On the Fundamental Limits of Reconstruction-Based Super-resolution Algorithms

Zhouchen Lin Heung-Yeung Shum
Microsoft Research, China
{zhoulin|hshum}@microsoft.com

Abstract

Super-resolution is a technique that produces higher resolution images from low resolution images (LRIs). In practice, people have found that the improvement in resolution is limited. The aim of this paper is to address the problem "do fundamental limits exist for super-resolution?". *Specifically, this paper provides explicit limits for a major class of super-resolution algorithms, called the reconstruction-based algorithms, under both real and synthetic conditions. Our analysis is based on perturbation theory of linear systems. We also show that a sufficient number of LRIs can be determined to reach the limit. Both real and synthetic experiments are carried out to verify our analysis.*

1. Introduction

Super-resolution is a technique that combines low resolution images (LRIs) to produce higher resolution ones, where not only the sharpness is improved, but more importantly, the pixel sizes are also smaller than those of the LRIs. Namely, it is a combination of deblurring and fine resolving. Various algorithms [1, 2, 3, 5, 8, 9] have been proposed for super-resolution. Most of them are so called reconstruction-based algorithms (RBAs) (coined in [2]).

RBAs start from the continuous image formation equation:

$$L(\mathbf{y}) = \int PSF(\mathbf{y},\mathbf{x})H(\mathbf{x})d\mathbf{x} + E(\mathbf{y}), \qquad (1)$$

where $L(\mathbf{y})$ is the continuous low resolution irradiance field on the image plane, $PSF(\mathbf{y},\mathbf{x})$ is the point spread function (PSF) of the system, $H(\mathbf{x})$ is the continuous high resolution irradiance light-field that would have reached the image plane and $E(\mathbf{y})$ is the noise. Equation (1) is a general formulation if it is considered locally, including motion deblurring where the PSF has been integrated in time. Then (1) is discretized to form linear systems, where the sampling density of $L(\mathbf{y})$ is lower than that of $H(\mathbf{x})$. A smoothness prior, such as maximum *a posteriori* (MAP) [6, 5, 10] or maximum likelihood (ML) [5], is often required in order to remove noise and solve the system stably.

The performance of RBAs is affected by several factors: the level of noise that exists in the LRIs, the accuracy of the PSF estimation and the accuracy of registration, including the correction of geometric distortion. The higher level of noise or the poorer PSF estimation and registration, the less improvement in resolution.

In practice people have found that the improvement in the resolution is quite limited. Current algorithms, not limited to RBAs, can only produce overly smooth images or images with undesirable details when the magnification factor becomes large.

It is important to address the problem *"do fundamental limits exist for super-resolution?"* so that practitioners can avoid trying those unduly large magnification factors. Furthermore, a better understanding would help people choose good magnification factors under the limits so as to make better use of system resources. Towards this problem, Baker and Kanade [2] showed that both the condition number of the system and the volume of solutions grow fast with the increment of the magnification factor, and an RBA can only generate an overly smoothed solution. However, they only pointed out the deteriorating tendency of the RBAs. Elad and Feuer [5] also briefly discussed the choice of magnification factors, but their assumption that the coefficient matrix is block-Toeplitz cannot be justified. Moreover, their results appear to permit arbitrarily large magnification factors.

In this paper, we will give explicit bounds of the magnification factor under both practical and synthetic situations. We further analyze the influence of increasing the number of LRIs and determine the optimal fractional part of the magnification factor and the corresponding sufficient number of LRIs. These conclusions result from our discovery that if some special low resolution pixels (LRPs) called the vertices set have been captured then the resolution will not be further improved.

The remainder of this paper is organized as follows. In Section 2, we introduce the main contributions of this paper and sketch the route of our proof. In Section 3, we introduce the detailed proof and present the fundamental limits of super-resolution under both practical and synthetic situations. In Section 4, we study the influence of increasing the number of LRIs. In Section 5, both real and synthetic experiments are presented to verify our theory. Finally, we give the conclusions and discuss some issues in Section 6.

2 Main Results

We prove that fundamental limits do exist for RBAs by deriving an inequality that an effective magnification factor, which is *possible* to produce higher resolution than smaller ones do, should satisfy. Specifically, we discuss two ex-

0-7695-1272-0/01 $10.00 © 2001 IEEE

treme cases and find that:

- The practical limit is 1.6, if the noise level is relatively high and is not removed effectively enough.
- The synthetic limit is 5.7. Moreover, the effective magnification factor can only distribute on some disjoint intervals.

