RATE-DISTORTION ANALYSIS OF SUPER-RESOLUTION IMAGE/VIDEO DECODING

Keita TAKAHASHI [†], *Takeshi NAEMURA* [†], *and Masayuki TANAKA* [‡]

[†] The University of Tokyo, IRT Research Initiative, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8656, Japan
[‡] Tokyo Institute of Technology, Graduate School of Engineering, Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8550, Japan

ABSTRACT

An image/video communication scenario with super-resolution (SR) decoding, where the decoded images are upsampled using SR reconstruction at the receiver side, is analyzed from a theoretical perspective. To formulate the rate-distortion performance for such cases, we propose a new numerical model that combines a frequency-domain SR model and a rate-distortion theory for lossy image compression. We considered several factors that affect the reconstruction quality, and revealed that SR decoding performs better in low bitrates. We also conducted real-image simulations and confirmed that both the numerical analysis and real-image simulations exhibit quite similar tendencies, which supports the effectiveness of our numerical model.

Index Terms— image coding, super resolution, rate-distortion theory, image reconstruction

1. INTRODUCTION

Super resolution (SR) reconstruction refers to the process of reconstructing a high-resolution (HR) image from multiple low-resolution (LR) images containing the same scene object [1]. Several researchers successfully combined SR technology with video coding schemes to enhance the resolution at the receiver side [2, 3, 4], because the same scene objects are observed multiple times in successive frames. This condition also applies to multiview images used for 3-D image communication. However, to our knowledge, not much attention has been paid on theoretical formulations of the rate-distortion performance with SR reconstruction. Such formulations are important to see the theoretical limitations and trade-offs in introducing SR technology to communication systems.

This paper presents a rate-distortion analysis of super-resolution (SR) decoding, where decoded images are upsampled using SR reconstruction at the receiver side. We first formulate a new numerical model that combines a frequency-domain SR model which is equivalent to [5], and a rate-distortion theory for lossy image compression which was presented in [6]. Our numerical analysis suggests that SR decoding has its advantage in low bitrates, which agrees with the literatures [3, 4]. We also present real-image simulations to confirm that both the numerical analysis and real-image simulations exhibit quite similar tendencies in rate-distortion performance.

The problem statement is given as follows. Assume that multiple observations of the same scene object, which can be video or multi-view images, are captured, then compressed and transferred to the receiver. Figure 1 illustrates two coding scenarios for this purpose. The first scenario is referred to as high-resolution (HR) coding, in which the images are sampled, compressed, and decompressed in a high resolution. The second scenario is named as low-resolution (LR) coding with super-resolution (SR) decoding, in which sampling, compression, and decompression are conducted in a lower resolution, but the images are upsampled using SR reconstruction at the receiver side. The question we address in this paper is which scenario is better in terms of rate-distortion performance.

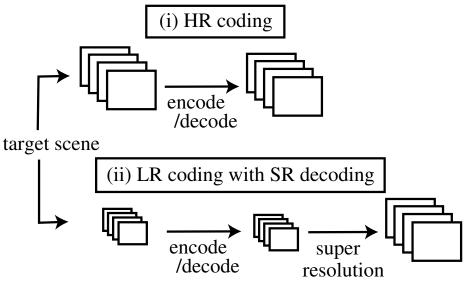

Fig. 1. Two coding scenarios compared in this paper.

To simplify the problem, let the resolution of the HR images be the twice of the LR images both in horizontal and vertical directions, and assume that each image is compressed independently without using inter-frame prediction. We also assume that the displacements between the LR images are modeled as global translations, and they can be registered without errors for SR reconstruction. Based on these assumptions, we compare rate-distortion performance of the two scenarios by numerical model analysis in Section 2 and real-image simulations in Section 3.

2. NUMERICAL MODEL ANALYSIS

In Sections 2.1–2.3, we formulate the processes of image formation, lossy image compression, and super-resolution reconstruction, in sequence. Section 2.4 describes the results of numerical analysis based on the theoretical model.

