Super-Resolution

Neil Alldrin[*]
University of California, San Diego

Abstract

Super-resolution is the process of recovering a high-resolution image from multiple low-resolution images of the same scene. An overview of existing super-resolution techniques is provided. This includes the formulation of an observation model and coverage of the dominant algorithms – frequency domain methods, deterministic regularization, and stochastic techniques. Also covered is recent work on the limits of super-resolution and a section on potential future directions for super-resolution algorithms.

1 Introduction

Super-resolution, loosely speaking, is the process of recovering a high-resolution image from a set of low-resolution input images. Such algorithms have long been portrayed in movies and television; a typical movie scene showing a computer operator repeatedly zoom in on a person's face or a license plate where the missing high-resolution detail magically appears on the computer screen after each successive zoom (see figure 1). Clearly this is pure fiction, after all there are an infinite number of higher-resolution images that could form the original low-resolution image, right? With a single image and no *a priori* knowledge, this is true; however, higher-resolution content *can* be recovered – to a point – if multiple low-resolution images are available of the same scene from slightly different poses. A closely related problem is that of *image restoration*, which utilizes *a priori* knowledge of the scene to recover missing detail from a *single* image[1]. So, while Hollywood almost always greatly exaggerates what can realistically be done, there actually is some merit to the "zoom in and enhance" scenario commonly depicted.

Applications for super-resolution abound. NASA has been using super-resolution techniques for years to obtain more detailed images of planets and other celestial objects. Closer to home, super-resolution can be used to enhance surveillance videos to more accurately identify objects in the scene. One particular example of this are systems capable of automatically reading license plate numbers from severely pixelated video streams. Another application is the conversion of standard NTSC television recordings to the newer HDTV format which is of a higher resolution.

A variety of approaches for solving the super-resolution problem have been proposed. Initial attempts worked in the frequency domain, typically recovering higher frequency components by taking advantage of the shifting and aliasing properties of the Fourier transform. Deterministic regularization approaches, which work in the spatial domain, enable easier inclusion of *a priori* constraints on the solution space (typically with a smoothness prior). Stochastic methods have received the most attention lately as they generalize the deterministic regularization approaches *and* enable more natural inclusion of prior knowledge. Other approaches include non-uniform interpolation, projection onto convex sets, iterative back
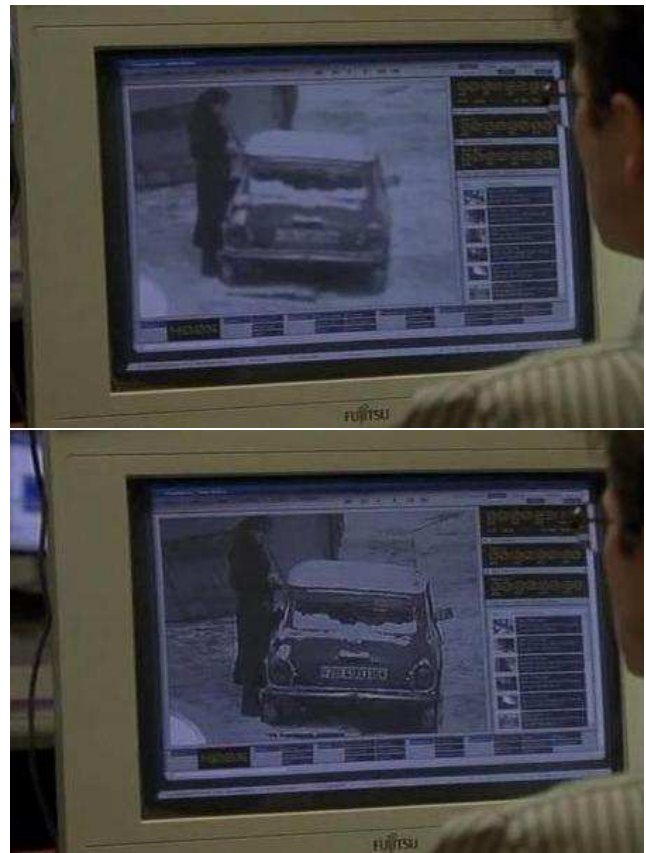
Figure 1: Two frames from the movie *The Bourne Identity* showing the enhancement of a surveillance video.

projection, and adaptive filtering. With the increased emphasis on stochastic techniques has also come increased emphasis on learning priors from from example data rather than relying on more heuristically derived information.

