

TOTAL VARIATION SUPER RESOLUTION USING A VARIATIONAL APPROACH

S. Derin Babacan¹, Rafael Molina², Aggelos K. Katsaggelos¹

¹Department of Electrical Engineering
and Computer Science
Northwestern University, Evanston, IL 60208, USA
sdb@northwestern.edu, aggk@eecs.northwestern.edu

²Departamento de Ciencias
de la Computación e I.A.
Universidad de Granada, 18071 Granada, Spain
rms@decsai.ugr.es

ABSTRACT

In this paper we propose a novel algorithm for super resolution based on total variation prior and variational distribution approximations. We formulate the problem using a hierarchical Bayesian model where the reconstructed high resolution image and the model parameters are estimated simultaneously from the low resolution observations. The algorithm resulting from this formulation utilizes variational inference and provides approximations to the posterior distributions of the latent variables. Due to the simultaneous parameter estimation, the algorithm is fully automated so parameter tuning is not required. Experimental results show that the proposed approach outperforms some of the state-of-the-art super resolution algorithms.

Index Terms— Super resolution, total variation, variational methods, parameter estimation, Bayesian methods.

1. INTRODUCTION

High resolution (HR) images can in some cases be obtained directly from high-resolution acquisition devices. However, due to theoretical and practical limitations, in most cases the resolution of the acquired images are lower than desired. These limitations include the increased cost, data transfer rate and the amount of shot noise due to the size of the digital sensor. Recently signal processing techniques have been utilized as an alternative to increase the resolution of digital images.

Super resolution describes the process of reconstructing an HR image from a set of low resolution (LR) observations. The LR images typically are undersampled, degraded and shifted versions of the HR image with subpixel displacements. The recovery of the HR image is possible through the use of the subpixel displacements between the observations.

Although the super resolution literature is rich (see [1] for an extensive review) it is still an open and widely investigated topic. Recently, motivated by its success in image recovery problems, the use of the total variation (TV) function and its variants has become popular in super resolution. Both regularization-based [2,3] and Bayesian [4] formulations have been proposed which utilize TV functions to characterize the HR images. However, both of these approaches involve certain model parameters to be set by the user, which is in general a difficult task. To our knowledge no work has been reported on simultaneously estimating the algorithm parameters and the HR image.

This work has been partially supported by the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018) and the Ministerio de Educacion y Ciencia under contract TIN2007-65533.

In this paper we propose a hierarchical Bayesian methodology for super resolution where the LR observations and the unknown HR image, as well as their associated hyperparameters (observation and acquisition noise, and the variance of the HR image) are modeled in two stages. We apply variational inference methods to this model and propose an algorithm which simultaneously provides estimates to the unknowns.

The rest of this paper is organized as follows. In Sec. 2 we formulate the LR image acquisition system by a linear time invariant model. The unknown variables in this model are cast into a hierarchical Bayesian framework as presented in Sec. 3. The variational inference to estimate the unknowns and the proposed algorithm are presented in Sec. 4. Experimental results are presented in Sec. 5 and conclusions are drawn in Sec. 6.

2. PROBLEM FORMULATION

Consider a set of L low resolution (LR) images $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_L\}$, where $\mathbf{y}_i, i = 1, \dots, L$ represents the i^{th} low resolution (LR) image. The goal is to reconstruct the high resolution (HR) image \mathbf{x} that would be observed under ideal conditions. The HR image is of size $P_1 N \times P_2 M$ and each of the LR images is of size $N \times M$, so that the horizontal and vertical magnification factors are P_1 and P_2 , respectively. Each image can be transformed to a column vector by lexicographically ordering the pixels, so that the $(P_1 N \times P_2 M) \times 1$ vector \mathbf{x} represents the HR image, and the $N M \times 1$ vector represents the i^{th} LR image.