2.1. Image Formation Model

Let (u, v) be the coordinate system of the 2-D image signal. Assume an image formation model:

$$y(u,v) = \{p(u,v) * x(u,v)\}\,\delta_\Delta(u,v) + n_{ob}(u,v) \tag{1}$$

$$\delta_\Delta(u,v) = \sum_{m,n\in\mathcal{Z}} \delta(u - m\Delta - \zeta, v - n\Delta - \eta) \tag{2}$$

where $x(u, v)$ is the underlying continuous signal, and $y(u, v)$ is a digital image generated from $x(u, v)$. $p(u, v)$ is a point spreading function (PSF), and $*$ denotes convolution. $\delta_\Delta(u, v)$ represents the sampling grid, where $\delta(u, v)$ is the Dirac's delta function, and Δ denotes the length of pixels. (ζ, η) represents the sampling offset, and without loss of generality, $\zeta, \eta \in [-\Delta/2, \Delta/2]$ can be assumed. $n_{ob}(u, v)$ is the observation noise. The Fourier transforms of Eqs. (1) and (2) can be described as

$$\hat{y}(\hat{u},\hat{v}) = \{\hat{p}(\hat{u},\hat{v})\hat{x}(\hat{u},\hat{v})\} * \hat{\delta}_\Delta(\hat{u},\hat{v}) + \hat{n}_{ob}(\hat{u},\hat{v}) \tag{3}$$

$$\hat{\delta}_\Delta(\hat{u},\hat{v}) = \frac{4\pi^2}{\Delta^2}\sum_{m,n\in\mathcal{Z}} \delta\left(\hat{u} - \frac{2m\pi}{\Delta}, \hat{v} - \frac{2n\pi}{\Delta}\right) e^{-j(\hat{u}\zeta + \hat{v}\eta)} \tag{4}$$

This research is supported by the Strategic Information and Communication R&D Promotion Programs (SCOPE) of the Ministry of Internal Affairs and Communications, Japan.

$$\overbrace{\begin{pmatrix} \hat{y}_{L,\theta,1}(\hat{u},\hat{v}) \\ \vdots \\ \hat{y}_{L,\theta,k}(\hat{u},\hat{v}) \\ \vdots \\ \hat{y}_{L,\theta,K}(\hat{u},\hat{v}) \end{pmatrix}}^{\hat{\mathbf{y}}_{L,\theta}} = \overbrace{\frac{1}{4}\begin{pmatrix} 1 & e^{-j\pi\zeta_1} & e^{-j\pi\eta_1} & e^{-j\pi(\zeta_1+\eta_1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\pi\zeta_k} & e^{-j\pi\eta_k} & e^{-j\pi(\zeta_k+\eta_k)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\pi\zeta_K} & e^{-j\pi\eta_K} & e^{-j\pi(\zeta_K+\eta_K)} \end{pmatrix}}^{\mathbf{M}} \cdot \overbrace{\mathrm{diag}\begin{pmatrix} g(\hat{u},\hat{v};\theta)\cdot\hat{p}_L(\hat{u},\hat{v})/\hat{p}_H(\hat{u},\hat{v}) \\ g(\hat{u},\hat{v};\theta)\cdot\hat{p}_L(\hat{u}-\pi,\hat{v})/\hat{p}_H(\hat{u}-\pi,\hat{v}) \\ g(\hat{u},\hat{v};\theta)\cdot\hat{p}_L(\hat{u},\hat{v}-\pi)/\hat{p}_H(\hat{u},\hat{v}-\pi) \\ g(\hat{u},\hat{v};\theta)\cdot\hat{p}_L(\hat{u}-\pi,\hat{v}-\pi)/\hat{p}_H(\hat{u}-\pi,\hat{v}-\pi) \end{pmatrix}}^{\mathbf{G}_\theta}$$

