

# SUPER-RESOLUTION BY GMM BASED CONVERSION USING SELF-REDUCTION IMAGE

Yuki Ogawa, Yasuo Ariki and Tetsuya Takiguchi

Graduate School of System Informatics, Kobe University  
Rokkodaicho 1-1, Nada-ku, Kobe, HyogoPref., Japan 657-8501

## ABSTRACT

In recent years, super-resolution techniques in the field of computer vision have been studied actively owing to the potential applicability in various fields. In this paper, we propose a single-image, super-resolution approach using GMM (Gaussian Mixture Model)-based conversion. The conversion function is constructed by GMM using the input image and its self-reduction image. The high-resolution image is obtained by applying the conversion function to the enlarged input image without any outside database. We confirmed the effectiveness of this proposed method through the experiments.

*Index Terms*— super-resolution, GMM

## 1. INTRODUCTION

The resolution of the digital camera installed in cellular phones has increased dramatically in recent years. On the other hand, due to price competition, the need to reduce the cost of the image sensor has been a serious problem. For this reason the technology for high-resolution digital image processing has been attracting much attention. If images are enlarged by using either linear interpolation or bicubic interpolation (a popular expansion processing technique), the resolution of the images decreases because their edge information is lost. Therefore, a method that assures the high resolution of the expanded images and adds an appropriate high-frequency component to the image is required.

Single-image super-resolution techniques in the field of computer vision have been studied actively. The typical method for single-image super-resolution is an example-based one [1]. Association between low- and high-resolution image patches is learned from a database with low- and high-resolution image pairs, and then it is applied to a new low-resolution image to restore the most likely high-resolution component. Some studies [2, 3] also propose methods of employing a conversion function so that low-resolution images can be converted into high-resolution ones.

Super-resolution techniques restore the high-frequency component of the original data that is lost for various reasons from the observed data. In this paper, we propose a method to restore the high frequency component by constructing a conversion function that converts the low-resolution image

feature preserved in the enlarged blur image to the lost higher-frequency component. This method was originally developed in GMM (Gaussian Mixture Model)-based voice conversion [4]. The voice conversion is a method that converts a speaker's voice into another speaker's voice. We apply this voice conversion to super-resolution so that low-resolution images can be converted into high-resolution ones. The conversion function is constructed between the original image and its self-reduction image using GMM. Then the conversion function is applied to the enlarged image to restore the higher frequency component.

This paper is structured as follows. Section 2 describes GMM based conversion. Section 3 describes the proposed super-resolution system using the conversion function. Section 4 describes our experimental results, and Section 5 summarizes the paper.

## 2. GMM-BASED CONVERSION FUNCTION

Let the source image and the target image be expressed as  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ ,  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$  respectively. The probability distribution of the source image  $x$  is expressed by mixture distribution as follows:

$$P(x) = \sum_{i=1}^m \alpha_i N(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad (1)$$

$$\sum_{i=1}^m \alpha_i = 1, \alpha \geq 0 \quad (2)$$

where  $\alpha_i$  is the weight of mixture  $i$ , and  $m$  is the number of mixtures.  $N(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  is the normal distribution with mean vector  $\boldsymbol{\mu}_i$  and variance-covariance matrix  $\boldsymbol{\Sigma}_i$  and expressed as follows:

$$N(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right] \quad (3)$$

The conversion function from the source image into the target one is expressed as follows [4].

$$\begin{aligned}
y &= F(x) = E[y|x] \\
&= \sum_{i=1}^m h_i(\mathbf{x}) [\boldsymbol{\mu}_i^y + \boldsymbol{\Sigma}_i^{yx} (\boldsymbol{\Sigma}_i^{xx})^{-1} (\mathbf{x} - \boldsymbol{\mu}_i^x)] \quad (4) \\
h_i(\mathbf{x}) &= \frac{\alpha_i N(\mathbf{x}; \boldsymbol{\mu}_i^x, \boldsymbol{\Sigma}_i^{xx})}{\sum_{j=1}^m \alpha_j N(\mathbf{x}; \boldsymbol{\mu}_j^x, \boldsymbol{\Sigma}_j^{xx})} \quad (5)
\end{aligned}$$

