathbf{q} \in W} I(\hat{\mathbf{x}} + \mathbf{q})^2}}. \tag{9}$$

2.3. Update of pixel values by minimizing energy function

The energy function E defined in Eq. (4) is minimized by using a framework of greedy algorithm. In our method, we pay attention to the fact that the energy function E for each pixel can be treated independently if similar pattern pairs $(\mathbf{x}, \hat{\mathbf{x}})$ calculated by Eq. (7) can be fixed and the change of coefficient $\alpha_{\mathbf{x}\hat{\mathbf{x}}}$ is very small. Thus, we repeat the following two processes until the energy converges: (I) parallel update of all the pixel values in the missing region for fixed similar pairs of windows in the missing region and the data region, and (II) update of pairs of windows for fixed pixel colors. In order to avoid local minima efficiently, a coarse-to-fine approach is also employed. In the following, we describe the processes (I) and (II) in detail.

2.3.1. Parallel update of pixel values in missing region

In the process (I), we update all the pixel values $I(\mathbf{x})$ in the missing regions in parallel by minimizing the energy defined by Eq. (4). In the following, we describe the method for calculating the pixel values $I(\mathbf{x})$ for fixed pairs of windows. First, the energy E is resolved into the element energy $E(\mathbf{x})$ for each pixel in the missing region. As shown in Figure 3, the target pixel to be updated is $\mathbf{x}$, and the pixel position inside a window W can be expressed as $\mathbf{x} + \mathbf{p}$ $(\mathbf{p} \in W)$ and is corresponded to $f(\mathbf{x} + \mathbf{p})$ by Eq. (7). Thus, the position of the pixel corresponding to the pixel $\mathbf{x}$ is $f(\mathbf{x} + \mathbf{p}) - \mathbf{p}$. Now, the element energy $E(\mathbf{x})$ can be defined as follows by the pixel values of $\mathbf{x}$ and $f(\mathbf{x} + \mathbf{p}) - \mathbf{p}$, the coefficient $\alpha_{\mathbf{x}\hat{\mathbf{x}}}$ and the Euclid distance between $\mathbf{x}$ and $f(\mathbf{x})$.

$$E(\mathbf{x}) = \sum_{\mathbf{p} \in W} w_{\mathbf{x}+\mathbf{p}} \{ I(\mathbf{x}) - \alpha_{\mathbf{x}+\mathbf{p} f(\mathbf{x}+\mathbf{p})} I(f(\mathbf{x} + \mathbf{p}) - \mathbf{p}) \}^2 + w_{dis} \parallel \mathbf{x} - f(\mathbf{x}) \parallel. \tag{10}$$

The relationship between the energy for all over the missing region E and the element energy for each pixel $E(\mathbf{x})$ can be written as follows.

$$E = \sum_{x \in \Omega} E(\mathbf{x}) + C. \tag{11}$$

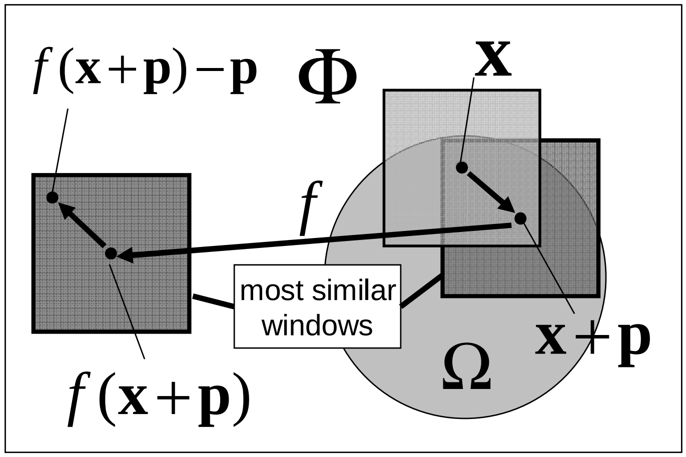


Figure 3. Relationship between pixels in energy calculation.

