NEW RESULTS ON PERFORMANCE ANALYSIS OF SUPER-RESOLUTION IMAGE RECONSTRUCTION

Junlan Yang, Dan Schonfeld

Department of Electrical and Computer Engineering,
University of Illinois, Chicago, IL, 60607
{jyang24,dans}@uic.edu

ABSTRACT

In this paper, we present new results in performance analysis of super-resolution (SR) image reconstruction. We investigate bounds on the improvement in resolution that can be achieved and its relation to the image sequence. We derive lower bounds on the resolution enhancement factor based on a frequency-domain SR algorithm. We subsequently show that the bounds remain valid for other SR algorithms. Moreover, we consider an image sequence model in the presence of affine motion. We demonstrate theoretically and experimentally that incorporation of affine motion into the image model can be used to increase the enhancement factor in comparison to purely translational motion. Finally, we discuss the extension of the performance bounds to temporal super-resolution methods and its implications on the image sequence.

1. INTRODUCTION

Super-Resolution (SR) in general refers to algorithms of processing multiple low-resolution (LR) images to reconstruct a high-resolution (HR) image. Super-Resolution algorithms can be divided into two main categories: spatial-domain methods [3] and frequency-domain methods [1]. Spatial-domain methods regard the LR images to be formed by applying a series of linear degradations to the original HR image. Frequency-domain methods introduced in [1] are considered to be one of the pioneering work in the content of SR. It justifies theoretically that super-resolution can be achieved by taking advantage of aliasing effects in the frequency domain. The limitation of the algorithm is it only considers frames with relative shifts, i.e., only translational motion. Among numerous SR theories and algorithms proposed, there are not many results concerning the performance limits of the SR problem. Authors in [4] state that when the magnification factor increases, the difficulty of SR reconstruction increases dramatically. In [5] authors discuss the statistical performance of SR algorithm using Cramer-Rao (CR) bounds. The results takes into account image registration and restoration simultaneously.

In this paper, we try to answer a rather direct question: given a sequence of images, how much resolution improvement can at least be achieved and how is it related to the specifications of the sequence? We start with the frequency-domain method proposed in [1] and extend it from considering only translational motions to accommodating general 2D affine motions. In the meanwhile, we identify conditions posed during the derivation of the proposed algorithm which can further translate to achievable bounds of the enhancement factors. We show that adding image frames with affine motions gives a new achievable bound for the enhancement factors. Under the framework, we study the factors affecting the reconstruction errors. In addition, we generalize the SR problem to include SR in temporal domain. The formulation will be especially useful for videos. The conditions on performing temporal SR is also discussed. Several experimental results are provided which confirm our proposition. It will also be shown with experiments that our results derived based on a particular frequency-domain method can also be observed when using other spatial-domain methods.

2. PERFORMANCE ANALYSIS OF SUPER-RESOLUTION WITH AFFINE MOTION

2.1. Theoretical Foundations of Frequency-Domain Analysis of Super-Resolution

We begin by assuming a moving camera is looking at a static scene and taking a video of it. We model the scene as a continuous function $f(x, y)$. In time t_0 and time $t = t_k$, the camera takes two images, frame 0 ($f^{(0)}$) and frame k ($f^{(k)}$), $k = 1...P-1$. P is the number of frames that the camera takes. In the camera's coordinate system, the scene is moving due to the motion of the camera. Denote $f_0(x, y)$ as the scene in t_0 and $f_k(x, y)$ in time t_k. Assuming the camera has a sampling period T and a sampling rate $w_s = 2\pi/T$, we have $f^{(k)}(i, j) = f_k(iT, jT)$, $k = 0, 1...P-1$. We denote two Fourier Transform pairs: Continuous Fourier Transform (CFT) of $f_k(x, y)$ as $F_k(w_1, w_2)$ and the Discrete Fourier Transform (DFT) of $f^{(k)}(i, j)$ as $F^{(k)}(m, n)$, $k = 0, 1...P-1$. w_1, w_2 are frequency indices and $m = 1...M, n = 1...N$. $M \times N$ is the size of the image. The problem of super-resolution in frequency domain is given DFT of all the P

frames to recovery as many samples of the original Fourier Transform $F_0(w_1, w_2)$ as possible so that they can be used to provide highly resolved image. When the sample numbers are larger than that of $F^{(k)}(m,n)$, we achieve a resolution magnification of the image.