$$\cdot \underbrace{\begin{pmatrix} \hat{y}_H(\hat{u},\hat{v}) \\ \hat{y}_H(\hat{u}-\pi,\hat{v}) \\ \hat{y}_H(\hat{u},\hat{v}-\pi) \\ \hat{y}_H(\hat{u}-\pi,\hat{v}-\pi) \end{pmatrix}}_{\hat{\mathbf{y}}_{\mathbf{H}}} + \underbrace{\begin{pmatrix} g(\hat{u},\hat{v};\theta)\,\hat{n}_{ob,1}(\hat{u},\hat{v}) + \hat{n}_{code,1}(\hat{u},\hat{v};\theta) \\ \vdots \\ g(\hat{u},\hat{v};\theta)\,\hat{n}_{ob,k}(\hat{u},\hat{v}) + \hat{n}_{code,k}(\hat{u},\hat{v};\theta) \\ \vdots \\ g(\hat{u},\hat{v};\theta)\,\hat{n}_{ob,K}(\hat{u},\hat{v}) + \hat{n}_{code,K}(\hat{u},\hat{v};\theta) \end{pmatrix}}_{\hat{\mathbf{n}}_\theta}$$

Fig. 2. Details of Eq. (16)

where $\hat{}$ is used to denote the frequency-domain representation of the corresponding symbol. Equation (3) means that the spectrum of $\hat{p}(\hat{u},\hat{v})\hat{x}(\hat{u},\hat{v})$ is replicated in constant intervals with phase shifts, $\exp(-j(\hat{u}\zeta+\hat{v}\eta))$, which depend on the sampling offset (ζ,η).

Let the pixel size of the HR images be 1×1, and assume that the underlying signal $\hat{x}(\hat{u},\hat{v})$ is bandlimited within $[-\pi,\pi]$: i.e. the Nyquist condition is satisfied for this resolution. Thus, the spectrum of a HR image for $\hat{u},\hat{v}\in[-\pi,\pi]$ is described as

$$\hat{y}_H(\hat{u},\hat{v}) = 4\pi^2\hat{p}_H(\hat{u},\hat{v})\hat{x}(\hat{u},\hat{v}) + \hat{n}_{ob}(\hat{u},\hat{v}) \tag{5}$$

where p_H is the PSF of the HR image. Meanwhile, the spectrum of a LR image, whose pixel size is 2×2, is periodic with cycles of (π,π) and can be described in the range $\hat{u},\hat{v}\in[0,\pi]$ as

$$\hat{y}_L(\hat{u},\hat{v}) = \sum_{m,n\in\{0,1\}} \frac{\hat{y}'_H(\hat{u}-m\pi,\hat{v}-n\pi)}{4}\, e^{-j(m\pi\zeta+n\pi\eta)} + \hat{n}_{ob}(\hat{u},\hat{v}) \tag{6}$$

where $\hat{y}'_H(\hat{u},\hat{v})$ $(\hat{u},\hat{v}\in[-\pi,\pi])$ is defined as

$$\hat{y}'_H(\hat{u},\hat{v}) = \{\hat{p}_L(\hat{u},\hat{v})/\hat{p}_H(\hat{u},\hat{v})\}\,\hat{y}_H(\hat{u},\hat{v}) \tag{7}$$

and p_L is the PSF of the LR image. Equation (6) means that $\hat{y}_L(\hat{u},\hat{v})$ consists of four spectral components of $\hat{y}'_H(\hat{u},\hat{v})$, and an additive noise $\hat{n}_{ob}(\hat{u},\hat{v})$, where all the observation noises are put together. This relation can be derived from Eqs. (3)–(5).

2.2. Lossy Image Compression Model

Assume that images are modeled as zero-mean wide-sense stationary Gaussian signals. Based on the rate distortion theory [6], the minimum rate R_H and distortion D_H of the HR image are given by

$$R_H(\theta) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\max\left[0, \frac{1}{2}\log_2\frac{\Phi_{y_H}(\hat{u},\hat{v})}{\theta}\right] d\hat{u}d\hat{v} \tag{8}$$

$$D_H(\theta) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\min[\theta, \Phi_{y_H}(\hat{u},\hat{v})] d\hat{u}d\hat{v} \tag{9}$$

