

SUPER-RESOLUTION USING MULTIPLE QUANTIZED IMAGES

Ayça Özçelikkale,^{*1} Gözde B. Akar,² Haldun M. Ozaktas^{1†}

¹Dep. of Electrical Eng.
Bilkent Univ., Ankara, Turkey
e-mail: ayca, haldun@ee.bilkent.edu.tr

²Dep. of Electrical Eng.
Middle East Technical Univ., Ankara, Turkey
e-mail: bozdagi@metu.edu.tr

ABSTRACT

In this paper, we study the effect of limited amplitude resolution (pixel depth) in super-resolution problem. The problem we address differs from the standard super-resolution problem in that amplitude resolution is considered as important as spatial resolution. We study the trade-off between the pixel depth and spatial resolution of low resolution (LR) images in order to obtain the best visual quality in the reconstructed high resolution (HR) image. The proposed framework reveals great flexibility in terms of pixel depth and number of LR images in super-resolution problem, and demonstrates that it is possible to obtain target visual qualities with different measurement scenarios including images with different amplitude and spatial resolutions.

Index Terms— super-resolution, quantization, amplitude resolution, pixel depth.

1. INTRODUCTION

In this paper, we study the effect of limited amplitude resolution in super-resolution problem, where multiple images with poor spatial resolution are used to reconstruct an image of the same scene with higher spatial resolution [1]. We note that in the standard super-resolution problems, researchers mostly focus on increasing resolution in space, whereas in our study both resolution in space and resolution in amplitude are substantial parameters of the framework.

Many applications in image processing will benefit from such a study including converting available low resolution content to high definition television (HDTV). This subject is not merely of interest for practical purposes but can also lead to a better understanding of the effect of pixel depth in super-resolution problem. We are concerned with questions such as "To obtain a target resolution, which is better, a high number of coarsely quantized images or a low number of densely quantized images?" or "What is the range of admissible pixel depths at a particular spatial resolution to obtain an image with a target spatial resolution with a target visual quality?". Admitting great flexibility in terms of number and accuracies of the LR images, our framework is similar to other constrained signal acquisition scenarios such as compressed sensing paradigm.

The measurement framework this research is based on was first proposed in [2, 3]. Here we study an application of this approach to super-resolution problem. Although super-resolution is a popular problem with a wide range of applications, no studies exist in

literature which address super-resolution in multiple domains, i.e. amplitude and space.

We emphasize that since both resolution in space and resolution in amplitude are variables, the term *low/high resolution* image is, in fact, ambiguous in our framework. Nevertheless, we use them to refer to images with *low/high spatial resolution* to be consistent with the literature.

2. MEASUREMENT MODEL

We consider the following image acquisition scenario where L low resolution images are obtained from a high resolution image \mathbf{x} according to the model:

$$\mathbf{y}_k = D_k H_k F_k \mathbf{x} + \mathbf{v}_k, \quad k = 1, \dots, L \quad (1)$$

where \mathbf{y}_k 's are LR images, \mathbf{v}_k 's denote the system noise, D_k represents the decimation operator, H_k represents the camera blur, F_k represents the motion operator, L is the number of available LR images. \mathbf{v}_k 's are independent of each other, and the components of each \mathbf{v}_k are i.i.d. All images are rearranged in lexicographic order. Here \mathbf{x} is of size $N_1 N_2$, and \mathbf{y}_k 's are of size $\bar{N}_1 \bar{N}_2$, where $N_1 = r_1 \bar{N}_1$, and $N_2 = r_2 \bar{N}_2$.

We assume that we only have access to quantized LR images;

$$\mathbf{y}_k^{b_{y_k}} = Q_{b_{y_k}}(\mathbf{y}_k), \quad k = 1, \dots, L \quad (2)$$

where $Q_{b_{y_k}}$ is the uniform quantizer with $2^{b_{y_k}}$ levels. In general, b_{y_k} may be different for different LR images. Here, for simplicity, we assume that all LR images are quantized with the same number of bits, i.e. $b_{y_k} = b_y$.

Our aim is to reconstruct the HR image from the observation of quantized LR images. We are interested in the trade-offs between resolution in amplitude and resolution in space.

We describe the spatial resolution of each LR image y_k relative to the spatial resolution of target high resolution image $\hat{\mathbf{x}}$, and it is given by $1/(r_1 r_2)$. The number of LR images may be thought as a part of spatial resolution, as well as a parameter associated with resolution in time when considered in a spatio-temporal framework. The resolution in amplitude associated with an image I is described by the number of bits used to represent pixel values b_I , which is the pixel depth.