We denote by the $P_1 P_2 N M \times P_1 P_2 N M$ matrix \mathbf{C}_i the warping matrix that maps the HR image \mathbf{x} to the high-resolution version \mathbf{x}_i of the LR image \mathbf{y}_i . The $N M \times P_1 P_2 N M$ matrix \mathbf{A} is the down-sampling matrix and the $P_1 P_2 N M \times P_1 P_2 N M$ matrices \mathbf{H}_i are the PSFs of the blurs. Then the LR image acquisition process can be modeled mathematically as follows

$$\mathbf{y}_i = \mathbf{A} \mathbf{H}_i \mathbf{C}_i \mathbf{x} + \mathbf{n}_i = \mathbf{B}_i \mathbf{x} + \mathbf{n}_i, \quad (1)$$

where \mathbf{n}_i is the combination of the registration and acquisition noise. In this work the matrices \mathbf{H}_i and \mathbf{A} are assumed known. Their estimation using variational methods is left as future work.

3. HIERARCHICAL BAYESIAN MODEL

Utilizing a Bayesian analysis, the unknown \mathbf{x} and the observed LR images $\mathbf{y}_i, i = 1, \dots, L$ are treated as stochastic quantities and prior probability distributions on them are defined. Since these distributions will also have their model parameters, called *hyperparameters*, we adopt a hierarchical Bayesian model with two stages. In the first stage, the HR image and the observation noise are modeled using

some unknown hyperparameters, and the hyperprior distributions of the hyperparameters are modeled in the second stage.

3.1. First Stage: LR Image Acquisition Model

Using the model in Eq. (1) and assuming that \mathbf{n}_i is a zero-mean white Gaussian noise with the variance β_i^{-1} , the likelihood of the LR image \mathbf{y}_i can be written as

$$p(\mathbf{y}_i|\mathbf{x}, \beta_i) \propto \beta_i^{NM/2} \exp\left[-\frac{\beta_i}{2} \|\mathbf{y}_i - \mathbf{B}_i\mathbf{x}\|^2\right]. \quad (2)$$

Assuming statistical independence of the noise between LR images, the probability distribution of the set of LR images \mathbf{y} given \mathbf{x} can be expressed as

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}, \underline{\beta}) &= \prod_i p(\mathbf{y}_i|\mathbf{x}, \beta_i) \\ &= \left[\prod_i \beta_i^{NM/2} \right] \exp\left(-\frac{1}{2} \sum_i \beta_i \|\mathbf{y}_i - \mathbf{B}_i\mathbf{x}\|^2\right), \end{aligned} \quad (3)$$

where $\underline{\beta} = (\beta_1, \dots, \beta_L)$.

3.2. First Stage: HR Image Model

As the prior model for the HR image \mathbf{x} we utilize the quadratic approximation of the TV prior

$$p(\mathbf{x}|\alpha) \propto \alpha^{P_1 P_2 NM/2} \exp\left[-\frac{1}{2} \alpha \text{TV}(\mathbf{x})\right], \quad (4)$$

where

$$\text{TV}(\mathbf{x}) = \sum_j \sqrt{(\Delta_j^h(\mathbf{x}))^2 + (\Delta_j^v(\mathbf{x}))^2}. \quad (5)$$

The operators $\Delta_j^h(\mathbf{x})$ and $\Delta_j^v(\mathbf{x})$ correspond to, respectively, horizontal and vertical first order differences, at pixel j , that is, $\Delta_j^h(\mathbf{x}) = x_j - x_{l(j)}$ and $\Delta_j^v(\mathbf{x}) = x_j - x_{a(j)}$, where $l(j)$ and $a(j)$ denote the nearest neighbors of j , to the left and above, respectively.

3.3. Second Stage: Hyperpriors on the Hyperparameters

In the second stage of the Bayesian model, we use flat improper hyperpriors on α and β_i , that is, we utilize

$$p(\alpha) \propto \text{const}, \quad p(\beta_i) \propto \text{const}, \quad i = 1, \dots, L. \quad (6)$$

Note that with this choice of the hyperpriors the hyperparameters α and β_i are solely estimated from the LR observations \mathbf{y} .