It is well known that DTFT for a sampled signal is the superposition of shifted replicas of CFT of original continuous signals, and DFT is a sampled version of DTFT. It is easy to verify the following relationship between DFT and CFT [1]:

$$F^{(k)}(m,n) = \frac{1}{T^2}\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} F_k(\frac{2\pi m}{MT}+rw_s, \frac{2\pi n}{NT}+sw_s), \tag{1}$$

where $m = 1, ...M, n = 1, ...N$, r, s are integers. To obtain relationships between $F^{(k)}(m,n)$ and $F_0(w_1, w_2)$, we need to relate $F_k(w_1, w_2)$ with $F_0(w_1, w_2)$, which can be obtained based on the motion transformations between the frames.

2.2. Frequency-Domain Analysis of Super-Resolution with Affine Motion

Instead of considering only translational motion as in [1], we assume the camera undergoes affine motions during the acquisition process. A point Z's coordinate in time t_0 (x_0, y_0) and in time t_k (x_k, y_k) are related by 2D affine transforms:

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} a_k & b_k \\ d_k & e_k \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} c_k \\ f_k \end{bmatrix}, \tag{2}$$

where a_k, b_k, d_k, e_k are linear transformation parameters relating to camera rotation and scaling. c_k, f_k are camera translation parameters. According to the affine theorem for 2D Fourier transform [2], we have the following relationship:

$$F_k(w_1,w_2) = \frac{1}{|\Delta_k|}F_0(\frac{e_kw_1 - d_kw_2}{\Delta_k}, \frac{-b_kw_1 + a_kw_2}{\Delta_k})$$
$$\cdot exp\{\frac{j}{\Delta_k}[(e_kc_k - b_kf_k)w_1 + (a_kf_k - c_kd_k)w_2]\}, \tag{3}$$

where $\Delta_k \equiv a_ke_k - b_kd_k$. We first normalize the linear transformation parameters to simplify notations by replacing a_k with a_k/Δ_k and likewise for b_k, e_k and d_k. It can be shown after combining (3) with (1) and some additional algebra that

$$F^{(k)}(m,n)$$
$$= \frac{1}{T^2}\sum_{r,s=-\infty}^{\infty} exp\{j(c_ku^{(k)}_{m,n} + f_kv^{(k)}_{m,n})\}F_0(u^{(k)}_{m,n}, v^{(k)}_{m,n})$$
$$u^{(k)}_{m,n} \triangleq \frac{2\pi}{MT}(e_km - d_kn\frac{M}{N}) + (e_kr - d_ks)w_s,$$
$$v^{(k)}_{m,n} \triangleq \frac{2\pi}{NT}(a_kn - b_km\frac{N}{M}) + (a_ks - b_kr)w_s. \tag{4}$$

To further proceed, we define the mappings $(m', n') = g_k(m,n)$ and $(r', s') = h_k(r,s)$ as follows:

$$m' = e_km - d_knM/N \quad and \quad n' = a_kn - b_kmN/M;$$
$$r' = e_kr - d_ks \quad and \quad s' = a_ks - b_kr. \tag{5}$$

Along with $w_s = 2\pi/T$, Eq. (4) is simplified as

$$F^{(k)}(g_k^{-1}(m',n')) = \frac{1}{T^2}\sum_{r',s'=-\infty}^{\infty} F_0[(\frac{m'}{M}+r')w_s, (\frac{n'}{N}+s')w_s]$$
$$\cdot exp\{j[c_k(\frac{m'}{M}+r')w_s + f_k(\frac{n'}{N}+s')w_s]\}, \tag{6}$$

where g_k^{-1} denotes the inverse mapping of g_k.We make the assumption that the mapped indices (m', n') and (r', s') are still integers. Otherwise we can interpolate them to the nearest integers as we usually do in image registration.