We prove the inequality by applying perturbation theory to the linear systems assembled by the RBAs. The cornerstone for proving the bounds is a theorem on the perturbation of the least-square-error (LSE) solutions to overdetermined systems, which states how much the LSE solution deviates from the real one if noise and error exist. The limits are found by bounding the deviation. To estimate the deviation, we succeed in finely estimating the norms of the pseudo-inverse of the coefficient matrix of the system and the perturbation coefficient matrix.

After analyzing the influence of increasing the number of LRIs, we find that:

- If the vertices set have been captured then adding more LRIs will not further improve the resolution.
- For synthetic data, the best choice of the fractional part of the magnification factor is 0.5 and the corresponding sufficient number of LRIs is four times the squared magnification factor.

3 Outline of the Proof

3.1 Fundamental Linear System and Its Perturbation

The discrete version of (1) gives the relation between LRPs and high resolution pixels (HRPs) via a linear system:

$$\mathbf{L} = \mathbf{P}\mathbf{H} + \mathbf{E}, \tag{2}$$

where $\mathbf{L}$ is the column vector of the intensities[1] of all LRPs considered, $\mathbf{H}$ is the vectorized (namely concatenation of matrix rows) high resolution image (HRI), $\mathbf{P}$ gives the weights of the HRPs in order to obtain the intensities of corresponding LRPs, and $\mathbf{E}$ is the noise. $\mathbf{P}$ is required to be of full rank, i.e., the rank of $\mathbf{P}$ equals the number of columns in $\mathbf{P}$. This requirement is non-essential in our analysis because otherwise the solution that an RBA finds is smoother than the one from full rank $\mathbf{P}$, due to the smoothness regularization.

No matter how the RBAs vary, they can be viewed as denoising first and then solving (2) by finding its LSE solution. Such a viewpoint frees us from dealing with the details of RBAs. More explicitly, suppose an RBA finds a solution $\mathbf{H}_0$, then it can be viewed that the algorithm actually estimates the noise to be $\mathbf{E} = \mathbf{L} - \mathbf{P}\mathbf{H}_0 + \mathbf{E}_1$, leaves the residual error $\mathbf{r}$, and then finds the LSE solution to (2), where $\mathbf{E}_1 + \mathbf{r}$ is in the null space of $\mathbf{P}^T$. This will also give the solution $\mathbf{H}_0$. Note that the LSE solution to a system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{A}^+\mathbf{b}$, where $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. Therefore $\mathbf{E}$ in (2) can be considered as the user-estimated equivalent noise. It is no longer a random variable throughout our analysis.

[1] Since current RBAs compute in the greylevel domain, we also assume that they are greylevels.

Except those considering motion blur [4] in the super-resolution of video sequences, most existing RBAs throw all possible errors into $\mathbf{E}$ and the coefficient matrix $\mathbf{P}$ is deemed to be precise. Actually, this is not true. Regarding the errors in $\mathbf{P}$, the correct system should be a perturbed version of (2):

$$\mathbf{L} = \tilde{\mathbf{P}}\tilde{\mathbf{H}} + \tilde{\mathbf{E}}, \tag{3}$$

where $\tilde{\mathbf{P}} = \mathbf{P} + \delta\mathbf{P}$ is the correct coefficient matrix, $\tilde{\mathbf{H}} = \mathbf{H} + \delta\mathbf{H}$ is the correct solution and $\tilde{\mathbf{E}}$ is the exact noise. The error in estimating $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{E}}$ causes the deviation $\delta\mathbf{H}$ between $\tilde{\mathbf{H}}$ and $\mathbf{H}$. If $\delta\mathbf{H}$ is too large, then the HRIs cannot present correct details or can even lack details.

3.2 The Perturbation Theorem

The deviation $\delta\mathbf{H}$ can be depicted by a theorem [11, 7]:

Theorem: If both $\tilde{\mathbf{P}}$ and $\mathbf{P}$ are of full rank and $\eta < 1$, then

$$||\delta\mathbf{H}|| \leq \frac{\kappa}{1-\eta}\left[\varepsilon_{\mathbf{P}}\left(||\mathbf{H}|| + \kappa\frac{||\mathbf{r}||}{||\mathbf{P}||}\right) + \frac{||\delta\mathbf{E}||}{||\mathbf{P}||}\right], \tag{4}$$

where $\kappa = ||\mathbf{P}||||\mathbf{P}^+||$ is the condition number of the system, $\varepsilon_{\mathbf{P}} = ||\delta\mathbf{P}||/||\mathbf{P}||$, $\eta = \kappa\varepsilon_{\mathbf{P}}$, $\mathbf{r} = \mathbf{L} - \mathbf{P}\mathbf{H} - \mathbf{E}$ is the residual error and $\delta\mathbf{E} = \mathbf{E} - \tilde{\mathbf{E}}$ is the error in estimating the noise. The norms of a vector and a matrix are defined as: $||\mathbf{v}|| = \sqrt{\mathbf{v}^T \cdot \mathbf{v}}$, and $||\mathbf{A}|| = \max\limits_{||\mathbf{x}||\neq 0} \frac{||\mathbf{A}\mathbf{x}||}{||\mathbf{x}||}$, respectively.

Inequality (4) is a relatively sharp estimate of $||\delta\mathbf{H}||$, i.e., the equality may hold in some circumstances. This rules out abused versions of the argument in this paper, namely enlarge the right hand side (RHS) of (4) in order to obtain a tighter bound. Unfortunately, it is not guaranteed that $||\delta\mathbf{H}||$ will reach the magnitude of the RHS, because at the extreme case it is theoretically possible that $\mathbf{H}$ happens to be the solution of both (2) and (3) (or equivalently $\mathbf{H}$ is also the solution to $\delta\mathbf{E} = \delta\mathbf{P}\mathbf{H}$). However, usually we are not that lucky because $\delta\mathbf{P}$ and $\delta\mathbf{E}$ are independent in that they come from completely different sources. In practice, the possibility that the RHS is large while $||\delta\mathbf{H}||$ is small seems to be slight. Therefore, we should view $||\delta\mathbf{H}||$ as being quite correlated with the RHS. If the RHS is large, it is expected that $||\delta\mathbf{H}||$ is also large. Hence, a most conservative estimation of the RHS is a good estimate of $||\delta\mathbf{H}||$. The "conservativeness" refers to: 1. assuming as little error as possible in the measurement in order to make the RHS small; 2. allowing $||\delta\mathbf{H}||$ to be as large as possible. Such treatment shrinks the gap between the actual $||\delta\mathbf{H}||$ and the RHS.

3.3 The Limits of Reconstruction-Based Super-resolution

3.3.1 Assumptions

We first make some practical and reasonable assumptions in order to analyze the limits.

- The super-resolution is done locally on small regions of interest (ROI). This not only speeds up the super-resolution and lowers the memory requirement, but also make the problem simpler because both the PSF and geometric distortion in a small region are uniform.

- The number of LRIs is unlimited but there are no multiple images at the same place. This issue will be discussed in section 4.
- The camera is not allowed to rotate along its optical axis. For most applications, this restriction is adequate since the camera movement is still quite free.
- The PSF is a box function, namely

$$PSF(\mathbf{x}) = \begin{cases} S^{-2}, & |x| \le S/2, |y| \le S/2, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

where $\mathbf{x} = (x, y)$ and S is the size of the LRP. Such PSF weights HRPs according to their portions of area inside the LRPs. More general PSF will be discussed in section 6.

With these assumptions, camera movement can be regarded as locally translational on a plane parallel to the image plane, and the sizes of the LRPs are identical.

3.3.2 The Limit of Super-resolution

Suppose we want to raise the number of pixels in the ROI to $N_h \times N_h$. Let the magnification factor be $B > 1$, which can be decomposed into integral and fractional parts: $B = M + \varepsilon$ $(0 \le \varepsilon < 1)$. If the most conservative estimation of the RHS is still larger than $\delta_h N_h$, then it is quite possible that $||\delta\mathbf{H}||$ is also close to $\delta_h N_h$. For a particular $\hat{\delta}_h = 128$, this means that in average every component of $\delta\mathbf{H}$ takes a value of 128 or -128. A large deviation can imply that the details in the real solution are totally lost so that the computed solution is overly smooth, or the computed solution contains details that are quite different from those in the real solution, or even the profile of the super-resolved image is quite different from the real one. Consequently, it is desirable that the RHS is less than $\hat{\delta}_h N_h$, or

$$||\mathbf{P}^+||[||\delta\mathbf{P}||(||\mathbf{H}||+||\mathbf{P}^+||||\mathbf{r}||+\hat{\delta}_h N_h)+||\delta\mathbf{E}||] < \hat{\delta}_h N_h. \quad (6)$$