Finally, combining the first and second stage of the problem modeling in Eqs. (3), (4) and (6) we find the joint probability distribution $p(\alpha, \underline{\beta}, \mathbf{x}, \mathbf{y}) = p(\alpha, \underline{\beta})p(\mathbf{x}|\alpha)p(\mathbf{y}|\mathbf{x}, \underline{\beta})$.

4. VARIATIONAL INFERENCE

The Bayesian paradigm dictates that inference on $(\alpha, \underline{\beta}, \mathbf{x})$ should be based on

$$p(\alpha, \underline{\beta}, \mathbf{x} | \mathbf{y}) = \frac{p(\alpha, \underline{\beta}, \mathbf{x}, \mathbf{y})}{p(\mathbf{y})}, \quad (7)$$

However, the posterior $p(\alpha, \underline{\beta}, \mathbf{x} | \mathbf{y})$ can not be found in closed form. We therefore apply variational methods to approximate this distribution by the distribution $q(\alpha, \underline{\beta}, \mathbf{x})$ and utilize a mean field

approximation so that $q(\alpha, \underline{\beta}, \mathbf{x}) = q(\alpha, \underline{\beta})q(\mathbf{x})$. Additionally, we assume that $q(\mathbf{x})$ is a degenerate distribution, that is, a distribution which takes one value, \mathbf{x}^k with probability one and the rest with probability zero, which can be expressed mathematically as

$$q^k(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}^k). \quad (8)$$

The variational criterion used to find this approximation is to minimize the Kullback-Leibler (KL) distance between $q(\alpha, \underline{\beta}, \mathbf{x})$ and the posterior $p(\alpha, \underline{\beta}, \mathbf{x} | \mathbf{y})$, which is given by

$$\begin{aligned} C_{KL}(q(\alpha, \underline{\beta}, \mathbf{x}) \| p(\alpha, \underline{\beta}, \mathbf{x}|\mathbf{y})) \\ &= \int \int \int q(\alpha, \underline{\beta}, \mathbf{x}) \log\left(\frac{q(\alpha, \underline{\beta}, \mathbf{x})}{p(\alpha, \underline{\beta}, \mathbf{x}|\mathbf{y})}\right) d\alpha d\underline{\beta} d\mathbf{x} \\ &= \int \int \int q(\alpha, \underline{\beta}, \mathbf{x}) \log\left(\frac{q(\alpha, \underline{\beta}, \mathbf{x})}{p(\alpha, \underline{\beta}, \mathbf{x}, \mathbf{y})}\right) d\alpha d\underline{\beta} d\mathbf{x} + \text{const}, \end{aligned} \quad (9)$$

which is always non negative and equal to zero only when $q(\alpha, \underline{\beta}, \mathbf{x}) = p(\alpha, \underline{\beta}, \mathbf{x}|\mathbf{y})$.

The use of the TV prior makes the integral in Eq. (9) difficult to evaluate so a minorization of the TV prior is utilized. By defining the following functional $M(\alpha, \mathbf{x}, \mathbf{u})$, for α, \mathbf{x} , and a $P_1 P_2 NM$ -dimensional vector $\mathbf{u} \in (R^+)^{P_1 P_2 NM}$

$$\begin{aligned} M(\alpha, \mathbf{x}, \mathbf{u}) &= \text{const} \times \alpha^{P_1 P_2 NM/2} \\ &\times \exp\left[-\frac{\alpha}{2} \sum_j \frac{(\Delta_j^h(\mathbf{x}))^2 + (\Delta_j^v(\mathbf{x}))^2 + u_j}{\sqrt{u_j}}\right], \end{aligned} \quad (10)$$

and using the following inequality, also used in [5,6], for $w \geq 0$ and $z > 0$

$$\sqrt{wz} \leq \frac{w+z}{2} \Rightarrow \sqrt{w} \leq \frac{w+z}{2\sqrt{z}}. \quad (11)$$