The following sections will hopefully serve to elucidate the super-resolution problem. We start by introducing a model for image formation, then cover various approaches to the super-resolution problem – frequency domain techniques, non-uniform interpolation, deterministic regularization, stochastic methods, projection onto convex sets, and iterative back projection. Next we cover recent work attempting to determine limits to super-resolution techniques. Finally, we conclude with a section on potential future research directions.

1.1 A Very Basic Super-Resolution Algorithm

To motivate the super-resolution problem in a more concrete manner, suppose we have have a camera with finite resolution $M \times N$, but we desire an image of higher resolution $2M \times 2N$. What can we do? One option is to buy a new camera, but this could get expensive or might not be possible. Another option is to take four images

with our existing camera, each image offset by half a pixel from the other images (see figure 2). If we now combine the images on a more dense grid, we have the $2M \times 2N$ image we desire. This of course neglects many things, most importantly the spatial averaging that occurs over each pixel, but in theory would work well.
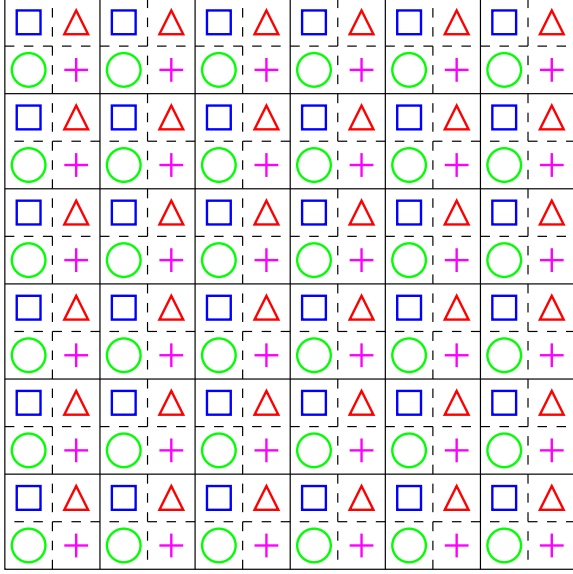

Figure 2: Higher density sampling resulting from four translated images. Blue squares correspond to pixels from the first image, red triangles the second, green circles the third, and purple pluses the fourth.

2 The Imaging Process and Motion Models

One of the most important parts of designing a super-resolution algorithm is choosing a good model for the imaging and motion process. Ultimately we are trying to remove various camera distortions – primarily decimation – from our input images. Therefore, at the very least, we need to model the decimation process. Other camera distortions that may be modeled are lens and motion blur, geometric distortions, and quantization. Many motion models have been proposed, ranging from simple translational offsets to complex non-linear transformations. In this survey, we use a few different observation models depending on the needs of the algorithm being reviewed.

2.1 Continuous Formation Model

Because light is continuous[2], it makes sense to model the formation of images in terms of the continuous spatial domain. The distortions arising from the camera system can be approximated by a point spread function (implying a linear, space invariant process) which is convolved with the image formed by an ideal pinhole camera. A formulation of the image formation equation under these assumptions is,

$$y_k(m) = (f_k * PSF_k)(m) = \int f_k(x)PSF_k(x-m)dx \qquad (1)$$

[2]Quantization of energy and particle nature of light aside.

Symbol	Definition
$f_k(x)$	kth continuous image formed by ideal pinhole camera
$y_k(m)$	kth observed (low-resolution) image
$g(p)$	Super-resolution image (what we want to recover)
$g(z)$	Continuous image corresponding to $g(p)$
$r_k(x)$	Maps kth image coordinates to SR image coordinates
$PSF_k(x)$	Point-spread function
K	Number of observed images
MN	Number of pixels in the observed LR images
PQ	Number of pixels in the SR image
L	Magnification factor

Table 1: Common symbols.

where $f_k(x)$ is the kth ideal image, $y_k(m)$ is the observed (low-resolution) image, $PSF_k(x)$ is the point spread function, $x \in \mathbb{R}^2$ represents real-valued coordinates on the image plane, and $m \in \mathbb{Z}^2$ represents pixel coordinates on the image plane for the observed images.

