Example-based Super Resolution to Achieve Fine Magnification of Low-Resolution Images

SENDA Shuji, SHIBATA Takashi, IKETANI Akihiko

Abstract

"Super Resolution" (SR) is a technique to restore a low-resolution image into a clear, high resolution image. It is necessary to infer missing high frequency components in order to restore a low-resolution image to a high-resolution image correctly. This paper describes the learning-based SR technique that utilizes an example-based algorithm. This technique divides a large volume of training images into small rectangular pieces called "patches" and brings them together in a dictionary as patch pairs of low-resolution and high-resolution images. Experiments show impressive results that identify specific objects such as text characters and human faces etc.

Keywords

super resolution (SR), training, example-based, patch, image restoration

1. Introduction

"Super Resolution" (SR) is a technique for restoring a low resolution image (LR image) into a clear high resolution image (HR image). This technique is widely adopted to improve the definition of images shot by a digital camera. In order to restore an image into a high-resolution image correctly, it is necessary to infer missing high frequency components of a low-resolution image. The technique to achieve this is called "learning-based super resolution," which prepares data relevant to the target object in advance.

The learning-based method can be classified into two methods; 1) a regression-based method that requires model function parameters, and 2) an example-based method that uses samples as models. The regression-based method features a superior generalization capability to handle various objects, however, at the same time it features an inferior capability when handling complex models. This means that the regression-based method has a limited magnification ratio for achieving super resolution images (scaling up ratio between a low-resolution image and a high-resolution image). On the other hand, the example-based method is able to achieve a super resolution image with a significantly high magnification ratio, although this will depend on the specific object being processed. It divides a large amount of training images into small patches and uses them for in the analytical process.

This paper proposes a learning-based super resolution technique using an example-based approach to enable restoration of finely-magnified, high-resolution images of specific objects including texts and human faces. This technique features a fine balance between efficiency and performance and thereby provides significant results especially for specifically processed objects. It deals with object images that feature minute dislocation and blur by increasing the number of training images, and with object images that have minor distortions by applying searching and synthesizing processes with weighted Euclidean distance. The proposed technique achieves finely-magnified, high-precision image restoration, so that low-resolution images of the license plates of vehicles recorded by surveillance cameras become legible. This was impossible when using the conventional super-resolution processing technique. According to the recent increase in the numbers of installed surveillance cameras, requests have been raised to analyze the vast amount of recorded data as a Big Data. The super resolution that we propose in this paper is a processing technique to convert low-resolution images that are recorded during wide area surveillance into high-resolution images, so that a machine or a human may recognize and analyze the results. I would like also to state that this technique is adequately meeting popular expectations.

2. Proposed Method

The proposed method consists of two phases, 1) a dictionary construction phase that performs extraction of patch pairs from both HR and LR images, and then stores them as training data in the dictionary, and 2) an super resolution phase that performs synthesizing of HR images by searching patches stored in the dictionary that are best matched to the input images.

2.1 Dictionary Construction Phase

In the dictionary construction phase, different types of LR images classified by minute variations are generated out of a large volume of HR training images. Patches are extracted from HR and LR images respectively, and the pairs are stored in the dictionary (**Fig. 1**).

In order to produce LR images from the corresponding HR images, a certain amount of blurring is provided to an HR image, and sampling of pixels is then performed at certain intervals. Here we describe an example process to perform super resolution with 3 times magnification. This example aims to produce an LR image with a scaling ratio of 1/3. With this aim, blurring by using a Gaussian filter with a standard deviation of 1.5 pixels is provided initially and then sampling of pixels at 3 pixel intervals is performed. Three types of deviation of 1.5, 1.875 and 2.25 pixels are provided, so that the blur amount can be finely varied according to the settings of the digital cameras. This process enables the handling of various levels of input images. Moreover, minute displacements may occur depending on the extraction position to be set when extracting pixels at intervals of 3 pixels. In order to solve this issue, 4 types of images are extracted by changing the positions to 2 vertical and 2 horizontal. When 3 types of blur and 4 types of displacement are provided, a total of 12 LR images will be produced to combine such fine variations from a single HR image.

Patch pairs are composed of LR patches (e.g. 5 × 5 pixels) extracted from LR images and the corresponding HR patches (e.g. 15 × 15 pixels in case of 3x magnification). Those pairs are stored in the dictionary. If trying to store all sort of patches, it would exceed over million pairs. Deleting similar patches enables to reduce the size of the dictionary.

2.2 Super Resolution Phase

In the super resolution phase, an LR patch is extracted from the input images. The most alike patch pair is searched in the dictionary to synthesize an HR image using the searched HR patch in the pair (**Fig. 2**).