we obtain a lower bound for the image prior (c a constant)

$$p(\mathbf{x}|\alpha) \geq c \cdot M(\alpha, \mathbf{x}, \mathbf{u}), \quad (12)$$

and the following lower bound for the joint probability distribution

$$\begin{aligned} p(\alpha, \underline{\beta}, \mathbf{x}, \mathbf{y}) &\geq p(\alpha)p(\underline{\beta})M(\alpha, \mathbf{x}, \mathbf{u})p(\mathbf{y}|\mathbf{x}, \underline{\beta}) \\ &= F(\alpha, \underline{\beta}, \mathbf{x}, \mathbf{u}, \mathbf{y}). \end{aligned} \quad (13)$$

Utilizing Eq. (13) we obtain the following upper bound for the KL distance

$$\begin{aligned} &\int \int \int q(\alpha, \underline{\beta}, \mathbf{x}) \log\left(\frac{q(\alpha, \underline{\beta}, \mathbf{x})}{p(\alpha, \underline{\beta}, \mathbf{x}, \mathbf{y})}\right) d\alpha d\underline{\beta} d\mathbf{x} \\ &\leq \min_{\mathbf{u}} \int \int \int q(\alpha, \underline{\beta}, \mathbf{x}) \log\left(\frac{q(\alpha, \underline{\beta}, \mathbf{x})}{F(\alpha, \underline{\beta}, \mathbf{x}, \mathbf{u}, \mathbf{y})}\right) d\alpha d\underline{\beta} d\mathbf{x} \\ &= \mathcal{M}(q(\mathbf{x}, \alpha, \underline{\beta})). \end{aligned} \quad (14)$$

Therefore, instead of minimizing the KL distance, we minimize the upper bound $\mathcal{M}(q(\mathbf{x}, \alpha, \underline{\beta}))$ by finding a sequence of distributions $\{q^k(\alpha, \underline{\beta}, \mathbf{x})\}$. We adopt an alternating minimization approach where at each step this upper bound is minimized with respect to one of the variables $\{\mathbf{x}, \alpha, \underline{\beta}\}$ while holding others as constant, which leads to an iterative procedure shown in Algorithm 1.

Let us now develop the solutions for each of the steps within the while loop of Algorithm 1. The minimization is carried out

Algorithm 1 Variational TV Super Resolution

Given $\mathbf{u}^1 \in (R^+)^{P_1 P_2 N M}$ and $q^1(\alpha, \beta)$
while Convergence criterion is not met **do**
 1. Estimate HR image \mathbf{x} using (15).
 2. Estimate \mathbf{u} using (18).
 3. Estimate hyperparameters α, β using (21)-(22).

first with respect to the HR image \mathbf{x} . By taking the derivative of $\mathcal{M}(q(\mathbf{x}, \alpha, \beta))$ and setting it equal to zero we obtain at iteration k

$$\mathbf{Q}(\alpha, \beta) \mathbf{x}^k = \sum_i \beta_i \mathbf{B}_i^t \mathbf{y}_i, \quad (15)$$

$$\mathbf{Q}(\alpha, \beta) = \sum_i \beta_i \mathbf{B}_i^t \mathbf{B}_i + \alpha \Delta^t \begin{bmatrix} W(\mathbf{u}^k) & \mathbf{0} \\ \mathbf{0} & W(\mathbf{u}^k) \end{bmatrix} \Delta, \quad (16)$$

with $\Delta = [(\Delta^h)^t (\Delta^v)^t]^t$ and $W(\mathbf{u}^k)$ a $P_1 P_2 N M \times P_1 P_2 N M$ diagonal matrix of the form

$$W(\mathbf{u}^k) = \text{diag} \left(\left[u_j^k \right]^{-\frac{1}{2}} \right), \quad j = 1, \dots, P_1 P_2 N M. \quad (17)$$