where $\Phi_{y_H}(\hat{u},\hat{v})$ is the power spectral density (PSD) of y_H defined over $\hat{u},\hat{v}\in[-\pi,\pi]$, and θ is a parameter to control the tradeoff between the minimum rate and distortion. Similarly, the minimum rate R_L and distortion D_L of the LR image are given by

$$R_L(\theta) = \frac{1}{\pi^2}\int_{0}^{\pi}\int_{0}^{\pi}\max\left[0, \frac{1}{2}\log_2\frac{\Phi_{y_L}(\hat{u},\hat{v})}{\theta}\right] d\hat{u}d\hat{v} \tag{10}$$

$$D_L(\theta) = \frac{1}{\pi^2}\int_{0}^{\pi}\int_{0}^{\pi}\min[\theta, \Phi_{y_L}(\hat{u},\hat{v})] d\hat{u}d\hat{v}. \tag{11}$$

where $\Phi_{y_L}(\hat{u},\hat{v})$ is the PSD of y_L, whose explicit form is given from Eqs. (6) and (7) as

$$\Phi_{y_L}(\hat{u},\hat{v}) = \sum_{m,n\in\{0,1\}} \frac{\Phi_{y'_H}(\hat{u}-m\pi,\hat{v}-n\pi)}{16} + \Phi_{n_{ob}}(\hat{u},\hat{v}) \tag{12}$$

where $\Phi_{y'_H}(\hat{u},\hat{v})$ and $\Phi_{n_{ob}}(\hat{u},\hat{v})$ are the PSDs of y'_H and n_{ob}. Here, it is assumed that n_{ob} and y_H are independent, and auto-spectral correlations of y_H can be ignored.[1] Note that the integration ranges of Eqs. (10) and (11) are limited to $\hat{u},\hat{v}\in[0,\pi]$ following the domain of definition of $\hat{y}_L(\hat{u},\hat{v})$ given by Eq. (6).

According to the discussion in [8], the LR image after lossy compression, $y_{L,\theta}$, can be described as

$$\hat{y}_{L,\theta}(\hat{u},\hat{v}) = \hat{g}(\hat{u},\hat{v};\theta)\,\hat{y}_L(\hat{u},\hat{v}) + \hat{n}_{code}(\hat{u},\hat{v};\theta) \tag{13}$$

where the gain term, $g(\hat{u},\hat{v};\theta)$, and the PSD of the additive noise, $\hat{n}_{code}(\hat{u},\hat{v};\theta)$, are given by

$$\hat{g}(\hat{u},\hat{v};\theta) = \max[\,0,\ 1-\theta/\Phi_{y_L}(\hat{u},\hat{v})\,] \tag{14}$$

$$\Phi_{n_{code}}(\hat{u},\hat{v};\theta) = \max[\,0,\ \theta\,(1-\theta/\Phi_{y_L}(\hat{u},\hat{v}))\,]. \tag{15}$$

2.3. Super-Resolution Reconstruction

Assume that multiple LR images with the same compression quality are available at the receiver side. The goal of SR decoding is to recover the underlying HR image from these LR images.

Let k and K be the index and number of the LR images, respectively. Using Eqs. (6) and (13), we have K equations:

$$\hat{\mathbf{y}}_{L,\theta} = \mathbf{M}\,\mathbf{G}_\theta\,\hat{\mathbf{y}}_{\mathbf{H}} + \hat{\mathbf{n}}_\theta \quad (\hat{u},\hat{v}\in[0,\pi]) \tag{16}$$

whose explicit form is given in Fig. 2. The sampling offset (ζ_k,η_k), observation noise $n_{ob,k}$, and coding noise $n_{code,k}$ are denoted with a subscript k, because they depend on k. We assume the sampling offset of each image is known, but the noise terms are unknown except their statistical properties like PSDs. Thereby, in the above equation, we know $\hat{\mathbf{y}}_{L,\theta}$, $\mathbf{M}$, and $\mathbf{G}_\theta$. We also know the statistical property of $\hat{\mathbf{n}}_\theta$. The unknown to estimate is $\hat{\mathbf{y}}_{\mathbf{H}}$, the underlying HR image.