Following the similar argument in [1], we find a pair of integers (L_1, L_2) such that for all the k (including $k = 0$), the mapped pair $(L_1', L_2') = h_k(L_1, L_2)$ satisfies

$$F_0(w_1,w_2) \to 0 \quad \forall \quad |w_1| \geq L_1'w_s, \quad |w_2| \geq L_2'w_s. \tag{7}$$

Note that $h_0(L_1, L_2) = (L_1, L_2)$. It is then automatically satisfied that $F_0(w_1, w_2) \to 0, \forall |w_1| \geq L_1w_s$ and $|w_2| \geq L_2w_s$. Therefore after deciding on (L_1, L_2) according to (7), we can try to estimate $F(w_1, w_2)$ within the range of $[-L_1w_s, L_1w_s]\times[-L_2w_s, L_2w_s]$. Note for each $(m', n'), m' = 1...M, n' = 1...N$, (6) can be rewritten as the following matrix equation:

$$\mathbf{Y}_{P\times 1} = \mathbf{A}_{P\times 4L_1L_2}\mathbf{X}_{4L_1L_2\times 1}, \tag{8}$$

$$\mathbf{Y}(k) = F^{(k)}(g(m',n')), \quad k = 1, ..., P$$
$$\mathbf{X}(l) = F_0(\frac{m'}{M}w_s + r'w_s, \frac{n'}{N}w_s + s'w_s), \quad l = 1, ..., 4L_1L_2$$
$$\mathbf{A}(k,l) = 1/T^2 exp\{j[c_k(\frac{m'}{M}+r')w_s + f_k(\frac{n'}{N}+s')w_s]\},$$
$$where \quad r' = lmod(2L_2) - L_1, s' = l - 2L_2\lfloor l/2L_2\rfloor - L_2.$$

When $P \geq 4L_1L_2$, the system is over-determined where a least-square solution of $\mathbf{X}$ can be obtained. Solving the equation for each (m', n'), we obtain the estimate of $F_0(w_1, w_2)$ ranging in $[-L_1w_s, L_1w_s - w_s/M]\times[-L_2w_s, L_2w_s - w_s/N]$ with spacing w_s/M and w_s/N for each dimension. Applying the inverse DFT, we can obtain an estimate of $f(x,y)$ at discrete points of $x = 0, y = 0$ to $x = (M-1)T, y = (N-1)T$ with spacing $T/2L_1$ and $T/2L_2$. Comparing to $f^{(0)}(i,j)$, the resolution is increased by a factor of $2L_1 \times 2L_2$.

2.3. Performance Analysis of Super-Resolution with Affine Motion

As mentioned previously, the conditions on the magnification factors can be identified as finding a common pair of integers L_1, L_2 such that (7) is satisfied for every frame. The following theorem formalizes this results:

Theorem 1 *If L_1, L_2 satisfy the following two conditions:*

$$|e_kL_1 - d_kL_2|w_s \geq w_{max,1}, \quad \forall k = 0, 1, ...P-1; \tag{9}$$
$$|-b_kL_1 + a_kL_2|w_s \geq w_{max,2}, \quad \forall k = 0, 1, ...P-1; \tag{10}$$

where $w_{max,1}$, $w_{max,2}$ are the maximum frequency of F_0, i.e., $F_0(w_1, w_2) \to 0$, $\forall |w_1| \geq w_{max,1}$ and $|w_2| \geq w_{max,2}$, we can at least achieve a resolution enhancement of $2L_1 \times 2L_2$.