The vector \mathbf{u}^{k+1} is found in the second step as

$$u_j^{k+1} = \left[\Delta_j^h(\mathbf{x}^k) \right]^2 + \left[\Delta_j^v(\mathbf{x}^k) \right]^2, \quad j = 1, \dots, P_1 P_2 N M. \quad (18)$$

It is clear from this equation that the vector \mathbf{u}^{k+1} represents the local spatial activity in the HR image \mathbf{x}^k . Therefore, matrix $W(\mathbf{u}^k)$ in Eq. (17) can be interpreted as the *spatial adaptivity* matrix since it controls the amount of smoothing at each pixel location depending on the strength of the intensity variation at that pixel, as expressed by the horizontal and vertical intensity gradients.

The direct solution of the HR image estimate in Eq. (15) is practically hard to compute because of the inversion of the huge matrix $\mathbf{Q}(\alpha, \beta)$. Therefore we use a conjugate gradient algorithm to find it numerically.

Finally, in the last step of the algorithm, the distributions of the hyperparameters $q^{k+1}(\alpha)$ and $q^{k+1}(\beta_i)$ are found by differentiating $\mathcal{M}(q(\mathbf{x}, \alpha, \beta))$ with respect to $q(\alpha, \beta_i)$ and setting it equal to zero. These distributions are Gamma distributions given by

$$q^{k+1}(\alpha) \propto \alpha^{P_1 P_2 N M / 2} \exp \left[-\alpha \sum_j \sqrt{u_j^{k+1}} \right] \quad (19)$$

and

$$q^{k+1}(\beta_i) \propto \beta_i^{N M / 2} \exp \left[-\beta_i \frac{\| \mathbf{y}_i - \mathbf{B}_i \mathbf{x}^k \|^2}{2} \right] \quad (20)$$

The means of these distributions are given by

$$\alpha^{k+1} = E_{q^{k+1}(\alpha)}[\alpha] = \frac{P_1 P_2 N M / 2 + 1}{\sum_j \sqrt{u_j^{k+1}}}, \quad (21)$$

$$\beta_i^{k+1} = E_{q^{k+1}(\beta_i)}[\beta_i] = \frac{N M + 2}{\| \mathbf{y}_i - \mathbf{B}_i \mathbf{x}^k \|^2}, \quad (22)$$

The algorithm is summarized above in Algorithm 1.

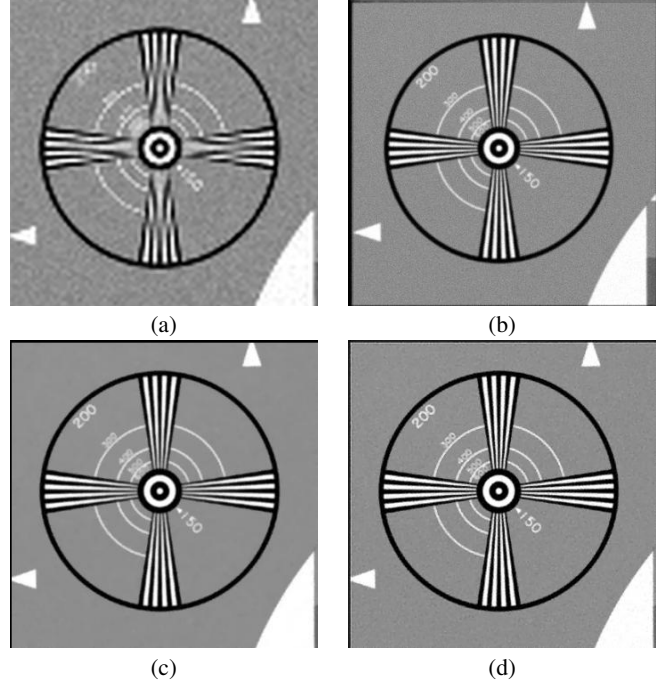


Fig. 1. Super resolution results (4x resolution increase) by (a) bicubic interpolation, (b) *MS*, (c) *FRS*, and (d) proposed algorithm.

