SUPER-RESOLUTION USING MOTION AND DEFOCUS CUES

K.V. Suresh

Department of Electronics and Communication,
Siddaganga Institute of Technology
Tumkur, Karnataka, India

A. N. Rajagopalan

Department of Electrical Engineering
Indian Institute of Technology, Madras
Chennai, Tamilnadu, India

ABSTRACT

Reconstruction-based super-resolution algorithms use either sub-pixel shifts or relative blur among low-resolution observations as a cue to obtain a high-resolution image. In this paper, we propose a super-resolution algorithm that exploits the information available in the low-resolution observations due to both sub-pixel shifts and relative blur to yield a better quality image. Performance analysis is carried out based on the Cramér-Rao lower bound. Several experimental results on synthetic and real images are given for validation.

Index Terms— Super-resolution, image enhancement, image resolution, image reconstruction, image sequence analysis

1. INTRODUCTION

Super-resolution is the method of obtaining a high-resolution (HR) image from several low-resolution (LR) observations of the same scene. The main idea behind super-resolution is dealiasing and deblurring. One class called motion-based techniques [1, 2, 3] assumes that there is a relative motion between the scene and the camera. Multiple, sub-pixel shifted LR images thus captured carry additional information about the scene which is then used to obtain a high-quality image. The quality of reconstructed image depends on the relative shifts among the observations. A second class called motion-free super-resolution [4, 5], on the other hand, make use of the information available from several defocused observations (blurred to different extents) to undo the effects of blurring and aliasing. They assume that there is no relative motion among the observations thus doing away with the correspondence problem. Here, the quality of the super-resolved image depends on the relative blur among the observations [4].

Existing motion super-resolution algorithms assume evenly spaced motion vectors [6] such that the LR images adequately sample the high-resolution image. Further, an adequate number of LR observations [7] are assumed to be available for reconstruction. In addition, they assume that accurate motion parameters are available. Blurring (if any) introduced either due to motion or camera defocusing is assumed to be same for all the observations and is usually treated as unwanted. A recent work [8] concluded that blur introduced due to camera motion limits super-resolution quality.

In a real scenario, motion super-resolution is confronted with the problem that the movement of either the camera or the object is not controlled. Hence, the assumption of availability of adequate number of evenly spaced observations is not realistic. Also, depending on the distance of the camera from the scene, the sub-pixel shifted observations may suffer from different amounts of blur. Knowing the fact that relative blur plays a role in motion-free super-resolution, it is then interesting to ask whether relative blurring can be used to our advantage in motion super-resolution also. Intuitively, one can say that relative blur among the observations could be used as a cue for better reconstruction, particularly when the motion shifts are not good and also for achieving higher resolution factors from less number of observations. In this paper, we show that this is indeed true. The Cramér-Rao lower bound (CRLB) for the covariance of the error in the estimate of the super-resolved image is derived to theoretically validate our argument. The image to be super-resolved is modeled as a Markov random field (MRF) and the maximum *a posteriori* (MAP) estimate of the image is derived. Several simulations on 1D sequence and images (both synthetic and real) are given to demonstrate the role of relative blur in the motion super-resolution problem.

2. PROBLEM FORMULATION

The relation between the lexicographically ordered LR observations and the original HR image can be expressed as

$$\underline{y_r} = DH_rW_r\underline{x} + \underline{n_r}, \qquad 1 \leq r \leq m \qquad (1)$$

Here, $\underline{x}$ is the original HR image of dimension N_1N_2 x 1, $\underline{y_r}$ is the r^{th} LR observation of dimension M_1M_2 x 1, D is the down-sampling matrix of dimension M_1M_2 x N_1N_2, and H_r is the camera defocus blur matrix of dimension N_1N_2 x N_1N_2. Existing super-resolution techniques assume the camera defocus blur to be Gaussian (an approximation) and same for all the observations ($H_r = H$). In this work, we assume the H_rs to be different for each observation which can be used as an additional cue for super-resolution. W_r matrix represents geometric warping and is of dimension N_1N_2 x N_1N_2

for the r^{th} frame. The term $\underline{n_r}$ is the noise in the r^{th} frame which is assumed to be Gaussian, independent, and identically distributed. We assume that there are m number of LR observations i.e., $1 \leq r \leq m$.

Eq. (1) can be expressed in matrix-vector form as

$$\underline{y} = A\underline{x} + \underline{n} \tag{2}$$

Here, $\underline{y} = [\underline{y_1}\, \underline{y_2}\, \cdots \underline{y_m}]^T$, the system matrix A=[A_1, A_2,$\cdots$ A_m]^T where $A_r = DH_rW_r$, and $\underline{n} = [\underline{n_1}\, \underline{n_2}\, \cdots \underline{n_m}]^T$. The matrix A_r is of dimension $(M_1M_2 \times N_1N_2)$ and hence A is of dimension $(m \cdot M_1M_2) \times N_1N_2$. When $m = N_1N_2/M_1M_2$, the system matrix A is a square matrix of size $N_1N_2 \times N_1N_2$.