Now suppose we want to recover a high-resolution image, $g(p)$, corresponding to some ideal image $g(z)$. Also consider a registration function $r_k(x)$ that maps coordinates of $g(z)$ onto the coordinate frame of the kth image. Then we can rewrite equation 1 as,

$$y_k(m) = \int f_k(r_k(z))PSF_k(r_k(z)-m)\left|\frac{dr_k}{dz}\right|dz \qquad (2)$$

where z denotes coordinates of the super-resolution image we wish to recover and $\left|\frac{dr_k}{dz}\right|$ is the determinant of the Jacobian of r_k.

To prevent confusion, note that we are using a little notational freedom with the coordinates of g. Wherever $g(p)$ is used we are referring to the pixel coordinates of g ($p \in \mathbb{Z}^2$) and wherever $g(z)$ is used, we are referring to the real coordinates of g ($z \in \mathbb{R}^2$).

One further simplification can be made to equation 2; if we assume the registration is correct for all images, then $f_k(r_k(z)) = g(z)$ for all k. We can now state,

$$y_k(m) = \int g(z)PSF_k(r_k(z)-m)\left|\frac{dr_k}{dz}\right|dz. \qquad (3)$$

2.2 Discretization and Noise

The next step is to discretize equation 3. We do this by replacing the continuous image $g(z)$ with the pixelized version, $g(p)$,

$$y_k(m) \simeq \sum_p \left[g(p) \cdot \int_p PSF_k(r_k(z)-m)\left|\frac{dr_k}{dz}\right|dz\right] \qquad (4)$$

where the integration is performed over each pixel p in g. This can be expressed more compactly as,

$$y_k(m) \simeq \sum_p W_k(m,p) \cdot g(p) \text{ where} \qquad (5)$$

$$W_k(m,p) = \int_p PSF_k(r_k(z)-m)\left|\frac{dr_k}{dz}\right|dz. \qquad (6)$$

We can further simplify the system using matrix notation,

$$y_k = W_k g. \qquad (7)$$

W_k is sometimes referred to as the warp matrix and provides a mapping from the high-resolution image to the kth low-resolution image.

Additive noise, η, is also easily added to the system of equations, yielding

$$y_k(m) \simeq \sum_p W_k(m,p) \cdot g(p) + \eta_k(m), \text{ and} \qquad (8)$$

$$y_k = W_k g + \eta_k. \qquad (9)$$

2.3 The Point Spread Function

The point spread function $PSF_k(x)$ can be decomposed as

$$PSF_k(x) = (w_k * a_k)(x) \qquad (10)$$

where $w_k(x)$ accounts for optical blurring and $a_k(x)$ accounts for the spatial integration over each pixel. w_k is usually further decomposed, but we omit the details here. One common form for a_k is a box function,

$$a_k(x) = \left\{ \begin{array}{ll} \frac{1}{S^2} & \text{if } |x| \leq S/2 \text{ and } |y| \leq S/2 \\ 0 & \text{otherwise} \end{array} \right\} \qquad (11)$$

where S is the length of a side of the box. It may be helpful to think of this as a square shaped photo-receptor that sums all the light arriving at its surface. Other common models for a_k are impulse and Gaussian functions, although these are not realistic for CCD sensors and are used primarily for mathematical convenience.

2.4 Registration

Registration, r_k, is often assumed to be known *a priori* in the super-resolution literature. Because registration is such a heavily researched topic, it is beyond the scope of this survey to cover it in full detail. However, it should be noted that equations 2 and 3 make the implicit assumption that r_k is invertible. Unfortunately, this is not a reasonable assumption for general camera motions (due to things like occlusion, failure of the constant brightness assumption, etc). While this can be overcome to some degree by considering the registration functions to be locally invertible, it generally means that super-resolution algorithms work best with input images taken from about the same pose.

