Super-Resolution Using GMM and PLS Regression

Yuki Ogawa, Takahiro Hori, Tetsuya Takiguchi, and Yasuo Ariki
Graduate School of System Informatics, Kobe University
1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan
{*ogawa,horitaka*}*@me.cs.scitec.kobe-u.ac.jp,* {*takigu,ariki*}*@kobe-u.ac.jp*

***Abstract*—In recent years, super-resolution techniques in the field of computer vision have been studied in earnest owing to the potential applicability of such technology in a variety of fields. In this paper, we propose a single-image, super-resolution approach using a Gaussian Mixture Model (GMM) and Partial Least Squares (PLS) regression. A GMM-based super-resolution technique is shown to be more efficient than previously known techniques, such as sparse-coding-based techniques. But the GMM-based conversion may result in overfitting. In this paper, an effective technique for preventing overfitting, which combines PLS regression with a GMM, is proposed. The conversion function is constructed 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 our experiments.**

***Keywords*-super-resolution; GMM; PLS regression;**

I. 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.

A considerable amount of research has been carried out on single-image super-resolution techniques in the field of computer vision. 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 it is then 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 for restoring 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]. This voice conversion is a method that converts a speaker's voice into another speaker's voice. We applied this voice conversion to super-resolution so that low-resolution images can be converted into high-resolution ones [5]. The conversion function is constructed between the original image and its self-reduction image using a GMM. Then the conversion function is applied to the enlarged image to restore the higher frequency component.

The GMM-based super-resolution technique may cause overfitting to occur when a model has too many degrees of freedom compared to the amount of training data available. Overfitting results in poor predicting ability regarding new data while giving very good results for the training data. To prevent overfitting, a technique to combine partial least squares (PLS) regression with a GMM has been proposed for voice conversion [6]. In this paper, we propose a super-resolution technique using a GMM and PLS regression.

This paper is structured as follows. Section II describes GMM-based conversion. Section III describes our super-resolution system. Section IV describes PLS-regression-based conversion. Section V describes our experimental results, and Section VI summarizes the paper.

II. GMM-based conversion function

Let the source image and the target image be expressed as $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$, $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]^T$, respectively. The joint probability distribution of vector $\boldsymbol{z} = [\ \boldsymbol{x}^T\ \boldsymbol{y}^T\ ]^T$ expressed by a GMM as follows:

$$P(z) = \sum_{m=1}^{M} \alpha_m N(\boldsymbol{z}; \boldsymbol{\mu}_m^z, \boldsymbol{\Sigma}_m^z) \tag{1}$$

where α_m is the weight of the m-th mixture, and M is the number of mixtures. $N(z; \mu_m^z, \Sigma_m^z)$ represents the normal distribution with mean vector μ_m^z and variance-covariance matrix Σ_m^z. $\boldsymbol{\mu}_m^z$ and $\boldsymbol{\Sigma}_m^z$ are expressed as follows:

$$\boldsymbol{\Sigma}_m^z = \begin{bmatrix} \boldsymbol{\Sigma}_m^{xx} & \boldsymbol{\Sigma}_m^{xy} \\ \boldsymbol{\Sigma}_m^{yx} & \boldsymbol{\Sigma}_m^{yy} \end{bmatrix}, \quad \boldsymbol{\mu}_m^z = \begin{bmatrix} \boldsymbol{\mu}_m^x \\ \boldsymbol{\mu}_m^y \end{bmatrix} \tag{2}$$

These parameters are trained using the EM algorithm.

978-0-7695-4875-3/12 $26.00 © 2012 IEEE
DOI 10.1109/ISM.2012.62

Next, the conversion function from the source image into the target one is expressed on the basis of the minimum mean-square error as follows [4]:

$$\begin{aligned} y &= F(x) = E[y|x] \\ &= \sum_{m=1}^{M} w_m(\boldsymbol{x})[\boldsymbol{\mu}_m^y + \boldsymbol{\Sigma}_m^{yx}(\boldsymbol{\Sigma}_m^{xx})^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_m^x)] \quad (3) \end{aligned}$$

$$w_m(\boldsymbol{x}) = \frac{\alpha_m N(\boldsymbol{x}; \boldsymbol{\mu}_m^x, \boldsymbol{\Sigma}_m^{xx})}{\sum_{j=1}^{M} \alpha_j N(\boldsymbol{x}; \boldsymbol{\mu}_j^x, \boldsymbol{\Sigma}_j^{xx})} \quad (4)$$

where $\boldsymbol{\mu}_m^x$ and $\boldsymbol{\mu}_m^y$ represent the mean vectors at the m-th mixture of the source image and the target image, respectively. $\boldsymbol{\Sigma}_m^{xx}$ and $\boldsymbol{\Sigma}_m^{yx}$ represent the covariance matrix and the cross-covariance matrix at the m-th mixture of the source image and target image, respectively. The source image is converted into the target image using Eq. (3).