In searching the LR patch process, by defining the LR and HR patches as the gray image vectors, the contrast of the LR and HR patches is normalized by dividing them by those norms, so that any degradation due to brightness may be avoided. Also, in order to achieve a robust comparison of minute variations, the distance between patches is calculated by weighted Euclidean distance, so the further the distance from the patch center point becomes, the smaller the weighting that is applied. The LR patches that are closest to each other in terms of distance should be recognized as similar patches. An HR image is synthesized by using the HR patch of the searched patch pair. This process also employs the same process for searching LR patches: the further the distance from the patch center point becomes, the smaller the weighting that is applied. At the same time, the contrast is corrected according to the average

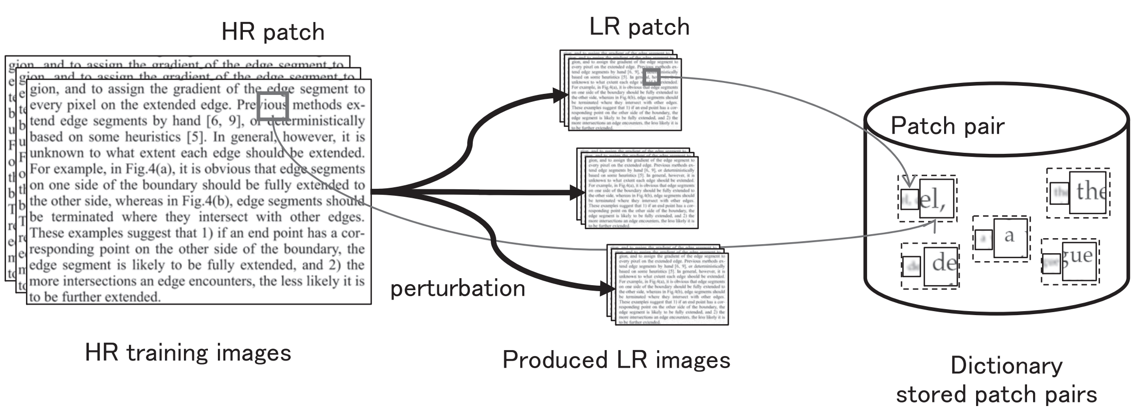

Fig. 1 Dictionary construction phase.

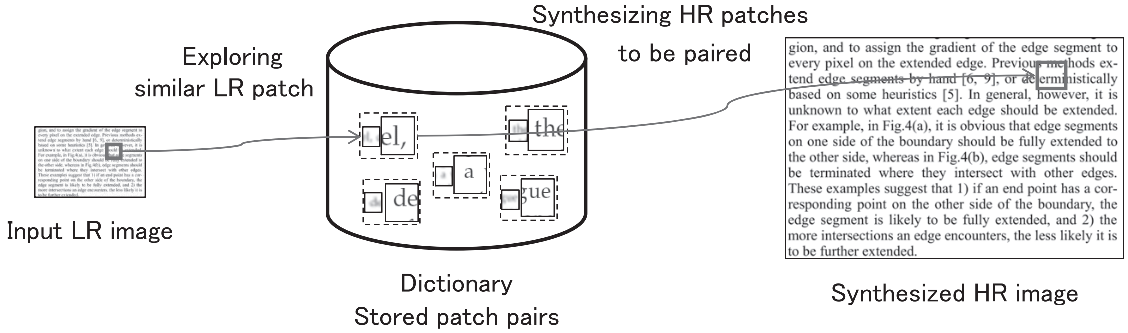

Fig. 2 Super-resolution phase.

value of the LR patches in input images. These processes such as searching of LR patches and synthesizing of HR patches are independently performed per pixel. Acquired results are also synthesized by applying weighting, so that high resolution images or "super resolution" results may be obtained.

The dictionary contains a vast number of patches so that it takes a long time to search the desired patches. Consequently LR patches that are normalized by norms and receive weights in advance are stored in the tree structure called "k-means tree." This is the process by which we carry out rapid exploration of the nearest neighbor patches.

3. Experimental Results

In order to show the effectiveness of our proposed method, we introduce the results when our proposed method was applied to a low-resolution text image shot by a digital camera, a car license plate image and a human face image.