We associate a cost with a particular representation of a scene: cost of a quantized image is given by the total number of bits needed to represent this particular representation, i.e. number of pixels in the image \times number of bits used to represent each pixel value. For example the representation cost of the HR image \mathbf{x} is $C_{\mathbf{x}} = N_1 \times N_2 \times b_{\mathbf{x}}$, and similarly the representation cost of a LR image $\mathbf{y}_k^{b_y}$

^{*}A. Özçelikkale was supported by TÜBİTAK Doctoral Scholarship.

[†]Haldun M. Ozaktas was supported in part by the Turkish Academy of Sciences



Fig. 1: Some of the images used in the experiments.

whose pixel values are quantized with b_y bits is $C_{\mathbf{y}_k^{b_y}} = \bar{N}_1 \times \bar{N}_2 \times b_y$. The total representation cost of L low resolution images is $L \times C_{\mathbf{y}_k^{b_y}}$.

The cost parameter provides a way of expressing the combined effect of the resolution in space, resolution in amplitude, and number of LR images for a given image acquisition scenario (given set of LR images) with a single number. We note that the actual number of bits needed to effectively store or transmit the images may be quite different from C . Our notion of cost should be considered as a part of acquisition rather than the coding of information.

The ratio of the total representation cost of L low resolution images to the representation cost of the target HR image $\hat{\mathbf{x}}$ is a useful parameter and is given by

$$C_r = \frac{L \times \bar{N}_1 \times \bar{N}_2 \times b_y}{N_1 \times N_2 \times b_{\hat{\mathbf{x}}}} = \frac{L \times b_y}{r_1 \times r_2 \times b_{\hat{\mathbf{x}}}}. \quad (3)$$

C may be seen as a measure of information in a particular representation of scene. Hence it may be argued that if $C_r < 1$, there is not as much as information in the LR images as in the target HR image, and the problem is underdetermined in the sense of number of bits available. In a typical image, the values of different pixels are neither independent, nor necessarily identically and uniformly distributed. Yet C provides an upper bound, and still may be useful in interpretation of the results. We finally note that in a typical super-resolution problem effective bit depths of the HR image, and the LR images and achievable bit depths for the target HR image may take different but related values, which puts constraints on the values C_r can take.

3. METHODOLOGY

To study the trade-off between amplitude resolution and spatial resolution within the given framework, we will consider different image acquisition scenarios and compare their success in generating HR images with a particular super-resolution method.

As super-resolution method, we use the norm approximation method recently proposed in [4]. We note that one could use other image reconstruction methods as well. Although the specifics of these methods may differ, we believe that the nature of the trade-offs observed and the general conclusions and insights that will be presented in this paper will remain useful with a wide variety of methods. In [4], the reconstructed image $\hat{\mathbf{x}}$ is given as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \sum_{k=1}^L \|\mathbf{y}_k - D_k H_k F_k \mathbf{x}\|_1 + \lambda \sum_{l=-P}^P \sum_{m=-P}^P \alpha^{|m|+|l|} \|\mathbf{x} - S_h^m S_v^l \mathbf{x}\|_1 \right\}, \quad (4)$$

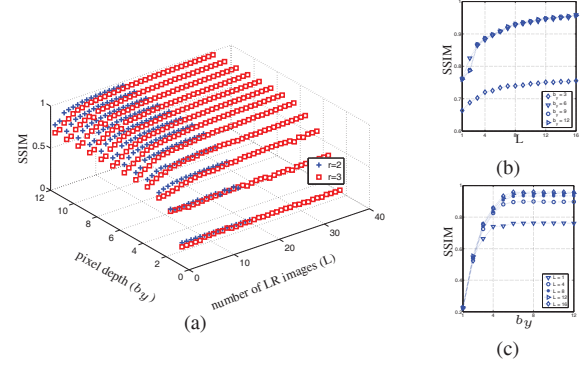


Fig. 2: SSIM for different image acquisition set-ups, HR image is used to select λ . (a) SSIM vs the number (L) and pixel depth (b_y) of LR images, upsampling factor r variable (b) SSIM vs L for $r = 2$ with varying b_y (c) SSIM vs b_y for $r = 2$ with varying L .

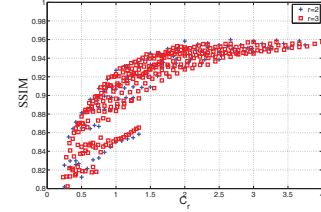


Fig. 3: SSIM vs C_r , upsampling factor r variable, HR image is used to select λ .

where operators S_h^m and S_v^l shift \mathbf{x} by m and l pixels in the horizontal and vertical directions, respectively. We have used $\alpha = 0.6$, and $P = 2$, which are one of the typical values used in [4]. Here $\lambda > 0$ is a scalar parameter used to control the amount of regularization. The method used to determine λ is explained in each experiment.