We usually choose L_1, L_2 as the smallest integers that satisfy (9) and (10) since larger L_1, L_2 in linear equation (8) only introduce zeros to the vector of X. Padding zeros to the frequency domain can result more samples in time domain, however they are interpolated values from the previous estimates. There is no new information introduced by larger L_1, L_2 to the system. Therefore we will see in the experiments forcing larger L_1, L_2 usually results in large errors.

The following two corollaries are easy to observe from Theorem 1. Note that Corollary 2 shows the result corresponding to the special case of pure translational motions, which coincides with results in [1].

Corollary 1 *If we try to increase the resolution by a same factor for each dimension, i.e.,* $L_1 = L_2 = L$, (9) *and* (10) *can be simplified as*

$$L \geq max(\frac{w_{max,1}}{\min_k(|e_k - d_k|)w_s}, \frac{w_{max,2}}{\min_k(|a_k - b_k|)w_s}). \quad (11)$$

Corollary 2 *(Tsai and Huang [1]) If motions are pure translational, i.e.,* $a_k = e_k = 1$, *and* $b_k = d_k = 0\ \forall k$, (9) *and* (10) *are reduced to* $L_1 \geq w_{max,1}/w_s, L_2 \geq w_{max,2}/w_s$.

3. ERROR ANALYSIS AND EXTENSION TO TEMPORAL SUPER-RESOLUTION

In this section, we discuss issues relating to error analysis, temporal super-resolution and its implications on the sequence.

3.1. Reconstruction Error Analysis

Notice the reconstructed $\hat{F}(w_1, w_2)$ is a sampled version of the continuous frequency $F(w_1, w_2)$. Assuming the estimate of the samples are exact, we have

$$\hat{F}(w_1, w_2) = \sum_{l_1, l_2 = -\infty}^{\infty} F(l_1\frac{w_s}{M}, l_2\frac{w_s}{N})\delta(w_1 - l_1\frac{w_s}{M}, w_2 - l_2\frac{w_s}{N}).$$

According to convolution theory, the inverse DFT is given by

$$\hat{f}[i,j] = \frac{MNT^2}{4\pi^2} \sum_{l_1, l_2 = -\infty}^{\infty} f(iT - l_1MT, jT - l_2NT).$$

When we assume the signal is band-limited which suggests that it is spatially-unlimited, aliasing effects shown in above equation is inevitable due to frequency domain sampling. Therefore, although not clearly stated in [1], we see that perfect reconstruction cannot be achieved in general, i.e., $4\pi^2/(MNT^2) \cdot \hat{f}[i,j] \neq f(iT, jT)$. In fact, such a limitation is expected since super-resolution is inherently an inverse problem from partial information. Nonetheless, it is reasonable to assume in practice, the signal is bandlimited while at the same time tails off rapidly outside of a certain range in space, i.e., $f(x,y) \rightarrow 0$ for (x,y) outside of $[-x_m, x_m] \times [-y_m, y_m]$. In this case, aliasing error is bounded and the proposed algorithm can be used to provide a good approximation of the original signal. When the sample interval $2\pi/(MT)$ of the estimate decreases, the reconstruction error will decrease. When $x_m < MT/2, y_m < NT/2$, there will be no aliasing.

3.2. Temporal Super-Resolution

Regard a video as a three-dimensional discrete signal, which is sampled both spatially and temporarily from a continuous changing scene. We denote a particular frame k as $f^{(k)}(i,j) = f(iT_1, jT_1, kT_2)$, where T_1 is the spatial sampling period and T_2 is the temporal sampling period which is the reciprocal of the frame rate. Temporal Super-Resolution is to increase the frame rate of the video, producing intermediate frames. For the temporal dimension, a single video is only one set of regular samples. Therefore from sampling theory, the requirement to have perfect interpolations in time is that the frame rate has to be larger than the Nyquist rate, i.e., $2\pi/T_2 \geq 2w_{max,3}$, where the $w_{max,3}$ is the maximum frequency for temporal dimension. $w_{max,3}$ is affected by several factors such as the moving velocity of the camera and abrupt scene changes.