3 Frequency Domain Approaches

The super-resolution problem was posed, along with a frequency domain solution, by Tsai and Huang[Tsai and Huang 1984]. Prior to their paper, interpolation was the best technique for increasing the resolution of images. Tsai and Huang showed that with multiple offset images of the same scene – and proper registration – restoration better than cubic spline interpolation could be achieved. Their motivation was to improve spatial resolution of satellite images of earth, where a large set of translated images of the same scene are available.

3.1 Theory

In one dimension, Tsai and Huang's original formulation of the super-resolution problem considered each low-resolution image, y_k,

as a discrete uniform sampling of some unknown continuous image, $g(x)$, offset by some amount δ_k. If we denote the sample spacing by T, then pixel i in the kth low resolution image can be written in terms of the original image as

$$y_{ki} = g(iT + \delta_k) \qquad (12)$$

where $k \in \{1, ..., K\}$ and $i \in \{1, ..., P\}$; K and P being the number of images and number of pixels per image respectively[3].

Now define $y_k(x) = g(x + \delta_k)$. Using the shifting property of the continuous Fourier transform (CFT), we have

$$Y_k(\omega) = e^{j\delta_k \omega} G(\omega), \qquad (13)$$

where $Y_k(\omega)$ and $G(\omega)$ are the CFT of $y_k(x)$ and $g(x)$ respectively. Letting Y_{kn} be the discrete Fourier transform (DFT) at discrete frequency n of the kth frame f_{ki}, by definition of the DFT we have

$$Y_{kn} = \sum_{i=0}^{N-1} y_{ki} e^{-j2\pi \frac{in}{N}}, n = 0, ..., N-1. \qquad (14)$$

From the aliasing relationship between the DFT and CFT we have

$$Y_{kn} = \frac{1}{T} \sum_{m=-\infty}^{\infty} Y_k(\frac{n}{NT} + m\omega_s), \qquad (15)$$

where ω_s is the sampling frequency of each image. If we assume the original image is band-limited, then $G(\omega) = 0$ for $|\omega| \geq L\omega_s$ for some L and we can write a matrix equation for each discrete frequency n,

$$Q_n = \phi_n G_n, \qquad (16)$$

where $Q_n \in \mathbb{C}^{K \times 1}$ with kth element Y_{kn}; $G_n \in \mathbb{C}^{2L \times 1}$ with ith element $G(\frac{n}{NT} + (i - L - 1)\omega_s)$; $\phi_n \in \mathbb{C}^{K \times 2L}$ with (k, j)th element $T^{-1} \exp\{j2\pi\delta_{k-1}\left(\frac{n}{NT} + (j - L - 1)\omega_s\right)\}$.

When δ_k is known, equation 16 is independent for different n and can be solved separately for the unknown column vector G_n, which contains the CFT of $g(x)$ at $2L$ equidistant frequency points with spacing ω_s. After this is done for each n we have an estimate of $G(\omega)$ at LN frequency points ranging from $(-L\omega_s)$ to $(L\omega_s - \frac{1}{NT})$ with spacing $\frac{1}{NT}$. From this we can estimate $g(x)$ at a resolution increased by a factor of $2L$.

As stated, we are solving a $K \times 2L$ matrix equation for each frequency component. By taking advantage of the structure of ϕ_n we can vastly reduce the complexity of the problem as well as prove some nice properties of the system. First note that ϕ_n can be decomposed as

$$\phi_n = D_n H, \qquad (17)$$

where D_n is a diagonal matrix with kth diagonal element

$$T^{-1} \exp\left\{j2\pi\delta_{k-1}\left(\frac{n}{NT} - L\omega_s\right)\right\} \qquad (18)$$

and $[H]_{ij} = \exp\{j2\pi\delta_i(j-1)\omega_s\}$. Using equation 16 we can write

$$D_n^{-1} Q_n = H G_n \qquad (19)$$

where D_n^{-1} is a diagonal matrix with kth diagonal element

$$T \exp\left\{-j2\pi\delta_{k-1}\left(\frac{n}{NT} - L\omega_s\right)\right\}. \qquad (20)$$

---

[3]The derivations in this section are taken from [Tsai and Huang 1984].