III. Super-Resolution System

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

A. Learning of the Conversion Function

1) A low- and high-resolution image pair (I_L and I) is prepared by reducing the high-resolution input image I to the self-reduction image I_R, and then enlarging image I_R to image I_L using bicubic interpolation.
2) The high-frequency components ($I_{LH1}, I_{LH2}, I_{LH3}$ and I_{LH4}) are extracted from 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.

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

3) The difference image I_F is produced as a high-resolution image feature by subtracting the enlarged image I_L (low resolution) from the input image I (high resolution).
4) The low-resolution image feature (I_{LH1}, I_{LH2}, I_{LH3}, I_{LH4}) and the high-resolution image feature I_F are decomposed into small image patches, respectively. Therefore, the k-th associated image patch is produced as follows:

$$\boldsymbol{x}_k = [\boldsymbol{I}_{LH1k}^T, \boldsymbol{I}_{LH2k}^T, \boldsymbol{I}_{LH3k}^T, \boldsymbol{I}_{LH4k}^T]^T \quad (5)$$

$$\boldsymbol{y}_k = [\boldsymbol{I}_{Fk}]^T \quad (6)$$

where k represents the patch index. The joint vector $\boldsymbol{z}$ is obtained by concatenating $\boldsymbol{x}$ and $\boldsymbol{y}$. Based on the GMM of the joint vector, the conversion function is constructed according to Eq. (3) in the learning phase.

B. Estimation of Super-Resolution

1) The input image I is enlarged to the image I'_L using bicubic interpolation. High-pass filters are applied to the enlarged input imageI'_L. The obtained images are decomposed into patches, and a set of image patches with the low-resolution image feature, $\boldsymbol{x}_k$, is obtained. The lost high-resolution image feature I'_F is restored by applying the conversion function Eq. (3) to the image patches $\boldsymbol{x}_k$ 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.

IV. PLS-Regression-Based Conversion Function

A well-known drawback of GMM-based conversion is overfitting. It may occur when a model has too many degrees of freedom compared to the amount of training data available. To prevent overfitting in voice conversion, a technique that combines partial least squares (PLS) regression with a GMM has been proposed [6].

In this paper, we propose a super-resolution technique using a GMM and PLS regression. PLS regression generates an observed variable using a small number of latent variables, which explains most of the variation in the target. The regression form is given by

$$y_k = \beta x_k + e_k \quad (7)$$

where β and e represent the regression matrix and the regression residual, respectively. Many variants exist for solving the PLS regression problem. In this paper, the SIMPLS (simple partial least squares) algorithm [7] is used for obtaining the regression matrix β.

Similar to voice conversion using PLS regression [6], in this paper, PLS regression is extended for a GMM using the posterior probabilities in Eq. (4), and the k-th target image patch $\hat{y}_k$ is given as follows:

$$\hat{y}_k = \sum_{m=1}^{M} w_m(x_k) \cdot \beta_m \cdot x_k + e_k \quad (8)$$

Learning of the conversion function and estimation of super-resolution are performed as described in Section III-A and Section III-B, respectively, using the conversion function of Eq. (8).

V. Experiments

A. Quantitative Measurements

In order to evaluate impartially, various evaluation techniques were employed in the super-resolution restoration experiment, including PSNR (Peak Signal to Noise Ratio), SSIM (Structural SIMilarity) [8] and VSNR (Visual SNR)