Solving for $\underline{x}$ in Eq. (2) given the observations $\underline{y}$ requires that the system matrix A be invertible. If the down-sampling matrix D and the blur matrix H_r are assumed to be the same for all the observations, the matrix A will be invertible only when the warping matrices W_r are different. In the event that any two observations have either identical or very close motion shifts (which is a possibility in a real scenario), the matrix A becomes ill-conditioned. If the H_rs are different, then relative blur among the observations can be expected to play a major role in reducing the condition number of the system matrix so as to yield a better output.

3. ROLE OF BLUR

Blurring is inherent during the formation of the image due to low-resolution sensor point spread spectrum (PSF). It may also be present due to relative motion between the camera and the scene. In the case of real aperture cameras, the defocus blur at a point is a function of depth of the scene [9]. If the scene is planar, the blur will be space-invariant but depending on the distance from the camera, the observations may have different blur. For a 3D scene, the blur PSF will be space-variant. Estimation of blur kernel even of a smaller size of say 3×3 involves estimation of 9 unknowns and hence is a very difficult task. Usually the blur will be modeled as parametric either as pill box or Gaussian [9]. In this paper, we assume that the defocus blur is space-invariant which is modeled as circularly symmetric 2-D Gaussian with blur parameter σ. Since we assume sub-pixel motion, motion blur is neglected.

3.1. Cramér-Rao Lower Bound

We now quantitatively analyze the effect of relative blur on the accuracy of the estimate of the reconstructed super-resolved image. The analysis is based on the Cramér-Rao lower bound (CRLB) [10], which provides a fundamental limit on the variance of the error attainable with an estimator for an unknown parameter. The CRLB expresses the minimum error variance of any estimator $\hat{x}(y)$ of x in terms of the conditional density $p(y|x)$ of the data. We now derive the CRLB for the motion super-resolution problem.

For an unbiased estimator, the CRLB [10] on the covariance of the estimate of $\underline{x}$ is given by

$$E[(\underline{x} - \underline{\hat{x}})(\underline{x} - \underline{\hat{x}})^T] \geq J^{-1} \tag{3}$$

where E is the expectation operator, $\underline{\hat{x}}$ is an unbiased estimator of $\underline{x}$, and

$$J = -E\left[\frac{\partial^2}{\partial \underline{x}^2} \log p(\underline{y}|\underline{x})\right] \tag{4}$$

Since the $\underline{n_r}$'s are assumed to be Gaussian and independent with variance σ_n^2, we have

$$p(\underline{y}|\underline{x}) = \frac{1}{(2\pi\sigma_n^2)^{m \cdot \frac{M_1M_2}{2}}} \exp\left[-\sum_{r=1}^{m} \frac{(\underline{y_r} - A_r\underline{x})^T(\underline{y_r} - A_r\underline{x})}{2\sigma_n^2}\right]$$

Thus

$$\frac{\partial}{\partial \underline{x}} \log p(\underline{y}|\underline{x}) = -\frac{1}{2\sigma_n^2}\left(\sum_{r=1}^{m} -2A_r^T\underline{y_r} + 2A_r^TA_r\underline{x}\right), \tag{5}$$

$$\frac{\partial^2}{\partial \underline{x}^2} \log p(\underline{y}|\underline{x}) = -\frac{1}{\sigma_n^2}\sum_{r=1}^{m} A_r^TA_r \tag{6}$$

Hence

$$J = -E\left[\frac{\partial^2}{\partial \underline{x}^2} \log p(\underline{y}|\underline{x})\right] = \frac{1}{\sigma_n^2}\sum_{r=1}^{m} A_r^TA_r \tag{7}$$

and the CRLB turns out to be

$$E[(\underline{x} - \underline{\hat{x}})(\underline{x} - \underline{\hat{x}})^T] \geq \sigma_n^2\left(\sum_{r=1}^{m} A_r^TA_r\right)^{-1} \tag{8}$$

Invertibility of the matrix $\sum_{r=1}^{m} A_r^TA_r$ depends on the warping matrices W_r and the blur matrices H_r. From Eq. (8), for a given noise variance, the CRLB depends on the condition number of $\sum_{r=1}^{m} A_r^TA_r$. Hence, when the motion vectors are not evenly spaced or the observations have identical motion shifts, relative blur among the observations can contribute towards reducing the condition number.