Obviously, $||\mathbf{P}^+||^{-1} = \min_{||\mathbf{x}|| \ne 0} ||\mathbf{P}\mathbf{x}||/||\mathbf{x}||$ and is upper-bounded by $G(B)N_l/N_h$, where $G(B)$ is some function of B and N_l is the square root of the number of LRPs considered in the system. In [2], by setting the value of pixel (p, q) of the HRI to be $(-1)^{p+q}$, $G(B)$ is roughly estimated as B^{-2}. In our much finer analysis, we set the value to be ω_i^{p+q} $(i = 1, 2, 3)$ instead, where ω_i are the solutions of

$$\sum_{k=0}^{M-1} \omega^k + \varepsilon\omega^M = 0, \ \sum_{k=0}^{M-1} \omega^k = 0, \text{ and } \sum_{k=0}^{M} \omega^k = 0,$$

respectively. Our better estimate is $G(B) = g^2(B)B^{-2}$, where

$$g(B) = \begin{cases} 1 - \varepsilon, & \text{if } M = 1 \\ \min(\varepsilon|1 - \omega_1^{M+1}||\omega_1|^{-\frac{M}{2}}, \varepsilon, 1 - \varepsilon), & \text{if } M > 1 \end{cases}$$

The curve of $g(B)$ is shown in Figure 1 (a), where ω_1 should be chosen among the solutions in order to minimize $|1 - \omega^{M+1}||\omega|^{-\frac{M}{2}}$.

On the other hand, $||\delta\mathbf{P}||$ is related to the registration error and B. Our analysis gives $||\delta\mathbf{P}|| \ge \delta_p N_l/N_h$, where

$$\delta_p = \Delta_r^2/4 \left[M(1 + M)^{-(1+\frac{1}{M})}\right]^2,$$

in which Δ_r is the maximum registration error and we assume that the registration error is uniformly distributed between 0 and $B\Delta_r$. Detailed proofs of the above inequalities are lengthy, hence are omitted due to page budget.

Suppose

$$||\delta\mathbf{E}|| = \delta_e N_l, \ ||\mathbf{r}|| = \delta_r N_l, \ ||\mathbf{H}|| = \sigma_h N_h,$$

from (6), we obtain

$$B < f(B) \equiv g(B)\sqrt{\frac{2\hat{\delta}_h}{\alpha + \sqrt{\alpha^2 + 4\hat{\delta}_h\delta_r\delta_p}}}, \quad (7)$$

where $\alpha = \delta_e + (\sigma_h + \hat{\delta}_h)\delta_p$.

The above inequality demarcates the existence intervals for "effective" magnification factors. Those that fall outside these intervals can still produce resolutions higher than that of low resolution images, but may not be efficient in resolution improvement in that they require stronger smoothness regularization and hence waste the effort of increasing pixel numbers. At this time, using smaller magnification factors is more economical.

Now let us estimate the parameters in (7) from a practical point of view.

1. In practice eighth-pixel registration accuracy is the limit. Hence $\Delta_r \ge 0.125$.
2. For σ_h, since globally shifting the intensity does not affect $||\delta\mathbf{H}||$, it should be estimated by the variation of the super-resolved image. Estimating it to be $\sigma_h \ge 15$ is quite conservative.
3. For δ_r, we may estimate it to be $\delta_r \ge 0.5$.

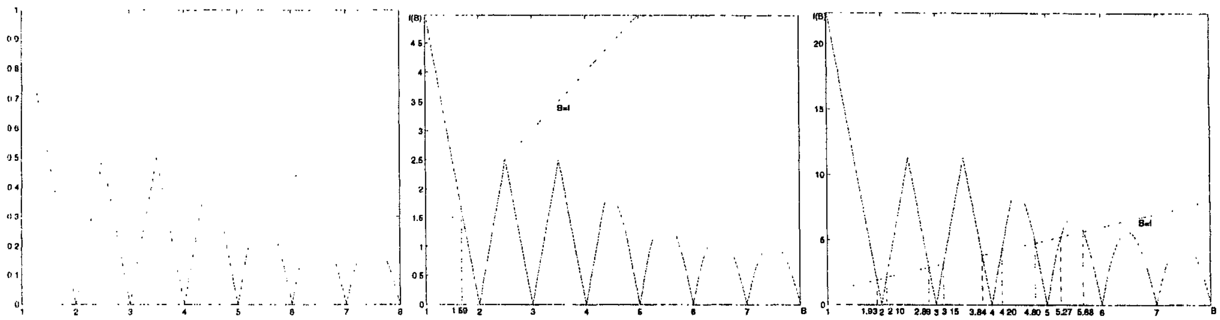
δ_e is nothing but the root mean square error (RMSE) after denoising. If the noise removal is unsuccessful, say $\delta_e \ge 5.0$, then with other practically optimal parameters, B must be in the interval where the curve of $f(B)$ is above the line $B = f$ in Figure 1(b), namely $B < 1.59$. $B = 2.5$ looks possible but can be easily ruled out by a bit more careful treatment in estimating $||\mathbf{P}^+||$.