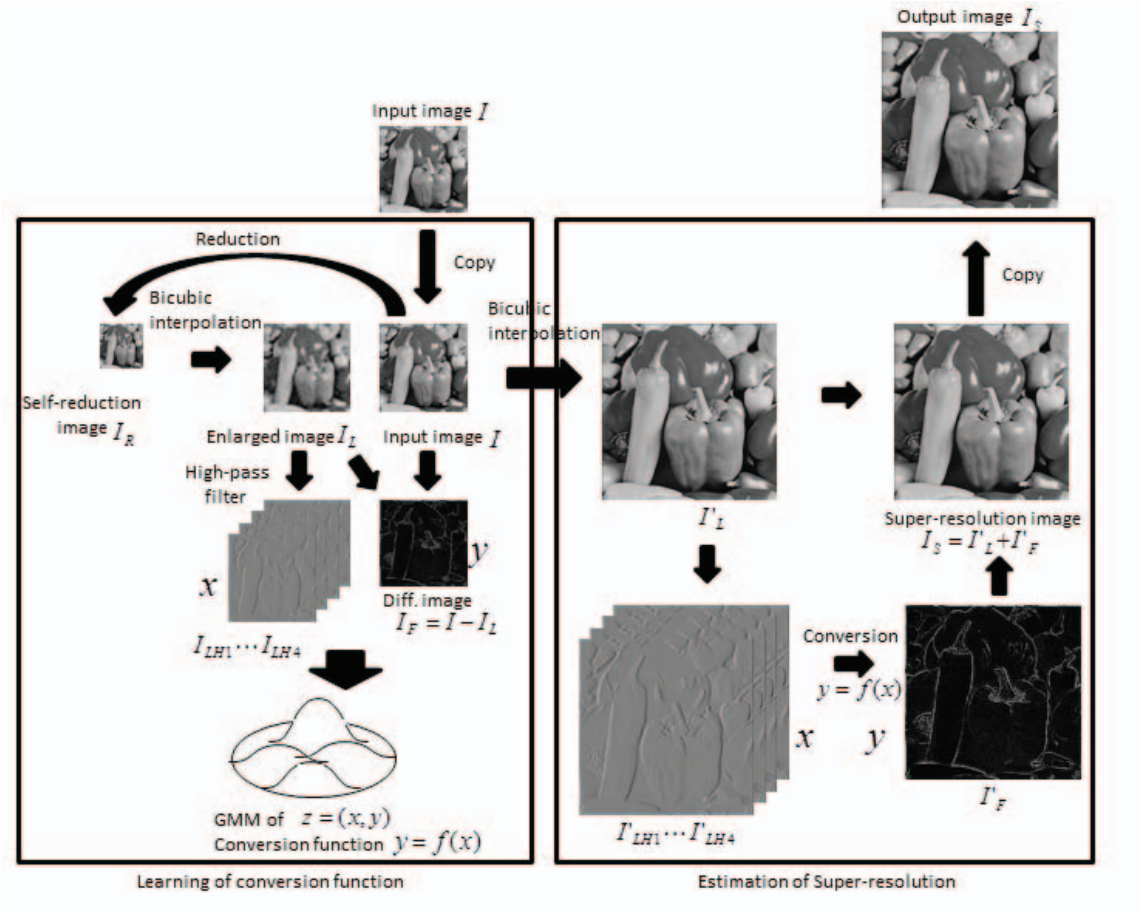

Figure 1. Super-resolution system

[9]. Given an original image and its processed image, PSNR, SSIM and 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.

1) PSNR: PSNR is obtained as follows:

$$PSNR = 10 \log_{10} \left(255^2 / MSE\right) \quad (9)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I_{i,j} - J_{i,j})^2 \quad (10)$$

where $I_{i,j}$ and $J_{i,j}$ are the original image and the processed image, respectively, and their size is $m \times n$.

2) SSIM: The image similarity is obtained using SSIM as follows:

$$SSIM(I, J) = \frac{(2\mu_I \mu_J)(2\sigma_{IJ})}{(\mu_I^2 + \mu_J^2)(\sigma_I^2 + \sigma_J^2)} \quad (11)$$

where μ_I and μ_J are the averages over the images I and J, respectively, σ_I and σ_J are the variances of I and J, respectively, and σ_{IJ} is the covariance of I and J.

3) VSNR: VSNR is given as follows:

$$VSNR = 10 \log_{10} \left(\frac{C(I)}{VD} \right)^2 \quad (12)$$

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

Figure 2. Images used in our experiments

where $C(I)$ denotes the contrast of the original image I, and VD denotes the visual distortion which is described for details in [9].

B. Experiment Results

Four images (shown in Fig. 2) were used in our experiments. The original images (512 $\times$ 512) were reduced by half in the horizontal and vertical directions, and they were used as input images I. The input images were enlarged by two times in the horizontal and vertical directions using the proposed method.