4. EXPERIMENTAL RESULTS

In this section, we study the effect of relative blur on the quality of the super-resolved output. Initially, to demonstrate the role of blur, the trace of the right hand side of Eq. (8) (a scalar performance bound [6]) for different relative blur was plotted for the case of a 1D signal of length 66 samples. We used the resolution factor to be 3 and the noise standard deviation $\sigma_n = 2.25$. Three LR observation sequences with shifts of [0 0.5 0.6] were used for the simulation. The third observation was intentionally chosen to be close to the second

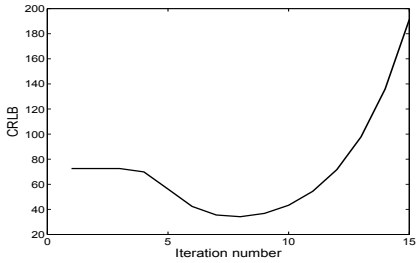

Fig. 1. CRLB versus relative blur.

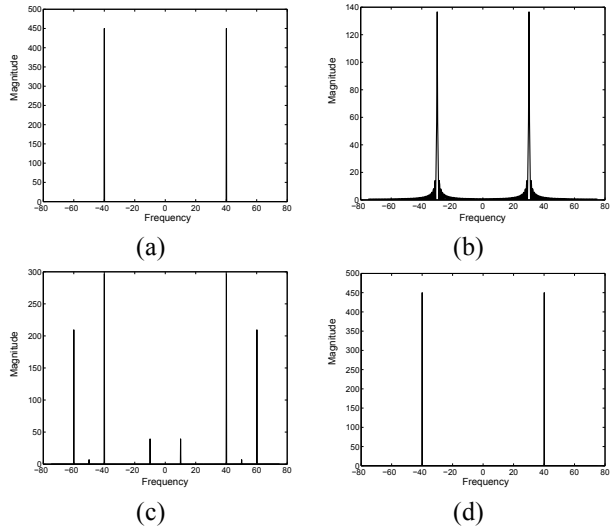

Fig. 2. A 1D example. Magnitude spectrum of (a) original sequence and (b) an LR sequence. Spectrum of super-resolved sequence with (c) no relative blur and (d) good relative blur.

one. The blur parameters of these observations were varied from $[\sigma_1 = 0.02, \sigma_2 = 0.02, \sigma_3 = 0.02]$ to $[\sigma_1 = 0.17, \sigma_2 = 0.77, \sigma_3 = 1.52]$ by increment of $[0.01, 0.05, 0.1]$ in each iteration. The plot of CRLB versus the iteration number (corresponding to different relative blur) is shown in Fig. 1. It can be seen from the figure that there exists an optimal relative blur at which the mean squared error (MSE) is minimum (in this case the minimum occurred at $[\sigma_1 = 0.09, \sigma_2 = 0.37, \sigma_3 = 0.72]$). The plot also demonstrates that beyond a certain limit, relative blur does not help in better reconstruction because deblurring capability becomes poor when the absolute blur is very high.

To get additional insight, we initially conducted simulations on a 1D sequence. We considered a single tone sinusoidal signal of frequency 40 Hz sampled at 150 Hz as the original sequence. Fig. 2(a) shows the plot of the magnitude spectrum of the sequence. Three LR sequences were generated from this sequence by down-sampling by a factor of 3 and warping by $[0\ 0.5\ 0.5]$ samples. Fig. 2(b) gives the magnitude spectrum of the LR observation showing the aliased

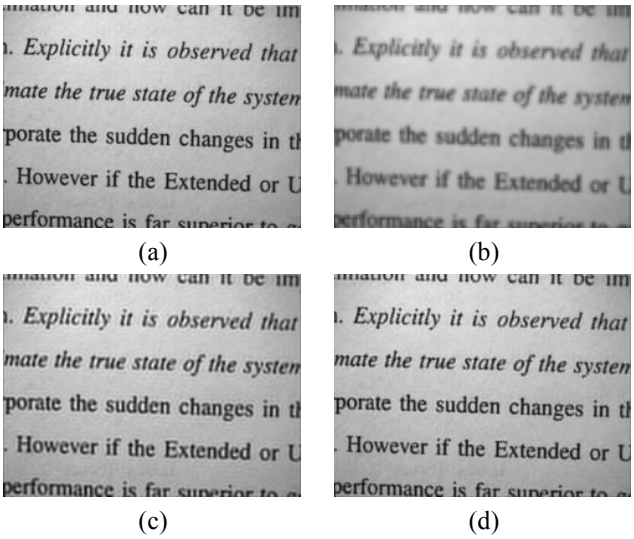

Fig. 3. A synthetic example. (a) Original image. (b) Interpolated result. Super-resolved image with (c) no relative blur (MSE = 110.81) and (d) good relative blur (MSE = 35.04).

frequency. The second and the third sequence are intentionally assumed to have same shifts rendering the system matrix A to be non-invertible when the blur kernels are assumed to be the same for all the observations. The blur parameter for the observations is assumed to be $[0.4\ 0.4\ 0.4]$ and an inverse in the least square sense yields an output whose spectrum is as shown in Fig. 2(c). It can be observed that there is no dealiasing in this case. We then introduced relative blur among the observations by choosing the blur parameter as $[0.4\ 0.6\ 0.8]$. The magnitude spectrum of the reconstructed signal is plotted in Fig. 2(d) which is the same as the original spectrum in Fig. 2(a). This experiment clearly demonstrates that relative blur plays an important role in dealiasing in motion super-resolution.