Structural similarity (SSIM) index [5] and peak signal to noise ratio (PSNR) are used as the quality metrics to report the success of different image acquisition scenarios. SSIM index between two images $\hat{\mathbf{x}}$ and \mathbf{x} are given as the mean of SSIM over aligned image patches, where the SSIM between image patches from $\hat{\mathbf{x}}$ and \mathbf{x} is given as

$$\text{SSIM} = \frac{(2 \mu_{\mathbf{x}} \mu_{\hat{\mathbf{x}}} + C_1) (2 \sigma_{\mathbf{x}\hat{\mathbf{x}}} + C_2)}{(\mu_{\hat{\mathbf{x}}}^2 + \mu_{\mathbf{x}}^2 + C_1) (\sigma_{\hat{\mathbf{x}}}^2 + \sigma_{\mathbf{x}}^2 + C_2)}. \quad (5)$$

Here $\mu_{\mathbf{x}}$, $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{x}\hat{\mathbf{x}}}$ denote the local estimates of the mean, variance and cross correlation respectively. We have used the implementation offered by [5], and reported SSIM over a dynamic range of 1 using C_1 and C_2 as $(0.01)^2$ and $(0.03)^2$ in accordance with [5].

Finally we give some of the parameters used in the experiments: The upsampling factors in two dimensions are assumed to be the same, i.e. $r_1 = r_2 = r$. Camera point spread function (p.s.f.) is assumed to be 3×3 Gaussian filter. Gaussian noise with a standard deviation of 0.02 is used to simulate the system noise. Camera p.s.f. and motion vectors are assumed to be known in the reconstruction.

4. EXPERIMENTAL RESULTS

We will now study the relationship between resolution in amplitude and resolution in space in super-resolution scenarios by ex-

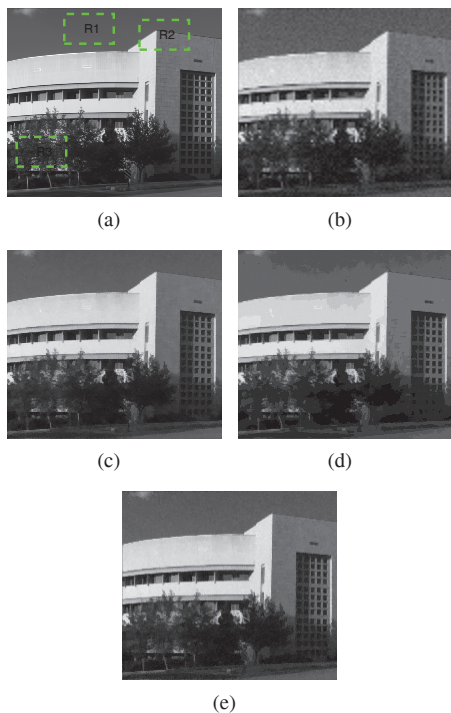


Fig. 4: (a) HR image, (b) bi-cubic interpolation of 1 LR image with 12 bit quantization, Images reconstructed from (c) 6 LR images with 8 bit quantization (P_1) (d) 12 LR images with 4 bit quantization (P_2) (e) 4 LR images with 12 bit quantization (P_3).

aming the success of different image acquisition set-ups. This study will also reveal the trade-off between the quality (SSIM of the reconstructed images) and cost (the representation costs of LR images) under the experiment parameters used. We use $C_r = (L \times b_y)/(r^2 \times b_x)$.

Exp. 1: This experiment investigates the case where HR image is assumed to be known in the reconstruction process and optimum λ to obtain the best SSIM is searched heuristically. This experiment serves the purpose of providing a benchmark for the best performance possible with the reconstruction method used. For this experiment the 12-bit grayscale image, shown in Fig. 1(a) is used. This image includes a fair amount of edges as well as textured, and smooth regions. We consider the image acquisition strategies with pixel depths $b_y \in \{1, \dots, 12\}$ and the number of LR images $L \in \{1, \dots, 4r^2\}$ with upsampling factors $r = 2, 3$.