Turning our attention to matrix H, if we let $W_i = e^{j2\pi\delta_{i-1}\omega_s}$ then H is a Vandamonde matrix of the form,

$$
\begin{pmatrix}
W_1^0 & W_1^1 & W_1^2 & \cdots & W_1^{2L-1} \\
W_2^0 & W_2^1 & W_2^2 & \cdots & W_2^{2L-1} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
W_K^0 & W_K^1 & W_K^2 & \cdots & W_K^{2L-1}
\end{pmatrix}
\tag{21}
$$

which means the rows of H are linearly independent so long as the offsets in each image satisfy $\delta_i \neq \delta_j + mT$, where $i \neq j$ and m is any integer. If this holds and $K \geq 2L$, then solving for G_n is well-posed. Another benefit of this decomposition is to note that H is independent of n, meaning its pseudo-inverse only needs to be computed once.

So far we have only been considering the one dimensional case. If we extend the analysis to two dimensions, the results are almost identical. Using the 2D Fourier transforms instead of the 1D Fourier transforms, we arrive at a system analogous to equation 16,

$$
Q_{mn} = \phi_{mn}G^{mn}.
\tag{22}
$$

We can also do a similar decomposition of ϕ_{mn},

$$
\phi_{mn} = D_{mn}H
\tag{23}
$$

so that

$$
D_{mn}^{-1}Q_{mn} = HG_{mn}.
\tag{24}
$$

Like the 1D case, H is independent of m and n so it's pseudo-inverse only needs to be computed once for all frequencies m and n. Unlike the 1D case, however, well-posedness is no longer guaranteed even if no image coincides with another. The chance of this happening is low, so the technique still generally works well.

4 Spatial Domain Approaches

Approaching the super-resolution problem in the frequency domain makes a lot of sense because it is relatively simple and computationally efficient. However, there are some problems with a frequency domain formulation. For one, it restricts the inter-frame motion to be translational because the DFT assumes uniformly spaced samples. Another disadvantage is that prior knowledge that might be used to constrain or regularize the super-resolution problem is often difficult to express in the frequency domain. Since the super-resolution problem is fundamentally ill-posed[4], incorporation of prior knowledge is essential to achieve good results.

A variety of techniques exist for the super-resolution problem in the spatial domain. These solutions include interpolation, deterministic regularized techniques, stochastic methods, iterative back projection, and projection onto convex sets among others. The primary advantages to working in the spatial domain are support for unconstrained motion between frames and ease of incorporating prior knowledge into the solution.[5]

---

[4]At some magnification, the observations alone fail to contain enough information to support a unique super-resolution image.

[5]The following sections are primarily adapted from the following three papers : [Park et al. 2003; Borman and Stevenson 1998; Elad and Feuer 1997].

4.1 Non-Uniform Interpolation

Perhaps the most naive method for performing super-resolution is to map pixels from the low-resolution images onto a common plane (according to the motion model for each image) and then interpolate over a more finely sampled grid to obtain a higher-resolution image. Unfortunately, this technique generally works very poorly because of some inherent assumptions; the main problem being that camera sensors do not act as impulse functions, but instead spatially average the incident light across each pixel. Figure 3c highlights the lack of high frequency recovery using this technique. Notice that non-uniform interpolation looks only marginally better than bi-linear interpolation. Application of a deblurring kernel, however, yields nice looking results.
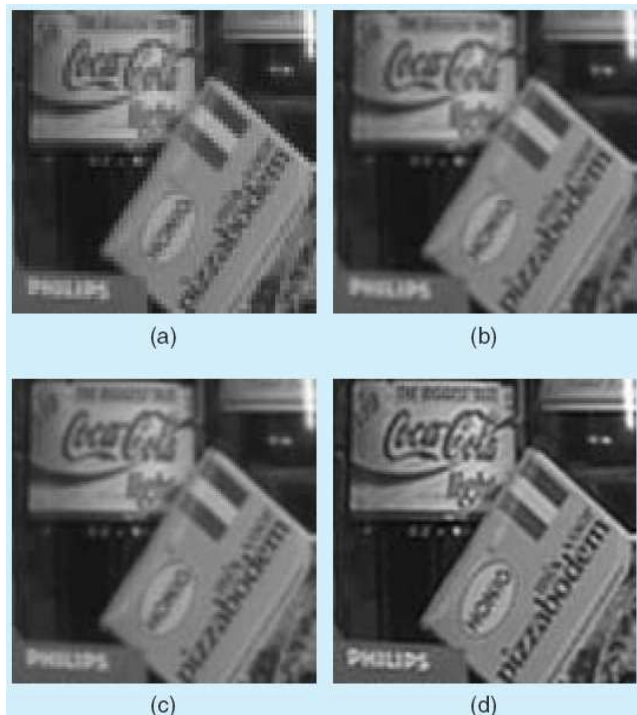