Next, we carried out simulations on a synthetic image. A high-quality "Text" image of dimension 180×240 pixels (Fig. 3(a)) was used to generate 4 LR observations by down-sampling by a factor of 2. The sub-pixel motion parameters were chosen to be $(0,\ 0)$, $(0,\ 0.5)$, $(0.5,\ 0)$, and $(0.5,\ 0)$. Here again, the shifts for the third and fourth frame were chosen to be same. Fig. 3(b) shows the bilinearly interpolated result. As expected, interpolation can not recover the high-frequency components. The super-resolved results corresponding to two different sets of blur parameters: $[\sigma_1 = 1.5, \sigma_2 = 1.5, \sigma_3 = 1.5, \sigma_4 = 1.5]$ and $[\sigma_1 = 0.8, \sigma_2 = 1.2, \sigma_3 = 1.4, \sigma_4 = 1.6]$ are shown in Figs. 3(c) and 3(d), respectively. The result with good relative blur is much better than the one with no relative blur. The mean squared error (MSE) for the case of good relative blur was much less than that with no relative blur.

Finally, we considered the important case of real data. An LV150-Nikon industrial microscope with a $2.5X$ objective

(a) (b) (c)

Fig. 4. A real example. (a) An LR image. Super-resolved image with (b) no relative blur and (c) good relative blur.

lens and a camera attached to it was used for capturing images in the lab. A planar object with text written on it was placed on the experimental-bed of the microscope and a small area of the object was imaged. The goal was to super-resolve text using the observations from the microscope. The experimental-bed was moved gently using the X-Y movement arrangement to introduce sub-pixel motion shifts while Z movement was used to introduce different defocusing among the observations. Several low-resolution observations of size 96×128 pixels of the object were captured. Fig. 4(a) shows one of the LR observations. Four LR observations captured with same defocus blur were then used to obtain a super-resolved image with a resolution factor of 3. The motion parameters were estimated using the technique proposed in [11]. By knowing the camera parameters and the distance of the object from the lens, the space-invariant blur parameter σ was determined. Ideally, we need 9 images for a magnification factor of 3 [7]. Hence the reconstructed result (Fig. 4(b)) with only 4 observations is not very good. We then used 4 sub-pixel shifted observations captured with different defocus blur. The resultant super-resolved image is shown in Fig. 4(c). Note that the quality of the output is very good with sharper features as compared to the earlier case.

5. CONCLUSIONS

A motion super-resolution technique which exploits the relative blur among the sub-pixel shifted LR observations was presented. When the motion vectors are not uniformly spaced or the number of observations is inadequate, the defocus information among the observations can be exploited to yield a good quality super-resolved image. Several synthetic and real results were given to demonstrate the role of relative blur.

6. REFERENCES

[1] T. S. Huang and R. Y. Tsai, "Multiple frame image restoration and registration," *Advances in Computer Vision and Image Process.*, vol. 1, pp. 317–339, 1984.

[2] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 742–746, 1988.

[3] R. L. Stevenson and R. R. Schultz, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, pp. 996–1011, 1996.

[4] A. N. Rajagopalan and V. P. Kiran, "Motion-free super-resolution and the role of relative blur," *J. Opt. Soc. Am. A*, vol. 20, pp. 2022–2032, 2003.

[5] D. Rajan and S. Chaudhuri, "Generation of super-resolution images from blurred observations using Markov random fields," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 3, pp. 1837–1840, 2001.

[6] D. Robinson and P. Milanfar, "Statistical performance analysis of super-resolution," *IEEE Trans. Image Process.*, vol. 15, pp. 1413–1428, 2006.

[7] Z. Lin and H. Shum, "Fundamental limits of reconstruction-based super-resolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, pp. 83–97, 2004.

[8] M. Ben-Ezra, A. Zomet, and S. K. Nayar, "Video super-resolution using controlled sub-pixel detector shifts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 977–987, 2005.

[9] S. Chaudhuri and A. N. Rajagopalan, "Depth from defocus: A real-aperture imaging approach," *Springer-Verlag, New York*, 1999.

[10] J. M. Mendel, "Lessons in digital estimation theory," *Prentice-Hall, Englewood Cliffs, NJ*, 1987.

[11] S. Peleg and M. Irani, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 231–239, 1991.