Fig. 2, and Fig. 3 gives the SSIM for different image acquisition scenarios and the associated trade-off between SSIM and C_r , respectively. We see that it is possible to obtain a given SSIM performance with different image acquisition strategies, and possibly different costs. In Fig. 3, the boundary of the achievable SSIM- C_r region shows that SSIM is very sensitive to increases in C_r for smaller values of C_r . Then it becomes less responsive, and eventually saturates at an asymptote for high values of C_r . We also note that in all of the measurement scenarios considered in this experiment, for a given pixel depth, if the total number of pixels available are the same for varying upsampling factors, SSIM values turn out to be very close. This also shows that under the image acquisition set-ups considered in this experiment, resolution in amplitude, not

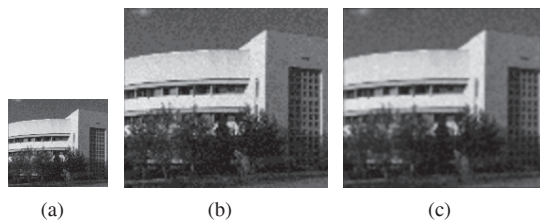


Fig. 5: (a) LR image with 4 bit quantization ($r = 2$) (b) bi-cubic interpolation (c) after noise removal

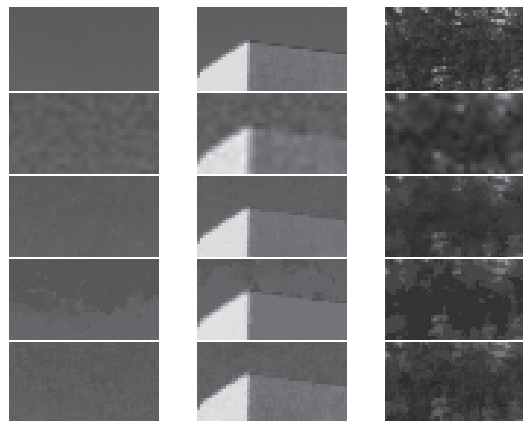


Fig. 6: Region 1(Left), Region 2 (Middle), Region 3 (Right): Patches from the images presented in Fig. 4

resolution in space (upsampling factor), is the key factor determining the quality of reconstructed images. This trend is strongly related to the size of camera p.s.f., the size of details in the images as well as the upsampling factors used in the experiment.

We observe that in general for a given pixel depth, SSIM increases as the number of available LR images increases (see for instance Fig. 2(b)). We also see that for a given number of available LR images, SSIM increases with increasing pixel depth (see for instance Fig. 2(c)). For low values of pixel depth, the information lost due to poor resolution in amplitude can be hardly recovered by acquiring more LR images, resulting in very close SSIM values for all values of L . The increase in SSIM with increasing L is lower for low values of pixel depth compared to high values. As pixel depth increases the number of available images becomes more important in determining the SSIM level that can be reached with a particular pixel depth. However for all values of pixel depth, the increase in SSIM with increasing L gradually becomes lower as L increases.

We now take a closer look on the following data points with $r = 2$: 6 LR images with 8-bit pixel depth (P_1), 12 LR images with 4-bit pixel depth (P_2), and 4 LR images with 12-bit pixel depth (P_3). The costs of these acquisition schemes are the same, so it is reasonable to use them to compare the following different sampling strategies: a high number of images with a coarse resolution in amplitude (P_2), a low number of images with a dense resolution in amplitude (P_3), and the strategy in between (P_1).

The actual HR image, and reconstructed images for P_1 , P_2 and P_3 are shown in Fig. 4(a), Fig. 4(c), Fig. 4(d), and Fig. 4(e) respectively. The regions indicated in Fig. 4(a) are shown in Fig. 6 with the corresponding SSIM and PSNR values in Table 1.

Table 1: SSIM and PSNR (dB) values for the image patches extracted from the image shown in Fig. 4(a) with different image acquisition scenarios corresponding to P_1 , P_2 , and P_3

	P_1	P_2	P_3
image	0.9135- 31.30	0.8540 - 29.33	0.8904 - 29.95
region 1	0.9629 - 43.48	0.8712 - 32.95	0.9300 - 40.88
region 2	0.9340 - 37.23	0.9015 - 33.14	0.9187 - 36.40
region 3	0.7879 - 27.86	0.7668 - 27.98	0.7610 - 27.19