C is the energy of pixels in the region $\overline{\Omega} \cap \Omega'$, and is treated as a constant because pairs of windows are fixed here. By differentiating E with respect to $I(\mathbf{x})$ in the missing region, the requirement for minimizing the energy E can be obtained as follows.

$$\frac{\partial E}{\partial I(\mathbf{x}_k)} = \sum_{\mathbf{x}\in\Omega} \frac{\partial E(\mathbf{x})}{\partial I(\mathbf{x}_k)} = 0. \tag{12}$$

Here, if it is assumed that the change of intensity modification coefficient $\alpha_{\mathbf{x}\hat{\mathbf{x}}}$ is smaller than that of the pixel intensity $I(\mathbf{x}_k)$, we can obtain the following equation.

$$\frac{\partial \alpha_{\mathbf{x}_i\mathbf{x}'}}{\partial I(\mathbf{x}_j)} = 0 \quad (\forall \mathbf{x}_i, \mathbf{x}_j \in \Omega, \forall \mathbf{x}' \in \Phi). \tag{13}$$

From this equation, the equation $\partial E(\mathbf{x})/\partial I(\mathbf{x}_k) = 0$ $(\mathbf{x} \neq \mathbf{x}_k)$ is formed. Thus, we can minimize the energy E by calculating $I(\mathbf{x}_k)$ that satisfies the following equation.

$$\frac{\partial E}{\partial I(\mathbf{x}_k)} = \frac{\partial E(\mathbf{x}_k)}{\partial I(\mathbf{x}_k)} = 0. \tag{14}$$

By generalizing Eq. (14) to all the pixels in the missing region, each pixel value $I(\mathbf{x})$ in the missing region can be calculated as follows.

$$I(\mathbf{x}) = \frac{\sum_{\mathbf{p}\in W} w_{\mathbf{x}+\mathbf{p}} \alpha_{\mathbf{x}+\mathbf{p} f(\mathbf{x}+\mathbf{p})} I(f(\mathbf{x}+\mathbf{p}) - \mathbf{p})}{\sum_{\mathbf{p}\in W} w_{\mathbf{x}+\mathbf{p}}}. \tag{15}$$

Eq. (15) is an approximate solution because we assume that Eq. (13) is satisfied. However, we can expect to obtain a good solution as the energy gets converged because the value of intensity modification coefficient $\alpha_{\mathbf{x}\hat{\mathbf{x}}}$ gets converged as $I(\mathbf{x})$ gets converged.

2.3.2. Update of pairs of windows

In the process (II), we update all the similar pattern pairs between the window in the missing region and that in the data region by using the pixel values calculated in the process (I). Basically, the update of the pair of windows can be simply performed by calculating SSD' and AD that satisfy Eq. (7). However, computational cost of exhaustively search over whole data region Φ is very expensive and inefficient. In this paper, in order to decrease computational cost, pattern candidates which can be the pattern satisfying Eq. (7) (the most similar pattern) are listed and search is performed effectively by using the candidate list. Figure 4 shows the procedure of searching for the corresponding pixel and making a list for each pixel in the missing region Ω. In the iteration of the process (II), first, every pair of windows is determined so as to satisfy Eq. (7) by computing SSD' and