Inequality (7) indicates that with small registration error, denoising is crucial to improve the resolution. In practice an RMSE$\le$5.0 is not undemanding for denoising, because:

- With a relatively high level of noise, noise removal cannot be very effective without strong *a priori* knowledge[2] on the image content and the noise type.
- Equation (1) is correct only in the radiometric domain, but all existing algorithms treat it in the greylevel domain directly.

However, one may try B =2.5 if he is sure that the noise level is quite low or he can remove the noise at an RMSE lower than 5.

[2]Otherwise, recognition techniques will replace super-resolution.



(a) (b) (c)

Figure 1: The plot of $g(B)$ and the intervals that B distributes. In (b) and (c) B must be in the intervals where the curve of $f(B)$ is above the line $B = f$. (a) The curve of $g(B)$. (b) Under practical situations, $1 < B < 1.59$. (c) For synthetic data, $B \in (1, 1.93) \cup (2.10, 2.89) \cup (3.15, 3.84) \cup (4.20, 4.80) \cup (5.27, 5.68)$.

In the case of synthetic data, the data may be noiseless except for the quantization error, and the registration is perfect ($\Delta_r = 0$). We may view the quantization error as uniformly distributed between 0 and 0.5; therefore, we may assume δ_e to be its expectation value 0.25, and the residual error is $\delta_r = 0.25$. With these parameters, B must be in the disjoint intervals shown in Figure 1(c), namely $B \in (1, 1.93) \cup (2.10, 2.89) \cup (3.15, 3.84) \cup (4.20, 4.80) \cup (5.27, 5.68)$.

4 The Sufficient Number of LRIs

It is well known that a convex linear combination (i.e., the coefficients are all non-negative) suppresses noise. Therefore, if an LRP is a convex combination of others, namely its coefficient vector is a convex linear combination of those of others, then the chance that its intensity is more accurate than the correspondingly linearly combined intensity is small, and it is quite unlikely that considering this LRP will improve the LSE solution. In this viewpoint, if we have already captured the vertices set of the set of all LRPs, whose convex hull includes all LRPs, then the chance that additionally more LRPs produce better HRPs is relatively small.

We define the relative displacement (RD) of an LRP, which occupies an area $[x, x+B] \times [y, y+B]$, to be $(x_\varepsilon, y_\varepsilon)$, where x_ε and y_ε are the fractional parts of x and y, respectively. It is easy to see that the LRPs with RDs at $(0,0)$, $(0, 1-\varepsilon)$, $(1-\varepsilon, 0)$ and $(1-\varepsilon, 1-\varepsilon)$ form the vertices set. Unfortunately, in practice, we cannot capture pixel by pixel. Instead, a set of LRPs that forms an LRI is acquired at the same time and in general, the RDs of these LRPs are quite different except for some special B. Therefore, we cannot obtain the vertices set conveniently and economically for an arbitrary B.

Fortunately it is not the case for B whose fractional part is $\varepsilon = 0.5$: one only need to shift the camera $(2B)^2$ times to make it run through an area of LRP, using a stepsize that equals half the size of the HRP. The number of LRIs is thus $4B^2$. Furthermore, as we see in Figure 1 (a), choosing such a B makes the system most stable and error-resistant when one wants to use a magnification factor that is greater than 2. Such magnification factor has good balance between improving the resolution and making the system stable.

