

An Example-based Prior Model for Text Image Super-resolution

Jangkyun Park, Younghee Kwon and Jin Hyung Kim

Division of Computer Science, KAIST

Guseong-dong, Yuseong-gu, Daejeon, Korea

E-mail: {jpark,kyhee,jkim}@ai.kaist.ac.kr

Abstract

This paper presents a prior model for text image super-resolution in the Bayesian framework. In contrast to generic image super-resolution task, super-resolution of text images can be benefited from strong prior knowledge of the image class: Firstly, low-resolution images are assumed to be generated from a high-resolution image by a sort of degradation which can be grasped through example pairs of the original and the corresponding degradation; Secondly, text images are composed of two homogeneous regions, text and background regions. These properties were represented in a Markov Random Field (MRF) framework. Experiments showed that our model is more appropriate to text image super-resolution than the other prior models.

1. Introduction

Super-resolution is a technique to produce a high-resolution image from a set of low-resolution images of the same scene [8]. Bayesian super-resolution provides a chance to utilize a prior knowledge over the high-resolution images as a stochastic approach for super-resolution [4].

Many works in Bayesian super-resolution have been focused on the super-resolution for general images, while some for domain-specific images like facial images. Although text images have an important role in image processing and OCR (Optical Character Recognition), the text-specific prior model is rare in Bayesian super-resolution. There are several works for text image super-resolution such as the prior model with edge-preserving functions like Huber penalty function [1], the prior model based on domain-specific sample images [2], and so on. However, these prior models didn't reflect any text image property, so the super-resolved results of text images were not satisfactory. On the other hand, Donaldson and Myers proposed a text-specific prior model which modeled the bimodality and the local smoothness with step discontinuity [3]. This

prior showed the best results among the previous prior models for text image super-resolution. However, as the low-resolution images are severely blurred, the super-resolved results get poorer and the stroke information gets omitted.

Here, we present a new image prior model based on examples for text image super-resolution in Bayesian super-resolution framework. Basic ideas have come from utilizing the underlying high-resolution image from training examples to obtain the extra information over the high-resolution images and modeling text image property that a text image is composed of two homogeneous regions, a text region and a background region. We model a prior model via an MRF to consider the local characteristics of images. According to this framework, an appropriate clique system and energy functions are proposed.

2. Image degradation model

Many researches in the Bayesian super-resolution framework assume a common image degradation model [1-5]. According to this model, each low-resolution image is assumed to be generated from the high-resolution image through transforming, blurring, sub-sampling and additive noise, independently to others. Because we focus on the construction of an image prior model, we adopt the image degradation model in [4] which is well established.

In this image degradation model, K low resolution images $\mathbf{y}^{(k)}$ are assumed to be generated from the high-resolution image \mathbf{x} :

$$\mathbf{y}^{(k)} = \mathbf{W}^{(k)} \mathbf{x} + \boldsymbol{\varepsilon}^{(k)}, \quad k = 1, \dots, K$$

where $\boldsymbol{\varepsilon}^{(k)}$ is a vector of i.i.d. Gaussian noise $\varepsilon_i^{(k)} \sim N(0, \beta_G^{-1})$ and the transformation matrix $\mathbf{W}^{(k)}$ includes transforming and blurring in the Gaussian form.

In this paper, we assume that the matrix $\mathbf{W}^{(k)}$ is known. Tipping have already suggested how to estimate $\mathbf{W}^{(k)}$ [4]. From this image degradation model, the likelihood of the high-resolution image is represented as

$$p(\mathbf{y}^{(k)} | \mathbf{x}) = \left(\frac{\beta_G}{2\pi}\right)^{M/2} \exp\left\{-\frac{\beta_G}{2} \|\mathbf{y}^{(k)} - \mathbf{W}^{(k)} \mathbf{x}\|^2\right\}$$

where each low-resolution image has M pixels.