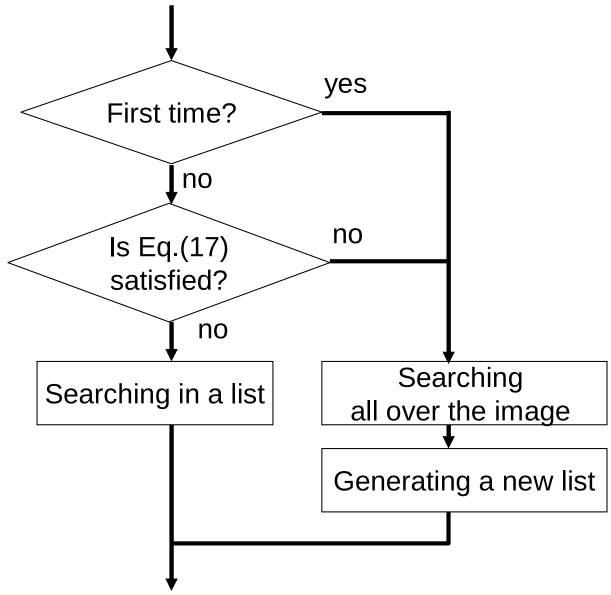


Figure 4. Procedure of making a list for each pixel

AD for all the pixels in the data region. In this exhaustive search, lists of candidates are also constructed. After the first iteration, as long as the condition that will be described later is satisfied, the list is searched for the most similar pattern, otherwise the list is updated by exhaustive search. By using similar pattern candidates, unnecessary search is excluded. The following describes efficient pattern searching in more detail.

First, we describe how to make a list in exhaustive search. Here, we set $S_{min} = SSD'(\mathbf{x}, \hat{\mathbf{x}}) + w_{dis}AD(\mathbf{x}, \hat{\mathbf{x}})$ where the pair $(\mathbf{x}, \hat{\mathbf{x}})$ is determined by Eq. (7). With respect to each $\mathbf{x}$ in the missing region, every $\mathbf{x}'$ in the data region satisfying the following equation is stored to the list as a candidate.

$$SSD'(\mathbf{x}, \mathbf{x}') + w_{dis}AD(\mathbf{x}, \mathbf{x}') < TS_{min}, \tag{16}$$

where T is a constant and we need to set T appropriately considering the tradeoff between the cost for searching a list and the frequency of exhaustive search.

After making the lists, we can update pairs of $(\mathbf{x}, \hat{\mathbf{x}})$ effectively by searching lists for the most similar pattern. However, as the image pattern around $\mathbf{x}$ is updated by the iteration, the most similar pattern may not exist in the list. Thus, it should be judged whether it is guaranteed that the most similar pattern exists in the list, and the list must be reconstructed if it is not guaranteed. Concretely, when the SSD between the pattern around $\mathbf{x}$ in making the list and the pattern in updating pixel values exceeds $\{(\sqrt{T}-1)/2\}^2 S_{min}$, it is not guaranteed that the most similar pattern exists in the list. Therefore, the list for $\mathbf{x}$ is updated if the following equation is not satisfied.

$$\sum_{\mathbf{p} \in W} \{I'(\mathbf{x}+\mathbf{p}) - I(\mathbf{x}+\mathbf{p})\}^2 < \left(\frac{\sqrt{T}-1}{2}\right)^2 S_{min}, \tag{17}$$

where, I' denotes the previous pixel value when the list has been updated by exhaustive search, and I does the current pixel value.

3. EXPERIMENTS

To demonstrate the effectiveness of the proposed method, we applied our method to several images (200 $\times$ 200 pixels) on a standard PC (Pentium-M 2GHz, Memory 512MB). Each image has different characteristics as shown in Figure 5(a). In this experiment, each parameter in the energy function is set as shown in Table 1, and the weight $w_{\mathbf{x}}$ which stands for the distance from the boundary of the missing region is set for each image so that the ratio of maximum to minimum weight is 10. The missing region was manually specified by painting objects in the input image as shown in Figure 5(b). Figure 5(c) illustrates the resultant image completed by the SSD-based conventional approach[6] described in Section 2.1. Figure 5(d) shows the resultant image completed by the proposed method. The difference in implementation between the conventional method and the proposed method is only about energy function, and the effective search described in Section 2.3.2 is implemented for both methods.

As for image A, it is shown that images generated by the conventional and proposed methods are both natural for the image including little change in intensity and pattern around the missing region. The subjective difference is very small. On the other hand, in experiments for images B to D that include large changes in intensity and continuous deformation of structure, the difference between resultant images by the conventional and proposed methods are obvious.

Image B includes large changes in intensity around the missing region. By the conventional method, the resultant image becomes unnatural due to the discontinuous edges between the seat and the seat back. Moreover, the pixel value on the floor changes unnaturally in the conventional method. Our method has successfully generated the image with continuous edge and continuous intensity changes on the floor.