4. SIMULATION RESULTS

We report some computer simulated experiments for testing the frequency-domain SR method with affine motions while also examining the proposed limits of magnification factors.

(1) Sequence PATTERN: A synthetic LR sequence of 16 images is generated by applying affine motion to a HR image followed by a 4×4 down-sampling. Rotated images are generated using bilinear interpolation. We pre-filter the images so that it is bandlimited by the sampling rate, i.e., $w_{max,1} = w_{max,2} = w_s$. The set of affine motions includes in-plane rotations of $0, 10, 15, 20$ degrees and translations of $[0,0]$,$[0,2]$,$[2,0]$ and $[2,2]$. According to Corollary 1 and 2, we find that when using only translational frames, we can achieve a magnification of 2×2 while a magnification of 4×4 can be achieved by processing all the affine frames. Figure 1 illustrates the frequency-domain SR method. Samples of affine transformed LR images are depicted in Fig.1(a). The original HR image is provided in Fig.1(b). The bilinear interpolation of LR by a factor of 2×2 and 4×4 are shown in Figs.1(c) and 1(d). The frequency-domain SR reconstructions of 2×2 and 4×4 magnification achieved by processing only translated frames are shown in Figs.1(e) and (f), respectively. Finally, the 4×4 magnification achieved by processing all 16 frames is represented in Fig.1(g). We observe that the presence of rotated frames increases the achievable magnification factor. The PSNR of the images in (c)-(g) are reported in Table I. We note that if we use only translated frames, the reconstruction quality decreases rapidly when the magnification factor exceeds 2. Whereas, if we rely on affine transformed frames, the quality of 4×4 reconstruction is greatly improved.

We also examine the effects of the number of frames used. We try to achieve a magnification of 2×2 by using 2 frames, 3 frames, 6 frames or 9 frames. The resulting PSNR are 10.613 dB, 17.442 dB, 19.509 dB and 20.234 dB correspondingly. Since $P < 4L_1L_2 = 4$ is not sufficient, the reconstruction quality when using 2 frames and 3 frames are impaired. Increasing the number of frames improves the reconstruction

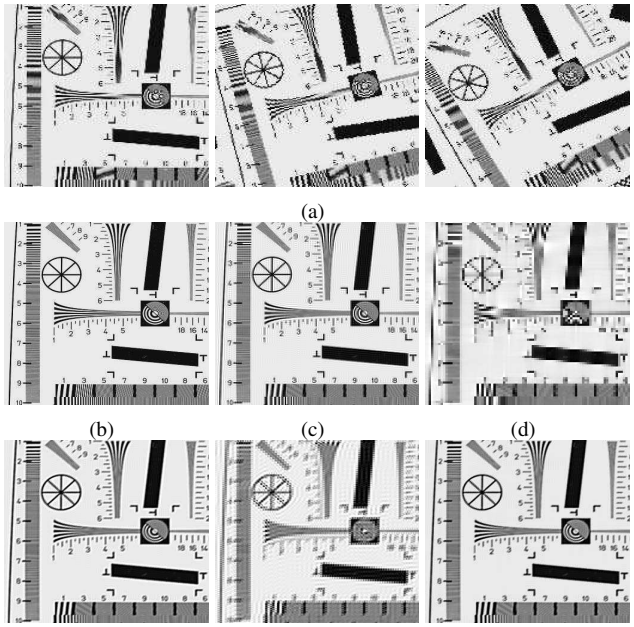

Fig. 1. Frequency-domain super-resolution (SR) methods: (a) low-resolution (LR) images; (b) original high-resolution (HR) image; (c) 2×2 interpolation; (d) 4×4 interpolation; (e) 2×2 magnification using only translational frames; (f) 4×4 magnification using only translational frames; and (g) 4×4 magnification using all 16 frames.

Table 1. PSNR for Reconstruction Results in Figure 1

Figure	(c)	(d)	(e)	(f)	(g)
PSNR(dB)	18.236	15.830	19.399	16.342	21.797

while it is also true that after we have sufficient number of frames, adding more frames only improves PSNR slightly.