Figure 3: Results of non-uniform interpolation. (a) Nearest neighbor interpolation (over a single image). (b) bi-linear interpolation (over a single image). (c) non-uniform interpolation (over four translated input images). (d) application of a deblurring algorithm to (c). (Figure taken from [Park et al. 2003])

4.2 Deterministic Regularization

When presented with an ill-posed problem it becomes necessary to impose prior knowledge on the solution space in order to obtain a unique solution. Many standard techniques exist for doing this, but perhaps the most common approach is to impose a smoothness prior via Tikhonov regularization on top of a least-squares optimization function[Park et al. 2003]. Recall our image formation equation (equation 9),

$$
y_k = W_k g + \eta_k.
\tag{25}
$$

Assuming the matrix W_k can be estimated for each input image y_k, we can estimate the high-resolution image $g(p)$ by minimizing the

following objective function,

$$\sum_{k=1}^{K} ||y_k - W_k g||^2 + \lambda ||Cg||^2. \qquad (26)$$

Here, $C \in \mathbb{R}^{PQ \times PQ}$ encodes our prior knowledge of what the high-resolution image should look like. The most common choices for C encourage smooth solutions by penalizing high frequency spatial variations in g. λ controls how much weight is given to the regularization constraint; large values of λ will result in overly constrained solutions that may not reflect the data, while small values could result in noisy solutions depending on the characteristics of the sampling noise, η.

The presence of the regularization term guarantees a convex and differentiable optimization function so long as C is invertible. Thus, a unique optimal value of g can be computed using a number of standard methods like gradient descent.

Figure 4 shows results of the deterministic regularization technique with a smoothness prior. Note the large effect of λ on the resulting images and how it compares to a stochastic based approach (covered in the next section). The results are certainly an improvement over the low-resolution image, but enforcing smoothness is not always the best option, especially if other priors can be formulated that preserve high-frequency details better (such as the edge preserving prior in figure 4d).

To give more insight on the regularization matrix C, we derive a matrix that can be used to enforce smoothness. A typical measure of smoothness is the discrete 2D Laplacian,

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad (27)$$

which approximates $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$. We want to apply Q to each pixel in g. Since the image g has been mapped to a $MN \times 1$ vector, we simply need to make sure Q is mapped onto the rows of C appropriately. Namely, the main diagonal of C should be -4 and the rows corresponding to the pixels adjacent to the pixel corresponding to the given row should be 1; all other entries will be 0. The same methodology can be used to apply other constraints on a per-pixel basis.

The above deterministic regularization approach can be further enhanced using multichannel theory[Hong et al. 1997b; Hong et al. 1997a; Kang 1998], which essentially enables constraints *between* images to be expressed. This is usually accomplished by concatenating the input images into one long vector and using a larger warp matrix W that contains cross image terms. Some work has also been done on estimating optimal regularization parameters λ by analyzing the L-curve[Hansen and O'Leary 1993].

4.3 Stochastic Reconstruction Methods

While deterministic regularization enables the use of prior knowledge to constrain the solution, it does so in a somewhat awkward way, by tacking on a regularization term to the optimization function. Another approach to the same problem is to use statistical techniques that explicitly handle prior information and noise. If certain conditions are met, the two formulations yield the same optimization functions; however, inclusion of prior knowledge is usually more natural using a stochastic approach.