Image C includes continuous deformation of structure that is caused by perspective projection around the missing region. By the conventional method, a part of the missing region is blurred. This is due to weakness of naive SSD function for the deformation of image structure. For the blurred region, the window has corresponded to the wall in the right in the image instead of the window around the missing region. This is because the texture in the missing region often corresponds to low frequency texture in the data region by the SSD-based pattern matching rather than high frequency texture when the deformation of structure exists around the missing region. It should be noted that there exists spatial locality of texture pattern around the missing region in the image C. By considering spatial locality of texture pattern, an appropriate texture pattern is selected for the texture in the missing region, and the texture in the missing region does not blur in the resultant image by the proposed method.

Image D includes both large changes in intensity and continuous deformation of structure around the missing region. In the resultant image of the conventional method, although the structure of road and stonewall is continuous, the unnatural texture is generated due to the discontinuous change in intensity in the missing region. On the other hand, continuous change in intensity and continuous structure is generated in the missing region by the proposed method.

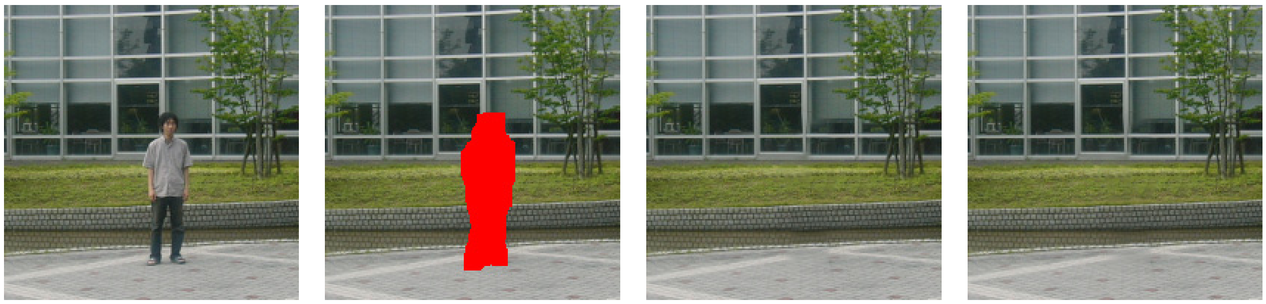
Next, we have compared the conventional and proposed methods with respect to computational cost. Table 2 shows the processing time of the conventional and proposed methods. The proposed method needs about three to seven times as much time as the conventional method. This is because the computational cost for calculating intensity modification coefficients is increased due to consideration of the intensity changes.

Table 1. Parameters in experiment.

Window size: N_w	9×9
Weight for distance: w_{dis}	0.002
Range of intensity modification coefficient α: D	0.1
Upper ratio for candidate for list: T	4

Table 2. Processing time.

	Conventional method	Proposed method
Image A	3'43"	10'38"
Image B	4'35"	12'12"
Image C	1'32"	5'06"
Image D	7'03"	50'40"



(a) (b) (c) (d)

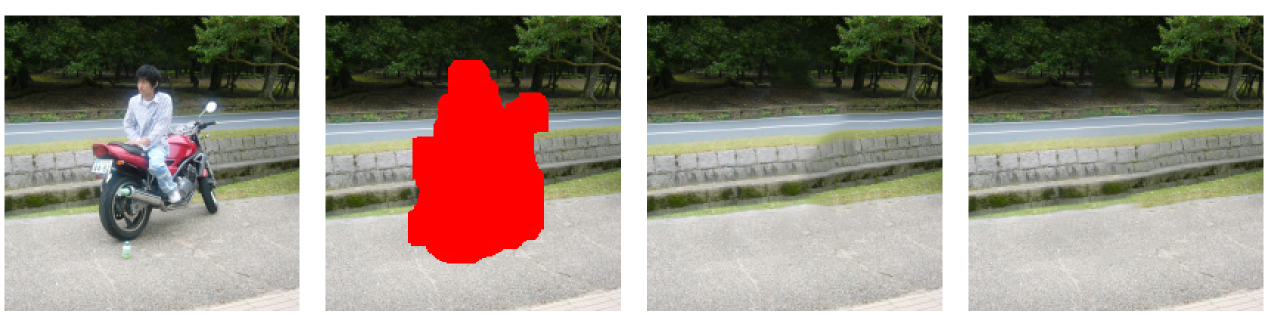
Image A: little change in intensity and pattern

(a) (b) (c) (d)

Image B: large changes in intensity

(a) (b) (c) (d)

Image C: continuous deformation of structure



(a) (b) (c) (d)

Image D: large changes in intensity and continuous deformation of structure

Figure 5. Image inpainting for various images. (a) Original image. (b) Missing region specified manually. (c) Resultant image by conventional method. (d) Resultant image by our method.