Estimation of $\hat{\mathbf{y}}_{\mathbf{H}}$ can be achieved by deconvolution of Eq. (16) with Tikhonov regularization as

$$\tilde{\hat{\mathbf{y}}}_{H,\theta} = \mathbf{F}_\theta\hat{\mathbf{y}}_{L,\theta},\quad \mathbf{F}_\theta = (\mathbf{G}_\theta^*\mathbf{M}^*\mathbf{M}\mathbf{G}_\theta + \lambda\mathbf{I})^{-1}\,\mathbf{G}_\theta^*\mathbf{M}^* \tag{17}$$

[1] The latter assumption is justified by the fact that Karhunen-Loeve transform converges to Fourier transform under some condition [7], thereby, the spectral components are nearly uncorrelated.

where $\mathbf{F}_\theta$ is a reconstruction filter, λ is a positive constant to avoid singularity, $\mathbf{I}$ is the identity matrix, and * denotes the conjugate transpose. The estimation error can be represented as a form of the covariance matrix

$$\begin{aligned}\mathbf{C}_{\theta,\hat{u},\hat{v}} &= E[(\tilde{\mathbf{y}}_{H,\theta} - \hat{\mathbf{y}}_H)(\tilde{\mathbf{y}}_{H,\theta} - \hat{\mathbf{y}}_H)^*] \\ &= (\mathbf{I}-\mathbf{F}_\theta\mathbf{M}\mathbf{G}_\theta)E[\hat{\mathbf{y}}_\mathbf{H}\hat{\mathbf{y}}_\mathbf{H}^*](\mathbf{I}-\mathbf{F}_\theta\mathbf{M}\mathbf{G}_\theta)^* + \mathbf{F}_\theta E[\hat{\mathbf{n}}\hat{\mathbf{n}}^*]\mathbf{F}_\theta^* \quad (18)\end{aligned}$$

where $E[\;]$ denotes the expectation. We assume that y_H, $n_{ob,k}$ and $n_{code,k}$ are independent and auto-spectral correlations of $\hat{y}_H$ can be ignored, so that non-diagonal elements of $E[\hat{\mathbf{y}}_\mathbf{H}\hat{\mathbf{y}}_\mathbf{H}^*]$ and $E[\hat{\mathbf{n}}\hat{\mathbf{n}}^*]$ are zero, which eases the calculation.

As shown by Eq. (17), the reconstruction filter $\mathbf{F}_\theta$ actually depends on the the compression quality θ. However, the compression quality is rarely considered in conducting SR reconstruction. To simulate this condition, we use $\mathbf{F}_0$ (meaning $\theta = 0$, corresponding to the reconstruction filter for non-compressed images) in Eqs. (17) and (18) regardless of the compression quality.

Given the error model of Eq. (18), we obtain the rate-distortion model for SR decoding. The minimum rate $R_{SR}(\theta)$ is

$$R_{SR}(\theta) = R_L(\theta)/4. \quad (19)$$

because the number of pixels becomes four times by the SR reconstruction. The minimum distortion $D_{SR}(\theta)$ is

$$D_{SR}(\theta) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} D_{SR}(\hat{u},\hat{v};\theta)d\hat{u}d\hat{v}$$

$$D_{SR}(\hat{u},\hat{v};\theta) = \left\{\begin{array}{lll}\mathbf{C}_{\theta,\hat{u}\quad,\hat{v}} & (1,1) & \hat{u} \geq 0, \hat{v} \geq 0 \\ \mathbf{C}_{\theta,\hat{u}+\pi,\hat{v}} & (2,2) & \hat{u} < 0, \hat{v} \geq 0 \\ \mathbf{C}_{\theta,\hat{u}\quad,\hat{v}+\pi} & (3,3) & \hat{u} \geq 0, \hat{v} < 0 \\ \mathbf{C}_{\theta,\hat{u}+\pi,\hat{v}+\pi} & (4,4) & \hat{u} < 0, \hat{v} < 0\end{array}\right. \quad (20)$$

where $\mathbf{C}_{\theta,\hat{u},\hat{v}}(1,1)$ denotes the $(1,1)$ element of the matrix $\mathbf{C}_{\theta,\hat{u},\hat{v}}$