3. Proposed process and prior model

Proposed super-resolution process embeds the Bayesian super-resolution framework through iteration (Fig. 1). The bicubic-interpolated image of an observed low-resolution image is given as an initial intermediate-resolved image. And the intermediate-resolved image is improved by MAP (Maximum A Posteriori) estimation using the conjugate gradient algorithm in the Bayesian framework. At each iteration step, the likelihood and the prior are calculated for the MAP estimation. The quality of the final super-resolved result varies with the number of iterations. The prior is calculated based on the underlying high-resolution image and the edge map. The underlying high-resolution image comes from training examples and the edge map is extracted from the underlying high-resolution image in the previous iteration by edge detection.

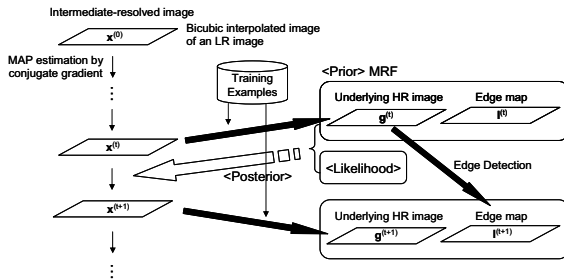


Fig. 1. Overall process of super-resolution.

In our model, we model the high-resolution image as an MRF. This means that the probability distribution of a node on the intermediate-resolved image is conditionally independent of all but the neighborhood of the node. Fig. 2 shows the neighborhood of a node of the MRF and Fig. 3 shows cliques in the neighborhood system. In Fig. 2, circles, rectangles and hexagons indicate intermediate-resolved image nodes \mathbf{x} , edge nodes \mathbf{l} and underlying high-resolution image nodes \mathbf{g} , respectively. And arcs represent statistical dependencies between nodes. In Fig. 3, we define two kinds of cliques $C_1 \in \mathbf{C}_1$ and $C_2 \in \mathbf{C}_2$. Therefore, for each node on intermediate-resolved image, there are five cliques related to it, one C_1 clique and four C_2 cliques. The clique C_1 represents the dependency between the intermediate-resolved image and the underlying high-resolution image. In clique C_1 , we get the extra information over the high-resolution image. And the clique C_2 represents the dependency between two neighbor nodes on the intermediate-

resolved image and an edge node between them. In clique C_2 , the selective smoothing using an edge map is performed to improve the local smoothness of each region of text region and background region.

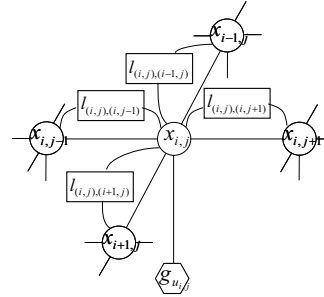


Fig. 2. Clique system in proposed MRF.

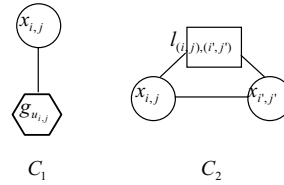


Fig. 3. Cliques.

The prior distribution of our model is represented by the joint probability of \mathbf{x} in MRF as the Gibbs distribution

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left\{-\frac{\sum_{c \in \mathbf{C}} E_c}{T}\right\} = \frac{1}{Z} \exp\left\{-\frac{\sum_{C_1 \in \mathbf{C}_1} E_{C_1}(\{x_{i,j}, g_{u_{i,j}}\}_{(i,j), u_{i,j} \in C_1}) + \sum_{C_2 \in \mathbf{C}_2} E_{C_2}(\{x_{i,j}, l_{(i,j),(i',j')}\}_{(i,j),(i',j') \in C_2})}{T}\right\}$$

where Z is the normalizing constant, T is the temperature and \mathbf{C} is a set of cliques. (i,j) , $u_{i,j}$, and (i',j') indicate the spatial position of \mathbf{x} , its one-to-one mapped spatial position of \mathbf{g} and its four neighbors, respectively. The energy function for the clique C_1 is given by

$$E_{C_1} = \beta(x_{i,j} - g_{u_{i,j}})^2$$

where β is a weighting parameter. And the energy function for the clique C_2 is given by

$$E_{C_2} = \alpha(x_{i,j} - x_{i',j'})^2 l_{(i,j),(i',j')}$$

where α is a weighting parameter and $l_{(i,j),(i',j')}$ represents whether (i,j) and (i',j') are in the same region or not.

$$l_{(i,j),(i',j')} = \begin{cases} 1, & \text{if } (i,j) \text{ and } (i',j') \text{ are in same region} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore in the clique C_2 , selective smoothing is performed. It is determined by an edge map whether (i,j) and (i',j') are in the same region or not.