We observe that there are quantization artifacts all over the image reconstructed from the set-up in P_2 (Fig. 4(d)). Some image details on textured regions are lost, and there are fake borders in smooth regions, which are particularly apparent in the sky region and on the building. After the noise removal, the low pixel depth of LR images causes banding in these regions, in which there is actually a smooth gray level transition. We note that these boundary effects are a result of successful noise removal. To illustrate this point, the LR image and bi-cubic interpolation of one LR image is shown in Fig. 5. We observe that with this naive approach the noise removal smoothes the edges and results in a blurred image. For P_3 (Fig. 4(e)), we observe that although most of the image details are successfully reconstructed, the image is noisy. In this case the number of available LR images is relatively low, hence they may not be sufficient to successfully remove noise without blurring. The noisy behaviour of the image suggests that using such a high pixel depth is a waste of resources, since the image pixels are already corrupted with a noise whose level is much higher than the quantization interval, and these bits could have been used to acquire more LR images. We note that by adjusting the parameter λ , it may be possible to obtain a smoother but blurred image. We also note that if the system noise had been lower, the number of LR images at hand could have been sufficient to construct a less noisy image without blur. Finally, Fig. 4(c) (P_1) presents the image reconstructed from the 6 images with 8-bit pixel depth. Among the three measurement strategies, this strategy is the one that gets the highest scores from both of the quality metrics, SSIM and PSNR. We see that there is still some noise in this image, but there are no quantization artifacts similar to the ones present in Fig. 4(d).

Exp. 2: In this experiment, we investigate the trade-off when another image with similar characteristics is used to select λ values: The image patch shown in Fig. 1(b) which is extracted from an outdoor image is used to learn the optimum λ for different image acquisition schemes. We run the experiments for the first 20 8-bit images in scene categories “CALsuburb” and “MITinsidicity” from the database introduced in [6] (examples shown in Fig. 1(c)) and report the mean SSIM values across each image category. We consider the image acquisition strategies with pixel depths $b_y \in \{1, \dots, 8\}$ and the number of LR images $L \in \{1, r^2, 2r^2, 3r^2, 4r^2\}$ with upsampling factors $r = 2, 3$.

Fig. 7 shows the trade off between SSIM and C_r . We observe that the nature of these plots are similar to the trade-off curve presented in Fig. 3, in which HR image is used to select the best λ is to obtain the best performance. The SSIM values that may be reached with the image acquisition scenarios under consideration does not change significantly. We may conclude that it is possible to reach the benchmark’s performance without knowing the HR image in advance, which is the case for a typical super-resolution application.

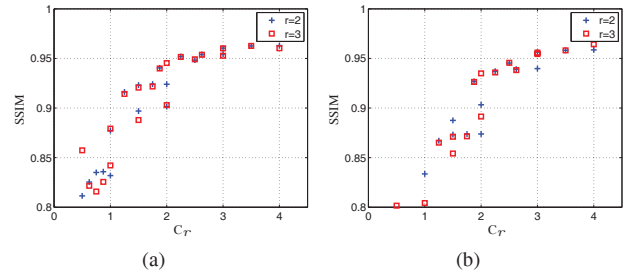


Fig. 7: SSIM vs C_r : upsampling factor variable, image patch shown in Fig. 1(b) is used to select λ . (a) database: “CALsuburban” (b) database: “MITinsidicity”

5. CONCLUSIONS

We have studied on understanding the relationship between resolution in amplitude and resolution in space in super-resolution problem. Unlike most previous work, amplitude resolution was considered as important part of the super-resolution problem as spatial resolution. We have studied the success of different measurement set-ups where the resolution in amplitude (pixel depth), resolution in space (upsampling factor) and the number of LR images are variable. Our study has revealed great flexibility in terms of spatial-amplitude resolutions in super-resolution problem. We have seen that it is possible to reach target visual qualities with different measurement scenarios including varying number of images with different amplitude and spatial resolutions. Our results illustrate how coarsely the images with low spatial resolution could be quantized in order to obtain images with high spatial resolution with good visual qualities. We believe that there is a great deal of exciting work to be done to understand the relationship between resolution in amplitude and resolution in space in super-resolution problem. An interesting example is finding all the achievable visual qualities and associated upsampling factors and amplitude resolutions under a given cost. This point is left as future work.

6. REFERENCES

- [1] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” *Signal Processing Magazine, IEEE*, vol. 20, pp. 21 – 36, May 2003.
- [2] A. Özçelikkale, “Structural and metrical information in linear systems,” Master’s thesis, Bilkent Univ., Ankara, Turkey, 2006.
- [3] A. Özçelikkale, H. M. Ozaktas, and E. Arıkan, “Optimal measurement under cost constraints for estimation of propagating wave fields,” in *Proc. 2007 IEEE Int. Symp. Information Theory*, pp. 696–700.
- [4] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, “Fast and robust multiframe super resolution,” *IEEE Trans. Image Process.*, vol. 13, pp. 1327 – 1344, Oct. 2004.
- [5] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, Apr. 2004.
- [6] L. Fei-Fei and P. Perona, “A Bayesian hierarchical model for learning natural scene categories,” *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 524 – 531.