5 Experiments

For the sake of visual comparison, in order to produce sharp HRIs, we only use the box function to filter a large image (Figure 2(a)) to obtain many LRIs (Figure 2(b)). Moreover, no noise except the quantization error is introduced to the LRIs and the registration is exact. Figures 2(c) through (n) are super-resolved images at various magnification factors, from B =1.5 to B =3.0. They are all scaled by bi-cubic interpolation to the size of Figure 2(n) for better visual comparison. The smoothness regularization is not applied to Figures 2(c)$\sim$(e) and (i)$\sim$(l), which corresponds to B =1.5, 1.8, 1.9, 2.15, 2.3, 2.5, 2.8, respectively, since the noise therein is unnoticeable. However, MAP regularization is applied to Figures 2 (f)$\sim$(h), (m) and (o) since the noise becomes unacceptable. The MAP algorithm follows [5] and [6]. We see that the resolution increases from B =1.5 (Figure 2(c)) to B =1.9 (Figure 2(e)), and does not increase until B =2.15 (Figure 2(i)). Then the resolution increases again until B =2.8 (Figure 2(l)) and then stops increasing again (Figures 2(m) and (n)). When B is outside those intervals, noise removal demands stronger smoothness regularization, hence the resolution is not further improved. When B is inside those intervals, we see that a larger B produces higher resolution. This phenomenon will persist for larger B, but the existence intervals we find may not be so accurate. This verifies the distribution intervals of B derived in section 3.3.2.

When noise increases, the existence intervals of B will shrink and the B's with $\varepsilon = 0.5$ (such as 2.5, 3.5) are the last to survive in the same interval. Figures 3(a)$\sim$(c) attest to this by adding small amount of uniform noise in the LRIs. We see that Figure 3(c) immediately becomes random noise and the noise levels in Figure 3(a) and (b) are the same. Note that they use the same number of LRIs and larger B should require more LRIs. Therefore, $\varepsilon = 0.5$ makes B

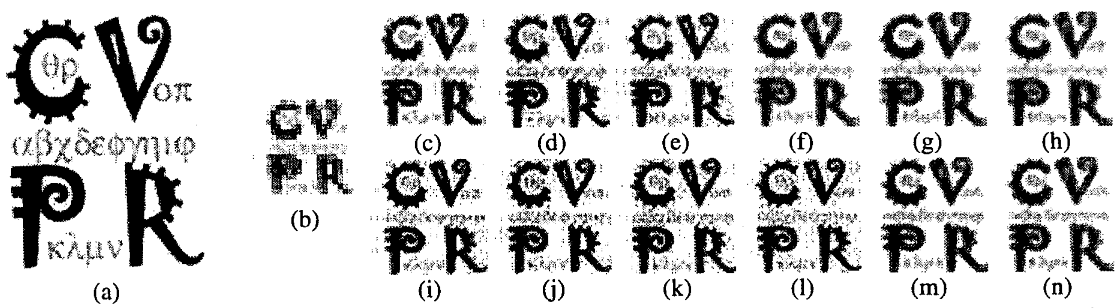


Figure 2: The images in the synthetic experiment. (a) The ultra-high resolution image for acquiring low resolution images. (b) One of the low resolution images. (c) The high resolution image at B =1.5. (d) B =1.8. (e) B =1.9. (f) B =1.98. (g) B =2.0. (h) B =2.02. (i) B =2.15. (j) B =2.3. (k) B =2.5. (l) B =2.8. (m) B =2.94. (n) B =3.0. Images (c)$\sim$(e) and (i)$\sim$(l) are not regularized. Images (f)$\sim$(h), (m) and (n) are regularized by MAP in order to remove the noise.

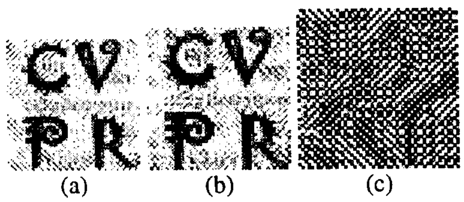


Figure 3: The comparison of error resistibility of different fractional parts. Uniform noise on [0,0.5] is added to the LRIs. These high resolution images all use 36 LRIs. (a) B =2.3; (b); B =2.5; (c), B =2.8.

most noise-resistant when $B > 2$. Furthermore, we also try to use the vertices set solely. Figures 4(a), (c) and (f) are the HRIs reconstructed without regularization from the vertices sets that correspond to B =2.3, 2.5 and 2.8, respectively. It can be seen that their resolutions can be deemed to be identical to those of Figures 4(b), (d) and (g) respectively, which are reconstructed from 36 full images. Figure 4(e) is reconstructed from exactly 25 full images. Figure 4(d) and (e) are also of the same resolution. These testify to our conclusions in section 4.