5. EXPERIMENTAL RESULTS

In this section we compare the performance of the proposed algorithm with bicubic interpolation, the method proposed in [7], denoted by *MS*, and the fast super resolution method based on bilateral TV in [4], which is denoted by *FRS*. The parameters of the *FRS* algorithm are chosen according to [4] which resulted in the visually best results. For all experiments, the criterion $\| \mathbf{x}^k - \mathbf{x}^{k-1} \|^2 / \| \mathbf{x}^{k-1} \|^2 < 10^{-4}$ is used to terminate the proposed algorithm, and the CG threshold is set equal to 10^{-4} .

In the first experiment, we run the algorithms on the EIA dataset obtained from [4]. In this data set, the original HR frame is shifted by 16 different motion vectors, blurred, and downsampled by 4 to obtain 16 LR observations. The methods are used to obtain an HR image with a factor of 4 resolution increase in each direction. The restoration results are shown in Fig. (1). As expected, all super resolution algorithms result in better reconstructions than bicubic interpolation. Comparing the results in Fig. 1(b)-(d), it is clear that the reconstruction of the proposed algorithm gives the most visually enhanced result. This is justified by looking at Fig. (2) where the middle sections of the reconstructions are shown in detail. Note that the numbers and the intersection of the lines with the circle are sharper in the image obtained by the proposed algorithm than by the other algorithms.

We present another experiment with the 20 real LR images taken from the disk dataset from [4]. The motion is estimated using the pyramidal Lucas-Kanade optical flow algorithm, and the blur is assumed to be a 6x6 Gaussian with variance 1. The reconstructed HR images have a factor of two resolution enhancement and are shown in Fig. (3). Although motion estimation errors are present, the proposed algorithm produces a sharp HR image with less ringing artifacts than the other approaches. This can be observed more clearly in the detailed areas shown in Fig. (4).

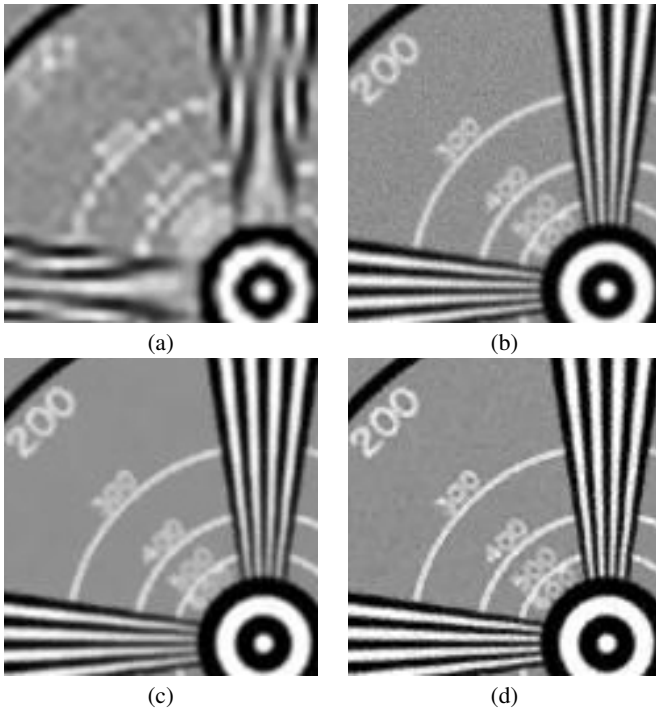


Fig. 2. Detailed areas of Fig. (1). (a) Bicubic interpolation, (b) *MS*, (c) *FRS*, and (d) proposed algorithm.