The underlying high-resolution image gives extra information over the high-resolution image. Pickup [2] used this strategy but their result of the underlying high-resolution image is poor for the blurred input images because they inferred the underlying high-resolution image with the direct matching between the severely blurred intermediate-resolved image and the high-resolution training examples. On the other hand, Baker and Kanade [5], and Freeman et al. [6] utilized pairs of high-resolution images and low-resolution images as a training example setting successfully. Inspired from this, in our prior model, training examples are composed of pairs of high-resolution images and their blurred images. Blurred training examples are made from high-resolution training examples by Gaussian blurring with 3×3 center weighted Gaussian mask. An underlying high-resolution image pixel $g_{x_{i,j}}$ is a center pixel intensity of the high-resolution training example patch which is one-to-one mapped patch of the blurred training example patch. And the blurred training example patch is determined to have the minimum squared error to the intermediate-resolved image patch of $x_{i,j}$.

Since the underlying high-resolution image pixels are determined independently, each region in the super-resolved result is not locally smooth. The selective smoothing in the clique C_2 improves the local smoothness of each region, as well as preserves edges. To indicate region discontinuity, the binary edge pixel map is generated from the underlying high-resolution image in the previous iteration by Canny edge detection algorithm. If the values in the edge map according to given two neighbor nodes are the same, those nodes are treated as in the same region.

4. Super-resolved result estimation

The posterior distribution over the high-resolution image \mathbf{x} is of the form,

$$p(\mathbf{x} | \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(K)}) \propto p(\mathbf{x}) \cdot p(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(K)} | \mathbf{x}).$$

Since low-resolution images are assumed to be generated independently, the likelihood is given by multiplication:

$$p(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(K)} | \mathbf{x}) = \prod_k p(\mathbf{y}^{(k)} | \mathbf{x}).$$

By taking the negative log to the posterior probability, we have

$$-\log p(\mathbf{x} | \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(K)}) \propto \sum_{C_1 \in C_1} E_{C_1}(\{x_{i,j}, g_{u_{i,j}}\}_{(i,j), u_{i,j} \in C_1}) + \sum_{C_2 \in C_2} E_{C_2}(\{x_{i,j}, l_{(i,j),(i',j')} \}_{(i,j),(i',j') \in C_2}) + \sum_k \|\mathbf{y}^{(k)} - \mathbf{W}^{(k)} \mathbf{x}\|^2.$$

The super-resolved result is estimated to minimize the above function with respect to \mathbf{x} . This function was optimized by the conjugate gradient descent algorithm.

5. Experimental results

To evaluate the performance of the proposed prior model, MLE (Maximum Likelihood Estimation), Gaussian process prior model [4], sampled texture prior model [2] and bimodal prior model [3] were compared to our model using two measure, RMSE (Root Mean Squared Error) between the true test image and the super-resolved result, and RMSEB (Root Mean Squared Error of Binarized images) between the binarized true test image and the binarized super-resolved result. All weight parameters were adjusted to have the minimum RMSE in each case. Niblack binarization method [7] was used for the binarization. The true test image was binarized once and fixed. And the binarization parameters for the super-resolved results were determined to have the minimum RMSE to the binarized true test image.

All images were gray scaled and normalized to $[0,1]$. High-resolution images were gathered from a journal by a 200 dpi scanner. Blurred training examples were created from high-resolution training examples by Gaussian blurring with blur diameter 0.4 pixel. Four 24×24 low-resolution images were created from the true test image by the image degradation model. For each intermediate-resolved pixel, 5×5 surrounding window was used to search the best matching training example patch to identify the underlying high-resolution pixel. In the experiments for the sampled texture prior model and our prior model, the same high-resolution training examples were used. The number of iterations of the conjugate gradient algorithm was limited to 30 iterations.