To carry out real experiments, we utilize a computer-controllable vertical XY-table shown in Figure 5(a), where a monochromic CCD camera (white box in Figure 5(a)) is attached vertically to the XY-plane. The camera can be moved at a stepsize of 0.025mm and the error is below 1%. In our experiments, we place a portrait (Figure 5(b)) about 3 meters away from the camera. When the camera moves 10,000 steps in both x and y directions, the disparities are 122 ± 1 and 113 ± 1 pixels, respectively. Therefore, we can register the images at a higher accuracy than any existing registration algorithms, such as [9]. On the other hand, multiple images (20 images in this experiment) can be captured at the same place in order to remove noise by simple averaging. Hence we waive the vision and image processing techniques for registration and denoising.

We set the gamma of the CCD camera to be 1 and captured 12×12 images (Figure 5(c)) for the portrait, in which most of the details are lost. After super-resolution using

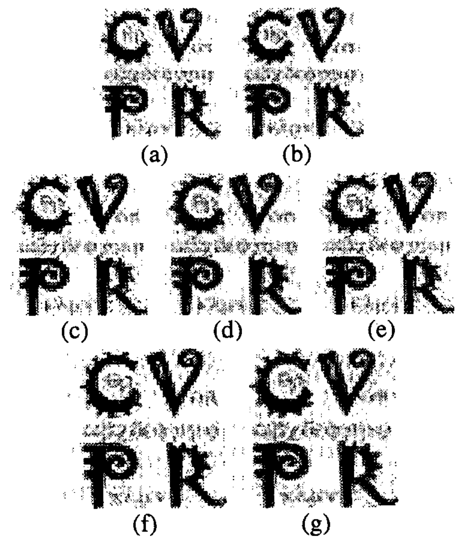


Figure 4: The comparison of super-resolution using only the vertices set and using many whole images. (a), (c) and (f) are the high resolution images using the vertices set only, while (b), (d) and (g) are the high resolution images from 36 whole images. (e) uses exactly 25 images. For (a) and (b), B =2.3; for (c)$\sim$(e), B =2.5; for (f) and (g), B =2.8.

the box PSF and MAP regularization, we obtain HRIs of the face. Figure 5(e) is the HRI at magnification factor 1.5. Figure 5(f) is reconstructed from the vertices set only. Since the available LRPs may not be exactly at RDs 0 or 0.5, we in fact pick out all the pixels whose RDs are close to 0, 0.5 or 1.0. We also try super-resolution at magnification factors 2 (Figure 5(g)) and 2.5 (Figure 5(h)). We see that Figure 5(e)$\sim$(h) are indistinguishable and they are a little sharper than the low resolution images (Figure 5(d)). Therefore, the resolution is not further improved for $B > 1.6$.

6 Conclusions and Discussions

We analyze reconstruction-based algorithms for super-resolution and give explicit limits under both practical and

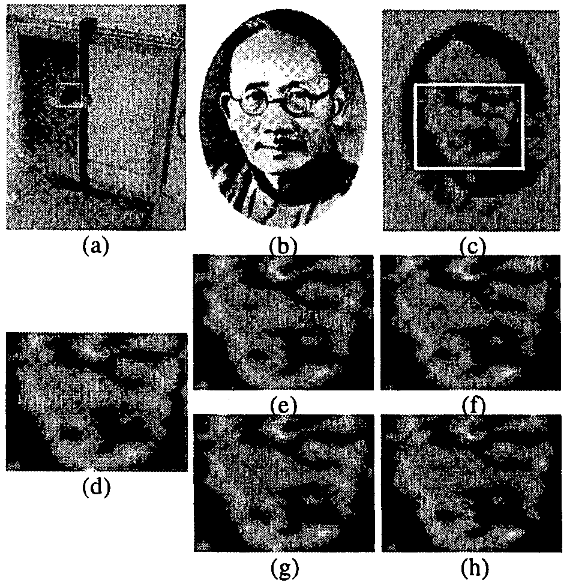


Figure 5: The device and images in the real experiment. (a) The vertical XY-table and the CCD camera (white box) in our experiments. (b) A portrait. (c) One of the low resolution images captured and the region of interest (white box). (d) Blow-up of the region of interest. (e) Super-resolution of the face area at $B = 1.5$ by MAP. (f) Super-resolution of the face area at $B = 1.5$ by MAP, using the vertices set only. (g) $B = 2.0$. (h) $B = 2.5$. (d)$\sim$(g) are enlarged to the size of (h) by bi-cubic interpolation.

synthetic conditions. Under practical conditions, the limit is found to be 1.6. While under synthetic conditions, it is 5.7. The highest resolution must be achieved below the limits. Hence larger magnification factors are wasteful. Moreover, for synthetic data, the effective magnification factors can only distribute on some disjoint intervals. We also find that if the vertices set has been captured, then adding more LRIs will not further improve the resolution. Finally, we conclude that the optimal fractional part of the magnification factor is 0.5 and the corresponding sufficient number of LRIs is four times the squared magnification factor. Our experiments have verified all these conclusions.