6. CONCLUSIONS

In this paper we presented a Bayesian super resolution method based on total variation image priors. Within a hierarchical Bayesian framework, the reconstructed HR image, the acquisition and motion estimation noise for each LR image is estimated simultaneously. Variational inference is applied to estimate the posterior distributions of the unknowns. The main novelty of the proposed algorithm is that the model parameters are estimated during the reconstruction so that the algorithm is fully automated whereas existing approaches require data-specific parameter tuning. Experimental results demonstrate that the proposed algorithm results in better HR reconstructions than existing approaches with both synthetic and real data.

7. REFERENCES

- [1] A. K. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video*, Morgan and Claypool, 2007.
- [2] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," *EURASIP Journal on Advances in Signal Processing*, no. 74585, 2007.
- [3] T. F. Chan, N. Ng, A. Yau, and A. Yip, "Superresolution image reconstruction using fast inpainting algorithms," *Applied and Computational Harmonic Analysis*, vol. 23, no. 1, pp. 3–24, July 2007.
- [4] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Processing*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
- [5] S. D. Babacan, R. Molina, and A.K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Processing*, vol. 17, no. 3, pp. 326–339, 2008.

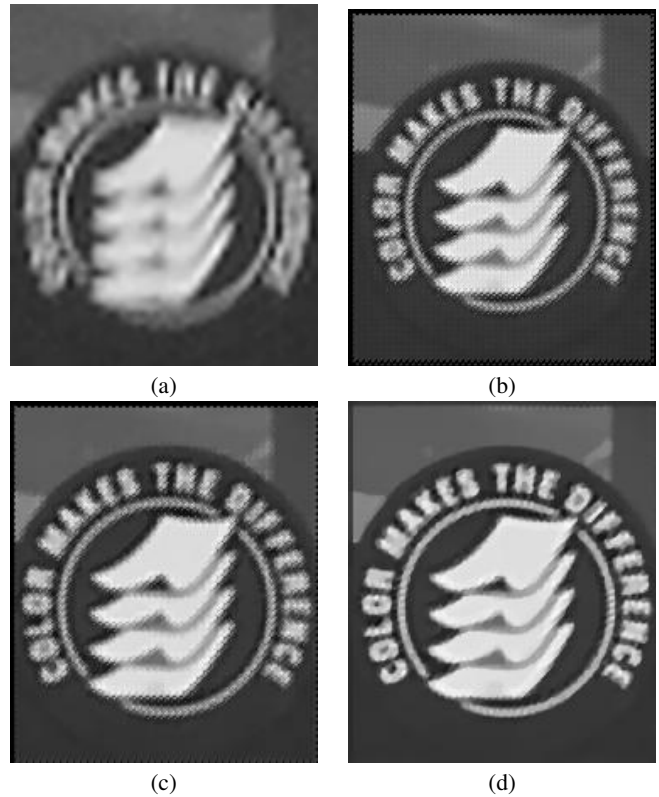


Fig. 3. Super resolution results (2x resolution increase) by (a) bicubic interpolation, (b) *MS*, (c) *FRS*, and (d) proposed algorithm.

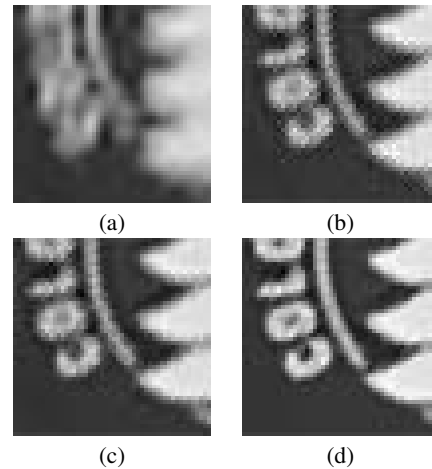


Fig. 4. Detailed areas of Fig. (3). (a) Bicubic interpolation, (b) *MS*, (c) *FRS*, and (d) proposed algorithm.

- [6] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Adaptive bayesian/total-variation image deconvolution: A majorization-minimization approach," in *EUSIPCO*, 2006.
- [7] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Processing*, vol. 10, no. 8, pp. 1187–1193, 2001.