Fig. 4 shows the 48×48 true test image and its binarized image. And Fig. 5 shows the results of the prior models. Training examples with the same font face and size as the true test image were used. γ denotes the width of PSF (Point Spread Function) as the degree of blurring in the image degradation model. As γ gets larger, the more severely blurred the low-resolution image gets. Bicubic interpolated result is similar to the low-resolution image. As far as the transformation matrix \mathbf{W} was calculated correctly, MLE gives plausible results. Gaussian process prior model gives better looking results than bicubic one

because the correlations between pixels were considered. Results of the sampled texture prior model look better than the results of Gaussian process prior model because they obtained extra information over the high-resolution image from sample images. However, as γ is increased, side effects are shown due to the direct comparison between intermediate-resolved image and the high-resolution sample image. Text-specific bimodal prior model gives better looking results than previous results. And the results of our model look best among six results. In Fig. 6, RMSE and RMSEB are shown. In the error sense, the results of our model are the best regardless of the degree of blurring.

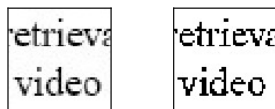


Fig. 4. True test image and its binarized image.

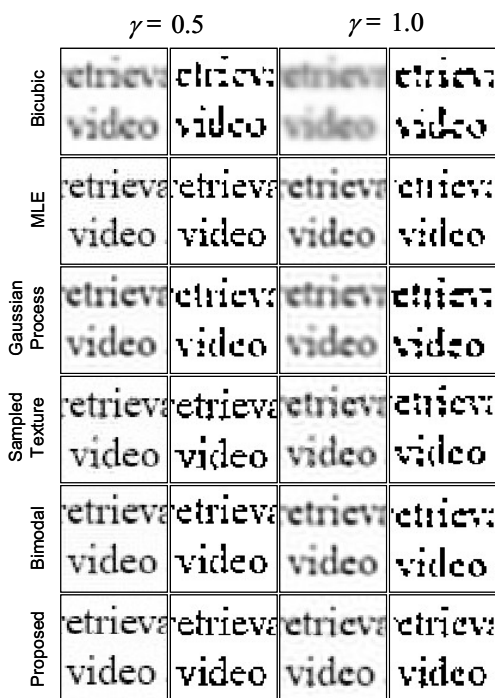


Fig. 5. Super-resolved results shown.

Fig. 7 shows the application of the proposed prior model with training examples whose font size is different from the true test image. Five font sizes were chosen: half size, ten percent smaller and larger, same size and double size. Training examples within ten-percent size variation give nearly same results as the training examples in the same font size, whereas training examples in half size or double size result poor. In Fig.

8, RMSE and RMSEB also reveal that the training examples with similar font size give good results.

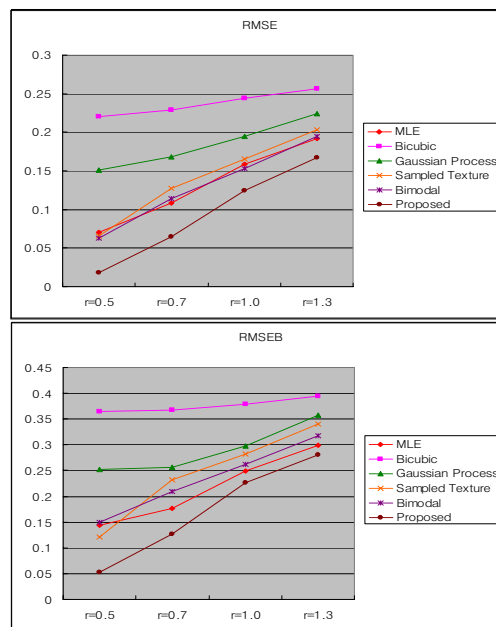


Fig. 6. RMSE and RMSE for binarized ones.

And Fig. 9 shows the results of our model with training examples in different font. Test image font is 'Times New Roman' and training example font is 'Arial'. In this case, training examples in different font gives worse results than training examples in same font. However, in Fig. 10, although the result is worse, it has still better RMSE than results of other prior models.

6. Conclusion

We have proposed a new example-based image prior model for text images in Bayesian super-resolution framework. In the proposed prior model, the underlying high-resolution image carries extra information about the high-resolution image successfully and the selective smoothing using an edge map strengthened the homogeneity of each region of text region and background region.