3.1 Text Image

50 pages of English texts for which the formats were roughly identical were used for the training HR images. They were saved as PDF files and were then converted into 150-dpi graphic images. LR images used for the experiment were produced by shooting the printed English texts with 50-dpi equivalent with a Web camera from a frontal angle. The printed pages were the different pages that were used for the training HR images. The result of the 3× magnified super resolution being applied is shown in **Fig. 3** . It is not known how much noise or blurring caused by the camera is contained in the image, even though the image was restored precisely. As shown in the results below, our proposed method proved to have significant effectiveness in restoring images when the target object contains strong limitations such as in fonts and types of texts, etc.

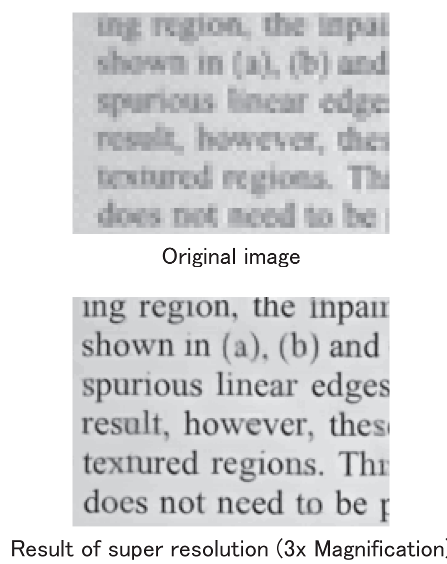

Fig. 3 Result when applying to a text image.

3.2 Car License Plate Image

The result when the proposed method was applied to a car license plate image is shown below. We constructed a dictionary with 2,000 car license plate images (of the same sizes) in order to prepare training HR images. The LR images used for the experiment were shot by a digital camera from a long distance. The result of 8× super resolution result acquired by the proposed method is shown in **Fig. 4** . Although the original image was only 13 by 7 pixels the four digit numbers in the image were correctly restored by making the best use of the car license plate dictionary.

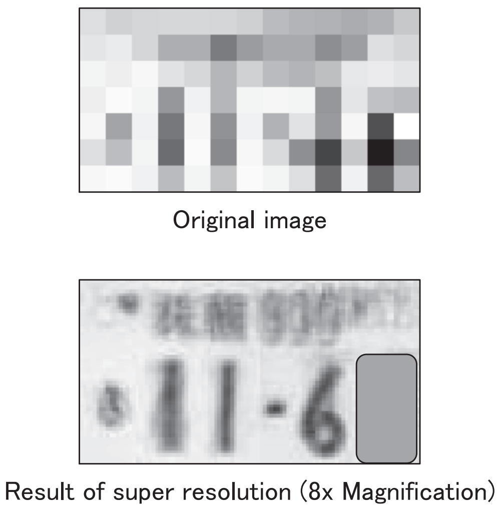

Fig. 4 Result when applied to a car license plate.

3.3 Human Face Image

The result when the proposed method was applied to the image of a human face is shown below. We constructed a dictionary with 450 human face images (of the same size) to prepare training HR images. The LR images used for the experiment were shot by a digital camera from a long distance. The 4× super resolution result acquired by the proposed method is shown in **Fig. 5** . Even though the training patches were produced from other people faces, the results show excellent restoration of the eyes and mouth.

Fig. 5 Result when applied to a human face.

4. Conclusion

In this paper we propose a learning-based super resolution technique using an example-based approach that enables restoration of finely-magnified, high-resolution images by specifying target objects. The proposed method features the unique properties of minutely varied, training images as well as applying weighted searching/synthesizing. Therefore, it can achieve finely-magnified, high-precision super resolution of images efficiently, including those of texts and human faces. The experimental results of text images, car license plate images and human face images demonstrate the great effectiveness of the proposed method.

With regard to issue solving in the future, we will study how to deal with images without prepared training images, and we will also try to improve our processing efficiency even further.

Reference

1) C. Liu, H. Y. Shum, and W. T. Freeman. “Face hallucination: theory and practice.” IJCV, Vol. 75, No. 1, pp. 115-134 2007.

Authors' Profiles

SENDA Shuji
Manager
C&C Innovation Initiative
Central Research Laboratories

SHIBATA Takashi
Information and Media Processing Laboratories
Central Research Laboratories

IKETANI Akihiko
Assistant Manager
Information and Media Processing Laboratories
Central Research Laboratories

The details about this paper can be seen at the following.

Related URL:

http://www.nec.co.jp/press/en/1112/1401.html

Information about the NEC Technical Journal

Thank you for reading the paper.
If you are interested in the NEC Technical Journal, you can also read other papers on our website.

Link to NEC Technical Journal website

Japanese

English

Vol.7 No.2 Big Data

Vol.7 No.2

September, 2012

Special Issue TOP