where  $\boldsymbol{\mu}_i^x$  and  $\boldsymbol{\mu}_i^y$  are the mean vectors at mixture  $i$  of the source image and the target image, respectively.  $\boldsymbol{\Sigma}_i^{xx}$  and  $\boldsymbol{\Sigma}_i^{yx}$  are the covariance matrix and the cross-covariance matrix at mixture  $i$  of the source image and target image, respectively. The source image is converted into the target image using Eq. (4). The parameter  $(\alpha_i, \boldsymbol{\mu}_i^x, \boldsymbol{\mu}_i^y, \boldsymbol{\Sigma}_i^{xx}, \boldsymbol{\Sigma}_i^{yx})$  of the conversion function is trained using the joint probability distribution of vector  $\mathbf{z} = [\mathbf{x}^T \ \mathbf{y}^T]^T$  expressed by GMM as follows:

$$P(\mathbf{z}) = \sum_{i=1}^m \alpha_i N(\mathbf{z}; \boldsymbol{\mu}_i^z, \boldsymbol{\Sigma}_i^z) \quad (6)$$

where  $\boldsymbol{\mu}_i^z$  and  $\boldsymbol{\Sigma}_i^z$  are expressed as follows:

$$\boldsymbol{\Sigma}_i^z = \begin{bmatrix} \boldsymbol{\Sigma}_i^{xx} & \boldsymbol{\Sigma}_i^{xy} \\ \boldsymbol{\Sigma}_i^{yx} & \boldsymbol{\Sigma}_i^{yy} \end{bmatrix}, \boldsymbol{\mu}_i^z = \begin{bmatrix} \boldsymbol{\mu}_i^x \\ \boldsymbol{\mu}_i^y \end{bmatrix} \quad (7)$$

These parameters are trained by EM algorithm.

### 3. SUPER-RESOLUTION SYSTEM

Fig. 1 shows the proposed super-resolution system. The conversion function is constructed using the input image and its self-reduction image by GMM in the learning phase. The high-resolution image is restored by using the conversion function in the estimation phase.

#### 3.1. Learning of Conversion Function

1. Low- and high-resolution image pair ( $I$  and  $I_L$ ) is prepared by reducing the input image  $I$  with high resolution to the self-reduction image  $I_R$ , and then enlarging the image  $I_R$  to the image  $I_L$  by bicubic interpolation.
2. The high-frequency components ( $I_{LH1}, I_{LH2}, I_{LH3}$  and  $I_{LH4}$ ) are extracted from the image  $I_L$  as a low-resolution image feature by applying various high-pass filters ( $H1, H2, H3$  and  $H4$ ). The high-pass filters are 6-dimensional first-order derivatives in the horizontal and vertical directions and 7-dimensional second-order derivatives in the horizontal and vertical direction, as

shown below.

$$\begin{aligned}
\mathbf{H1} &= [0, 0, 1, 0, 0, -1] \\
\mathbf{H2} &= [0, 0, 1, 0, 0, -1]^T \\
\mathbf{H3} &= [1/2, 0, 0, -1, 0, 0, 1/2] \\
\mathbf{H4} &= [1/2, 0, 0, -1, 0, 0, 1/2]^T
\end{aligned}$$

3. The difference image  $I_F$  is produced as a high-resolution image feature by subtracting the enlarged image  $I_L$  with low resolution from the input image  $I$  with high resolution.
4. Association between the low-resolution image feature  $I_{LH1}, I_{LH2}, I_{LH3}, I_{LH4}$  and the high-resolution image feature  $I_F$  is decomposed into the small image patches. Therefore, a set of associated image patches is produced as follows:

$$\mathbf{x}_m = [I_{LH1m}^T, I_{LH2m}^T, I_{LH3m}^T, I_{LH4m}^T]^T \quad (8)$$

$$\mathbf{y}_m = [I_{Fm}]^T \quad (9)$$

where  $m$  is the number of patches. The joint vector  $\mathbf{z}$  is obtained by concatenating  $\mathbf{x}$  and  $\mathbf{y}$ . Based on the GMM of the joint vector, the conversion function is constructed according to Eq. (4) in the learning phase.