2.4. Numerical Analysis

The purpose of the numerical analysis is to compare the theoretical rate-distortion (R-D) performances between the HR coding and LR coding with SR decoding. The R-D performance for the former case is given by $R_H(\theta)$ and $D_H(\theta)$ in Eqs. (10) and (11), and for the latter case, by $R_{SR}(\theta)$ and $D_{SR}(\theta)$ in Eqs. (19) and (20). By changing the value of θ, we can draw R-D curves.

As a widely used model of natural images [8],

$$\Phi_{y_H}(\hat{u},\hat{v}) = \frac{2\pi}{\omega^2}\left(1 + \frac{\hat{u}^2+\hat{v}^2}{\omega^2}\right)^{-\frac{3}{2}} \quad (\omega = -\ln(\rho)) \quad (21)$$

is adopted, where ρ denotes the correlation between adjacent pixels, which was set to 0.90. As the point-spreading function, we adopted $p_H(u,v) = \text{box}(u,v;1)$ and $p_L(u,v) = \text{box}(u,v;2)$ where

$$\text{box}(u,v;L) = \left\{\begin{array}{ll} 1/L^2 & u,v \in [-L/2, L/2] \\ 0 & \text{otherwise}\end{array}\right. \quad (22)$$

whose spatial supports correspond to the pixel sizes of the HR and LR images (1×1 and 2×2) , respectively. The observation noise is assumed to be white, denoted as $\Phi_{n_{ob}}(\hat{u},\hat{v}) = \sigma^2_{n_{ob}}$. The sampling offsets (ζ_k, η_k) were generated as uniform random values in $[-1,1]$. λ for Tikhonov regularization in Eq. (17) was set to 0.01.

Figure 3 shows the results of analysis. In the top graph, K (number of images used for SR decoding) was varied while $\sigma^2_{n_{ob}}$ (magnitude of observation noise) was fixed to 0. Meanwhile, in the bottom graph, $\sigma^2_{n_{ob}}$ was varied while K was fixed to 10. Both of the graphs show that SR decoding exhibits its advantage in low bitrates. However, in higher bitrates, SR decoding reaches the ceiling and is overtaken by the HR coding. The maximum quality achieved by SR decoding depends on the values of K and $\sigma_{n_{ob}}$; the larger K and the smaller $\sigma_{n_{ob}}$ have positive effects in the resulting quality.

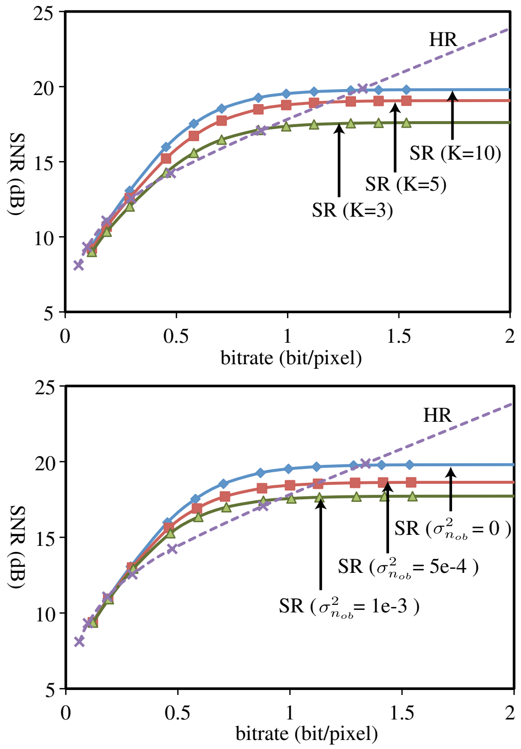

Fig. 3. Numerical analysis of R-D performance: (top) K is varied, $\sigma^2_{n_{ob}} = 0$, (bottom) $\sigma^2_{n_{ob}}$ is varied, $K = 10$.