Results showed that the proposed prior model gave significantly improved results compared with the other prior models in the sense of looking, as well as it showed minimum RMSE and minimum RMSEB. From the binarized results, we can expect that our model helps binarization for text images in low quality.

One possible argument about our model is that collecting training examples is an intensive work because of the variation of fonts. But experiments have

shown that even with the different font (in face and size) examples, our model still result better compared with the other prior models – the propose model is robust enough about fonts.

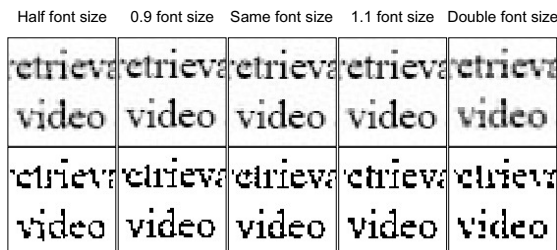


Fig. 7. Super-resolved results with training examples in different font size.

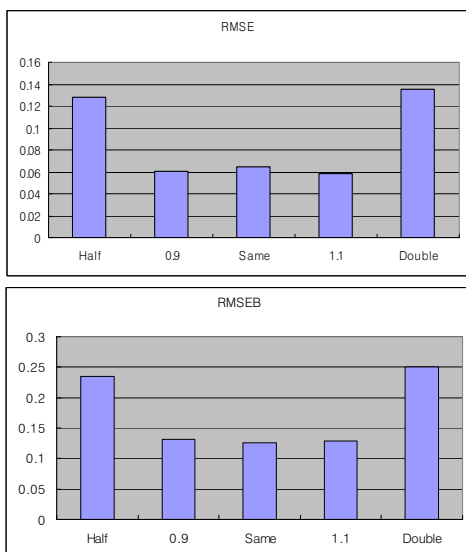


Fig. 8. RMSE and RMSEB.

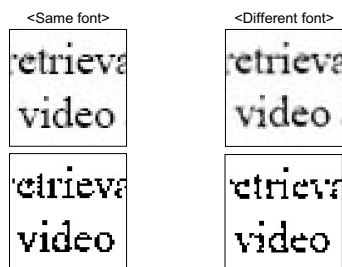


Fig. 9. Super-resolved results with training examples in different font.

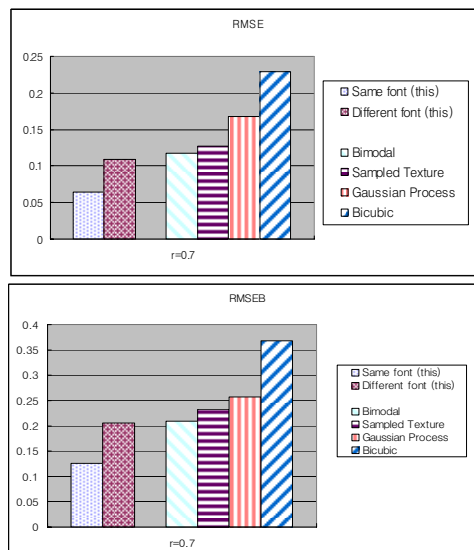


Fig. 10. RMSE and RMSEB.

7. References

- [1] D. P. Capel, *Image Mosaicing and Super-resolution*, PhD thesis, University of Oxford, 2001.
- [2] L. C. Pickup, S. J. Roberts, and A. Zisserman, "A sampled texture prior for image super-resolution," *Advances in Neural Information Processing Systems* 16, MIT Press, 2004.
- [3] K. Donaldson and G. K. Myers, "Bayesian super-resolution of text in video with a text-specific bimodal prior," website (<http://www.esd.sri.com/projects/vace/super-res.html>), 2003.
- [4] M. E. Tipping and C. M. Bishop, "Bayesian image super-resolution," *Advances in Neural Information Processing Systems* 15, MIT Press, 2003.
- [5] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167-1183, 2002.
- [6] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, issue 2, pp. 56-65, March 2002.
- [7] W. Niblack, *An Introduction to Digital Image Processing*, pp. 115-116, Englewood Cliffs, N. J., Prentice Hall, 1986.
- [8] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21-36, May 2003.