#### 3.2. Estimation of Super-resolution

1. The input image  $I$  is enlarged to the image  $I'_L$  using bicubic interpolation. The enlarged input image  $I'_L$  is high-pass filtered and decomposed into patches as in the learning phase. At this stage, a set of image patches with the low-resolution image feature  $\mathbf{x}_m$  is obtained. The lost high-resolution image feature  $I'_F$  is restored by applying the conversion function Eq. (4) to the image patches  $\mathbf{x}_m$  in the estimation phase.
2. Finally, the super-resolution image  $I_S$  is obtained by adding the lost high-resolution image feature  $I'_F$  to the enlarged input image  $I'_L$ .

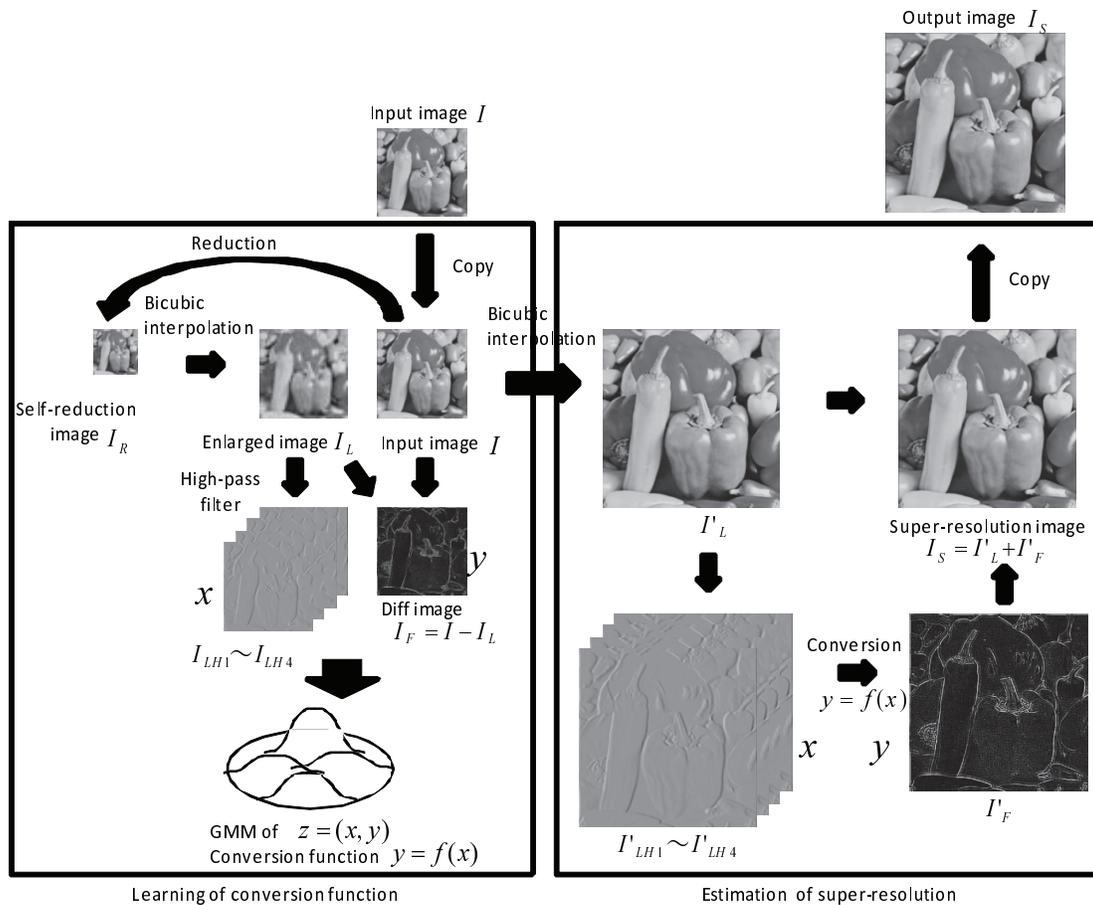
## 4. EXPERIMENTAL RESULTS

#### 4.1. The Input Image

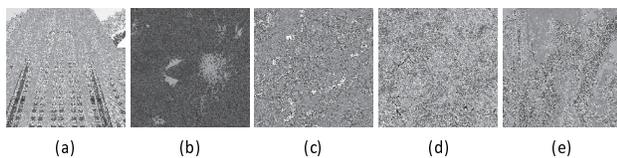
Five input images (shown in Fig. 2) were prepared for testing.