(2) Sequence STILLS: We wish to test our achievable bound for magnification factor with other time-domain SR, to verify that it is a general property of a image sequence independent of SR methods. We employ the time-domain SR algorithm proposed in [3]. We generat a sequence of 24 LR images by applying a mixture of $0, 10, 20$ degree rotations and 8 integer-pixel translations and then down-sampling by a factor of 4×4. The bounds given by Corollary 1 and 2 are once again 2×2 for translated frames and 4×4 for affine frames. Figure 2 and illustrates the performance of time-domain SR methods. The HR image is depicted in Fig.2(a). LR images are represented in Figs.2(b) and 2(c). The 4×4 magnification using 8 translational frames is shown in Fig.2(d). The 4×4 and 6×6 magnification achieved by processing all 24 frames are illustrated in Figs.2(e) and 2(f), respectively. For (f), since the size exceeds the original HR, We compare the reconstruction with interpolated HR. We note that using affine transformed frames can achieve a higher magnification factor. However, when we continue to increase the magnification factor, the reconstruction quality begins to degrade.

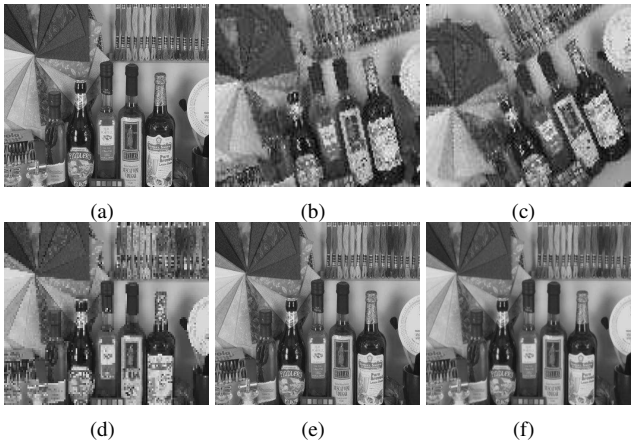

Fig. 2. Time-domain super-resolution (SR) methods: (a) original high-resolution (HR) image; (b)-(c) low-resolution (LR) images; (d) 4×4 magnification using 8 translational frames; (e) 4×4 magnification using 24 frames; and (f) 6×6 magnification using 24 frames.

Table 2. PSNR for Reconstruction Results in Figure 2

Figure	(d)	(e)	(f)
PSNR(dB)	19.103	28.494	26.445

5. CONCLUSIONS

In this paper, we introduced a frequency-domain super-resolution (SR) method for image sequences undergoing an affine transformation. We used this approach to derive achievable performance bounds on the resolution enhancement factor of SR techniques. We observed theoretically and experimentally that the presence of affine transformations in the image model results in an increase in the achievable bound and improves the quality of image reconstruction. We finally presented an error analysis in image reconstruction under this framework and discussed an extension of the performance analysis to temporal SR methods.

6. REFERENCES

[1] R. Tsai and T. Huang, "Multiframe Image Restoration and Registration", *Advances in Computer Vision and Image Processing*, vol. 1, pp. 317-339, 1984.

[2] R. Bracewell, K. Chang, A. Jha, and Y. Wang, "Affine theorem for two dimensional fourier transform", *Electronics Letters*, Feb. 1993.

[3] S. Farsiu, D. Robinson, M. Elad and P. Milanfar, "Fast and Robust Multiframe Super-Resolution", *IEEE Trans. on Image Processing*, vol. 13, pp. 1327-1344, 2004.

[4] S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them", *IEEE Trans. on PAMI*, vol. 24, No. 9, pp. 1167-1183, 2002.

[5] D. Robinson and P. Milanfar, "Statistical Performance Analysis of Super-Resolution", *IEEE Trans. on Image Processing*, vol. 15, No. 6, pp. 1413-1428, 2006.