3. REAL IMAGE SIMULATION

We conducted a controlled real-image simulation to confirm the numerical model presented in Section 2.

A grayscale image with very high resolution (4200×2800 pixels) shown in Fig. 4 was used as the source signal $x(u,v)$, from which HR and LR images, $y_{H,k}(u,v)$ and $y_{L,k}(u,v)$, whose pixel sizes are 10×10 and 20×20 pixels, respectively, were generated as:

$$\begin{aligned} y_{H,k}(u,v) &= (x * p_H)(10u + \zeta_k, 10v + \eta_k) \quad (23) \\ y_{L,k}(u,v) &= (x * p_L)(20u + \zeta_k, 20v + \eta_k) + n_{ob}(u,v) \quad (24)\end{aligned}$$

where k is the index, the sampling offsets $(\zeta_k, \eta_k) \in \mathcal{Z}^2$ $(k = 1, ...K)$ were generated in random, and p_H and p_L are the box-shaped PSFs whose supports correspond to the pixel sizes. Observation noise is Gaussian and added only to the LR images, because we focus on the relative observation difference between the HR and LR images. We compressed those images using *cjpeg* codec with a flat quantization matrix. The compressed images with quality factor Q are denoted as $y_{H,Q,k}(u,v)$ and $y_{L,Q,k}(u,v)$.

We then conduct SR reconstruction to estimate each HR image $\tilde{y}_{H,Q,k}(u,v)$ $(k = 1,\ldots,K)$ using all LR images compressed with quality factor Q by minimizing an energy function E_Q as

$$\begin{aligned}\tilde{y}_{H,Q,k}(u,v) &= \text{argmin}\; E_Q(y_{H,k}(u,v)) \quad (25) \\ E_Q(y_{H,k}(u,v)) &= \sum_{k'} \|y_{L,Q,k'}(u,v) - \mathbf{P}_{\mathbf{k}\rightarrow\mathbf{k'}}(y_{H,k}(u,v))\|^2 \\ &\quad + \lambda\|y_{H,k}(u,v) - \text{mean}(y_{H,k}(u,v))\|^2 \quad (26)\end{aligned}$$

where $\mathbf{P}_{\mathbf{k}\rightarrow\mathbf{k'}}$ denotes the geometrical mapping from the k-th latent HR image $y_{H,k}(u,v)$ to the k'-th LR image $y_{L,k'}(u,v)$. This mapping can be obtained from the shape of the PSFs p_H and p_L, and

Fig. 4. Input image used for the experiment.

the known sampling offsets (ζ_k, η_k). The regularization parameter λ was set to 0.1 according to empirical tests. Note that SR reconstruction of our real-image simulations is performed in the spatial domain, but the mathematical model is essentially equivalent with the frequency-domain deconvolution represented as Eq. (17).

Finally, the rates and distortions for the two coding scenarios are measured by the averages over K images as

$$R_H(Q) = \sum_k \text{bitrate}(y_{H,Q,k}(u,v))/K \quad (27)$$

$$D_H(Q) = \sum_k \text{mean}\left(\|y_{H,Q,k}(u,v) - y_{H,k}(u,v)\|^2\right)/K \quad (28)$$

$$R_{SR}(Q) = \sum_k \text{bitrate}(y_{L,Q,k}(u,v))/(4K) \quad (29)$$

$$D_{SR}(Q) = \sum_k \text{mean}\left(\|\tilde{y}_{H,Q,k}(u,v) - y_{H,k}(u,v)\|^2\right)/K \quad (30)$$

where the bitrates are obtained from the filesizes and the number of pixels. Plotting $(R(Q), D(Q))$ with different quality factors Q yields a rate-distortion curve.