Table I
COMPARISON OF DIFFERENCE IN IMAGE QUALITY IN RELATION TO PATCH SIZE

Patch size	PSNR	SSIM	VSNR
3×3	**38.92**	**0.9308**	**13.92**
6×6	38.62	0.9293	**13.92**
12×12	38.55	0.9246	**13.92**
18×18	38.13	0.9152	**13.92**
24×24	37.53	0.9006	**13.92**
30×30	36.52	0.8711	13.90

Table I shows the comparison of the difference in image quality in relation to patch size using GMM-based super resolution, where the image (b) was used, and the number of mixtures in Eq. (1) was 10. As shown in this table, the 3×3 image patch had the best value in the experiment.

Table II
COMPARISON OF DIFFERENCE IN IMAGE QUALITY IN RELATION TO THE NUMBER OF MIXTURES

Num. of mixtures	PSNR	SSIM	VSNR
3	38.89	0.9305	13.92
6	38.91	0.9309	13.92
9	38.92	0.9308	13.92
12	**39.00**	**0.9315**	**13.93**
15	38.71	0.9295	13.92

Table II shows the comparison of the difference in image quality in relation to the number of mixtures using GMM-based super resolution. The performance of 12 mixtures was slightly better than that of other numbers of mixtures. The following experiments were carried out using the 3×3 patch size and 12 mixtures of GMM.

Table III shows the evaluation results using PSNR, SSIM and VSNR. Comparing the GMM-PLS-based super-resolution with the previously known techniques, such as example-based super-resolution or sparse-coding-based super-resolution, it can be seen that the GMM-PLS-based method obtained good evaluation values. Also, the performance of the GMM-PLS-based super-resolution was slightly better than the GMM-based super-resolution.

VI. CONCLUSION

In this paper, a super-resolution technique was proposed that employs GMM-PLS-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. Necessary future work will include expanding the method to improve the conversion function by employing other correlation analysis.

ACKNOWLEDGMENTS

This research was supported in part by MIC SCOPE.

Table III
COMPARISON OF THE SUPER-RESOLUTION IMAGES USING PSNR, SSIM AND VSNR

Image	Method	PSNR	SSIM	VSNR
(a)	Bicubic	33.07	0.802	16.20
	Example-Based SR	32.23	0.754	15.35
	Sparse-Coding	34.20	0.861	17.05
	GMM-Based SR	35.57	**0.916**	**17.43**
	GMM+PLS SR	**36.74**	**0.916**	**17.32**
(b)	Bicubic	36.43	0.882	13.63
	Example-Based SR	35.42	0.849	13.32
	Sparse-Coding	37.71	0.908	13.87
	GMM-Based SR	39.00	**0.932**	**13.93**
	GMM+PLS SR	**39.50**	**0.932**	13.90
(c)	Bicubic	34.82	0.910	15.19
	Example-Based SR	33.56	0.883	14.21
	Sparse-Coding	36.91	0.945	15.75
	GMM-Based SR	39.22	**0.973**	**15.87**
	GMM+PLS SR	**41.08**	**0.973**	15.76
(d)	Bicubic	32.50	0.775	13.86
	Example-Based SR	32.03	0.745	13.44
	Sparse-Coding	33.36	0.835	14.35
	GMM-Based SR	**33.97**	0.880	14.52
	GMM+PLS SR	33.71	**0.884**	**14.57**

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, 2010.

[3] Takashiro Ogawa and Miki Haseyama, "Adaptive reconstruction method of missing textures based on perceptually optimized algorithm," *ICASSP*, pp. 1157–1160, 2011.

[4] Y. Stylianou, O. Cappe, and E. Moulines, "Statistical methods for voice quality transformation," *EUROSPEECH*, pp. 447–450, 1995.

[5] Y. Ogawa, Y. Ariki, and T. Takiguchi, "Super-Resolution by GMM Based Conversion Using Self-Reduction Image," *ICASSP*, pp. 1285–1288, 2012.

[6] E. Helander, T. Virtanen, and J. Nurminen, and M. Gabbouj, "Voice Conversion Using Partial Least Squares Regression," *IEEE Trans. on ASLP*, vol. 18, no. 5, pp. 912–921, 2010.

[7] S. de Jong, "SIMPLS: An alternative approach to partial least squares regression," *Chemometrics Intell. Lab. Syst.*, vol. 18, no. 3, pp. 251-263, 1993.

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

[9] D. M. Chandler and S. S. Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.