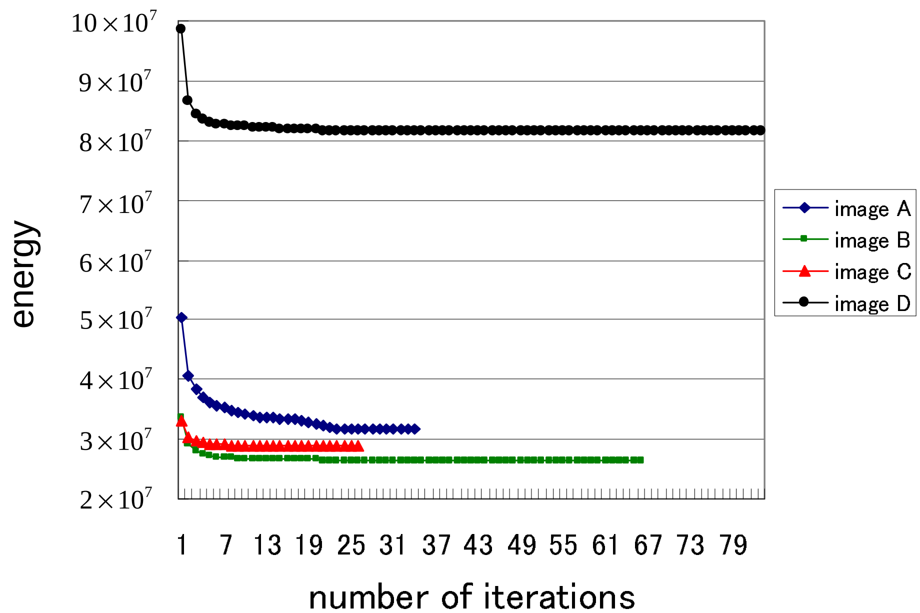


Figure 6. Relationship between energy and number of iteration

Finally, Figure 6 illustrates the change in energy with respect to the number of iterations. From this figure, we can confirm that the energy decreases monotonously although we use the approximated solution in Eq. (13). However, the energy in some images has hardly changed after several tens of iterations. This implies that the processing time can be shorten by easing the convergence condition.

4. CONCLUSION

In this paper, we have proposed a new method of image inpainting which completes missing image regions naturally by minimizing a newly defined energy function that considers intensity changes and spatial locality. In experiments, we have confirmed that the proposed method can prevent the unnatural intensity changes and blurs in the resultant image. However, parameters such as the size of window and the weight in the energy function were decided empirically. In future work, we should establish a method to decide optimum parameters. Furthermore, we should reduce the computational cost.

REFERENCES

1. S. Masnou and J.M. Morel: “Level Lines Based Dissocclusion”. Proc. IEEE Int. Conf. on Image Processing, 1998.
2. M. Bertalmio, G. Sapiro, V. Caselles and C.Ballester: “Image Inpainting”. Proc. ACM SIGGRAPH2000, pp.417-424, 2000.
3. T. Amano: “Image Interpolation by High Dimensional Projection based on Subspace Method”. Proc. IEEE Int. Conf. on Pattern and Recognition, pp.665-668, 2004.
4. 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, 2004.
5. J. Sun, L. Yuan, J. Jia and H. Shum: “Image Completion with Structure Propagation”. Proc. ACM SIGGRAPH2005, pp.861-868, 2005.
6. Y. Wexeler, E. Shechtman and M. Irani: “Space-Time Video Completion”. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.120-127, 2004.