The top graph in Fig. 5 shows the results with $K = 5$ and 10. Performance of bicubic upsampling (each LR image was independently upsampled) is also shown for reference. In the bottom graph, K was fixed to 10, while $\sigma_{n_{ob}}$ (the standard deviation of the observation noise) was varied. The overall tendency is similar to Fig. 3 which was derived from the numerical model analysis.

4. DISCUSSION AND CONCLUSION

The rate-distortion performance of super-resolution (SR) decoding was analyzed in this paper. We first presented a numerical model that combines a frequency-domain SR model and a rate-distortion theory to derive the theoretical performance of SR decoding. We also conducted real-image simulations to compare the results with the theory. We showed that both of them exhibit quite similar tendencies: SR decoding performs better in low bitrates, but reaches the ceiling in higher bitrates. Although our analysis was limited to an idealized configuration, we believe our approach opens a new vista to deductive analysis of SR decoding.

The future work should be focused on the generalization of the numerical model, because we ignored many practical factors for simplification: inter-frame prediction coding, non-global and non-translational registrations, registration errors, etc. We are also interested in the joint optimization of lossy compression and SR reconstruction, in which both algorithms might be tuned and modified to improve the overall performance. Furthermore, faster and more reliable implementation of SR reconstruction is an important issue for SR decoding to be applied to communication systems.

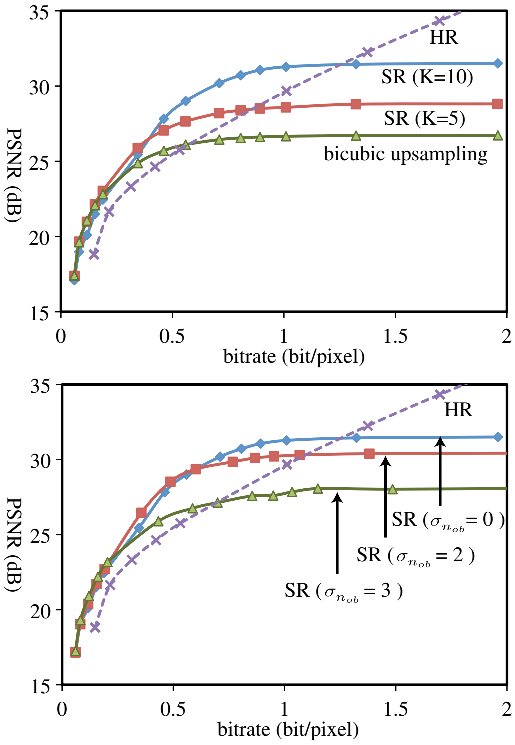

Fig. 5. Real image simulation of R-D performance: (top) K was varied, $\sigma_{n_{ob}} = 0$, (bottom) $\sigma_{n_{ob}}$ was varied, $K = 5$.

5. REFERENCES

[1] S.-C. Park, M.-K. Park, and M.-G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] G.M. Callico, A. Nunez, R.P. Llopis, R. Sethuraman, and M.O. de Beeck, "A low-cost implementation of super-resolution based on a video encoder," *Annual Conf. of the Industrial Elec. Society*, vol. 2, pp. 1439–1444, 2002.

[3] R. Molina, A.K. Katsaggelos, L.D. Alvarez, and J. Mateos, "Towards a new video compression scheme using super-resolution," *SPIE-IS&T Electronic Imaging, Visual Communications and Image Processing 2006*, vol. 6077, 2006.

[4] S. Ma, L. Zhang, X. Zhang, and Wen Gao, "Block adaptive super resolution video coding," *Advances in Multimedia Information Processing - PCM2009*, pp. 1048–1057, 2009.

[5] R.Y. Tsai and T.S. Huang, "Multipleframe image restoration and registration," *Advances in Computer Vision and Image Processing*, pp. 317–339, 1984.

[6] T. Berger, "Rate distortion theory," *Englewood Cliffs, NJ: Prentice-Hall*, 1971.

[7] U. Grenander and G. Szego, "Toeplitz forms and their applications," *Berkeley, Calif.: University of California Press*, 1958.

[8] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequence," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 7, pp. 1140–1154, 1987.