#### 4.2. Evaluation Method

In order to evaluate impartially, various evaluation techniques were employed in the super-resolution restoration experiment, including PSNR, SSIM [5] and VSNR [6]. Given an original image and its processed image, PSNR, SSIM and



**Fig. 1.** Super-resolution system



**Fig. 2.** Input images

VSNR measure the quality of the processed image. The larger the values of PSNR, SSIM and VSNR are, the higher the quality of the image will be.

### 4.3. Assessment of Experiment

The proposed method and bicubic interpolation were compared. The original images (576x576) were reduced by one-third in the horizontal and vertical directions and they were used as input images  $I$ . The input images were enlarged by three times in the horizontal and vertical directions by the proposed method and bicubic interpolation. The size of the im-

age patches was 3x3 as the best value in the experiment. The number of mixtures in GMM was set to 5 as the best value in the experiment. They were compared with the original image in terms of PSNR, SSIM and VSNR.

Table 1 shows the evaluation results by PSNR, SSIM and VSNR. Fig. 3 and Fig. 4 show part of the resultant images. Image (a) is the result enlarged by bicubic interpolation and image (b) is the result enlarged by the proposed method.

Comparing the “proposed” method with the “bicubic” method in Table 1, it can be seen that “proposed” obtained good evaluation values. In addition, the differences in clarity between the “(b) proposed method” images and bicubic method are noticeable in Fig. 3 and Fig. 4.

## 5. CONCLUSION

In this paper, a super-resolution technique was proposed by employing GMM-based conversion using the self-reduction image and the input image. The effectiveness of the proposed method was confirmed through the experiments in terms of three types of evaluation measures. The future work will in-

**Table 1.** Comparison of the super-resolution images shown in Fig. 2 by PSNR, SSIM and VSNR

Image	Measure	PSNR	SSIM	VSNR
(a)	Bicubic	32.08	0.7256	13.39
	Proposed	32.35	0.7584	13.61
(b)	Bicubic	37.62	0.8763	16.92
	Proposed	38.16	0.8907	17.15
(c)	Bicubic	34.40	0.8863	14.92
	Proposed	35.76	0.9163	15.27
(d)	Bicubic	32.55	0.7530	15.56
	Proposed	33.31	0.7961	16.00
(e)	Bicubic	36.08	0.8775	14.25
	Proposed	36.79	0.8941	14.50

clude expanding the method to improve the conversion function by employing canonical correlation analysis.

**Acknowledgment:** This research was supported in part by MIC SCOPE.

## 6. REFERENCES

- [1] Freeman W.T., Jones T.R., and Pasztor E.C., "Example-based super-resolution," *IEEE Computer Graphics and Applications*, pp. 56–65, 2002.
- [2] Michael Elad, Mario A. T. Figueiredo, and Yi Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, June 2010.
- [3] Takashiro Ogawa and Miki Haseyama, "Adaptive reconstruction method of missing textures based on perceptually optimized algorithm," *ICASSP*, pp. 1157–1160, 2011.
- [4] Yannis Stylianou, Olivier Cappe, and Eric Moulines, "Statistical methods for voice quality transformation," *EUROSPEECH*, pp. 447–450, September 1995.
- [5] Zhou Wang, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 13, no. 4, pp. 600–612, April 2004.
- [6] Damon M. Chandler and Sheila S. Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 16, no. 9, pp. 2284–2298, September 2007.



(a)Bicubic interpolation



(b)Proposed method

**Fig. 3.** Part of the enlarged image(d)



(a)Bicubic interpolation



(b)Proposed method

**Fig. 4.** Part of the enlarged image(e)