In our proof, we assume that the PSF is a box function. A general PSF can be decoupled into two parts [2]: $PSF = LPSF * SPSF$, where $LPSF$ and $SPSF$ are the PSFs of the lens and the sensor on the image plane, respectively. For motion deblurring, the $LPSF$ should also be convolved in time. Because of the localness assumption, it can be assumed to be shift-invariant. The $SPSF$ is the box function defined in (5). For most applications, recovering $U(\mathbf{x}) = LPSF * H(\mathbf{x})$ is more fundamental since estimating the $LPSF$ is not an easy task. More importantly, we may compute $U(\mathbf{x})$ first and then deconvolve it to obtain $H(\mathbf{x})$. However, unless the $LPSF$ is effectively super-resolved at higher magnification factors, after deconvolution larger magnification factors will not produce higher resolution than smaller ones do. Consequently, we need not take the $LPSF$ into consideration. As a result, our results are in fact quite general.

Among the four assumptions we make, the one that disallows rotation of the camera along the optical axis is essential for our analysis since it will destroy the separability of the $SPSF$ and make the estimate on $||\mathbf{P}^+||$ and $||\delta\mathbf{P}||$ much more difficult. We believe that allowing such rotation may make the limits larger but they are still bounded. We are working on proving this. Moreover, in recent years component-wise perturbation theory [7] has been developed for more accurate estimation. We hope this novel theory could improve our results.

References

[1] S. Baker and T. Kanade, "Hallucinating Faces," Technical Report CMU-RI-TR-99-32, The Robotics Institute, Carnegie Mellon University, 1999.

[2] S. Baker and T. Kanade, "Limits on Super-resolution and How to Break Them," *Proceedings of CVPR'2000*, Hilton Head, South Carolina, 2000, pp. 372-279.

[3] S. Borman and R.L. Stevenson, "Spatial Resolution Enhancement of Low-Resolution Image Sequences: A Comprehensive Review with Directions for Future Research," *Technical Report*, University of Notre Dame, 1998.

[4] N.K. Bose, H.C. Kim, and H.M. Valenzuela, "Recursive Implementation of Total Least Squares Algorithm for Image Reconstruction from Noisy, Undersampled Multiframes," *Proceedings of ASSP'93*, Minneapolis, MN, vol. 5, pp. 269-272.

[5] M. Elad and A. Feuer, "Restoration of Single Super-resolution Image from Several Blurred, Noisy and Down-sampled Measured Images," *IEEE Trans. on Image Processing*, Vol. 6, no. 12, pp. 1646-58, 1997.

[6] R.C. Hardie, K.J. Barnard and E.E. Amstrong, "Joint MAP Registration and High-resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Trans. on Image Processing*, Vol. 6, No. 12, pp. 1621-33, 1997.

[7] N.J. Higham, "A Survey of Componentwise Perturbation Theory in Numerical Linear Algebra," *Mathematics of Computation 1943–1993: A Half Century of Computational Mathematics, volume 48 of Proceedings of Symposia in Applied Mathematics (W. Gautschi eds.)*, pp. 49-77, American Mathematical Society, Providence, RI, USA, 1994.

[8] T.S. Huang and R. Tsai, "Multi-frame Image Restoration and Registration," *Advances in Computer Vision and Image Processing*, Vol. 1, pp. 317-339, 1984.

[9] M. Irani and S. Peleg, "Improving Resolution by Image Restoration," *Computer Vision, Graphics, and Image Processing*, Vol. 53, pp. 231-239, 1991.

[10] R. Schultz and R. Stevenson, "Extraction of High-resolution Frames from Video Sequences," *IEEE Trans. on Image Processing*, Vol. 5, No. 6, pp. 996-1011, 1996.

[11] S. Xu, *The Theory and Methods of Matrix Computation (in Chinese)*, Peking University Press, 1995.