The most common statistical approach for super-resolution is the Bayesian formulation. This encompasses maximum likelihood
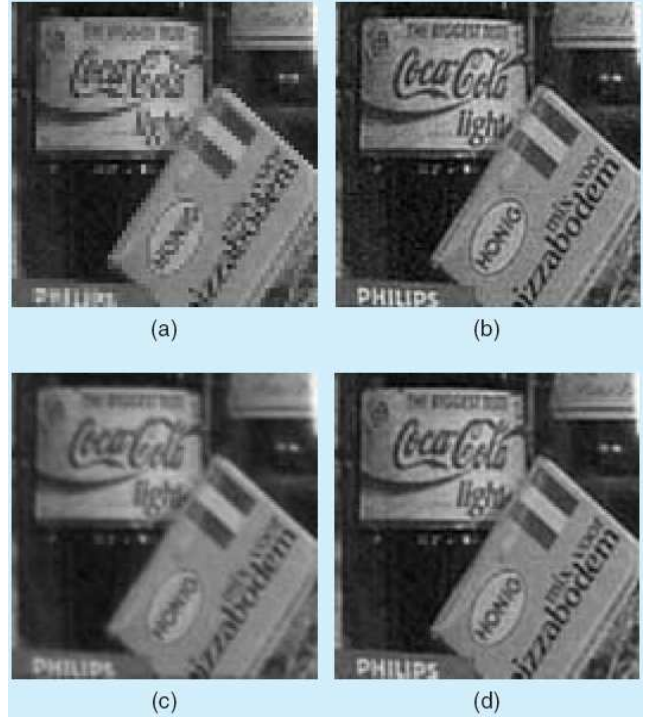

Figure 4: The effects of regularization on super-resolution. (a) Nearest neighbor interpolation (over a single image). (b) Least squares estimate with small λ. (c) Least squares estimate with large λ. (d) Maximum a posteriori technique with an edge-preserving prior. (Figure taken from [Park et al. 2003])

(ML) and maximum a-posteriori (MAP) techniques. We start with the standard observation equation,

$$y_k = W_k g + \eta_k \qquad (28)$$

where the observed data y_k, noise η_k, and super-resolution image g are assumed to be stochastic and the warp matrix W_k is known. Corresponding to the noise η_k is its autocorrelation matrix S_k. If we lack prior knowledge, we can set S_k^{-1} to the identity matrix, which corresponds to white noise. Likewise, we define Q to be the autocorrelation matrix of g.

The maximum a-posteriori approach, introduced by Schultz and Stevenson[Schultz and Stevenson 1996] to the super-resolution literature, tries to find the SR image g that maximizes the probability of the SR image given the observed images, $Pr\{g | \{y_k\}_{k=1}^{K}\}$; ie,

$$g_{MAP} = \arg\max g \left[Pr\{g | \{y_k\}_{k=1}^{K}\} \right] \qquad (29)$$

Applying Bayes rule and taking the logarithm of the result, we obtain

$$g_{MAP} = \arg\max g \left[Pr\{\{y_k\}_{k=1}^{K} | g\} \cdot Pr\{g\} \right] \qquad (30)$$

$$g_{MAP} = \arg\max g \left[\log(Pr\{\{y_k\}_{k=1}^{K} | g\}) + \log(Pr\{g\}) \right] \qquad (31)$$

where $\log(Pr\{\{y_k\}_{k=1}^{K} | g\})$ is the log-likelihood function, and $Pr\{g\}$ is the *a priori* density of g.

If we assume the noise and SR image correspond to zero mean Gaussian additive random processes with autocorrelation matrices

S and Q, then it can be shown that the MAP estimation is

$$g_{MAP} = \arg\max g \left[(Y - Wg)^T S^{-1} (Y - Wg) + g^T Q^{-1} g \right], \quad (32)$$

where we have combined the K images into a single linear system for notational convenience. Minimization yields

$$g = (W^T S^{-1} W + Q^{-1})^{-1} W^T S^{-1} Y. \quad (33)$$

This is exactly equivalent to the minimum mean squared error (MMSE) solution.

Freeman et al.[Freeman et al. 2000; Freeman et al. 2002] employ more powerful priors using Markov random fields (MRFs). The basic idea is to divide the observed images and the scene (ie, the SR image) into patches, assigning one node of a Markov network to each patch. Each scene patch is then connected to its corresponding image patch and to its spatially adjacent neighbors (see figure 5).
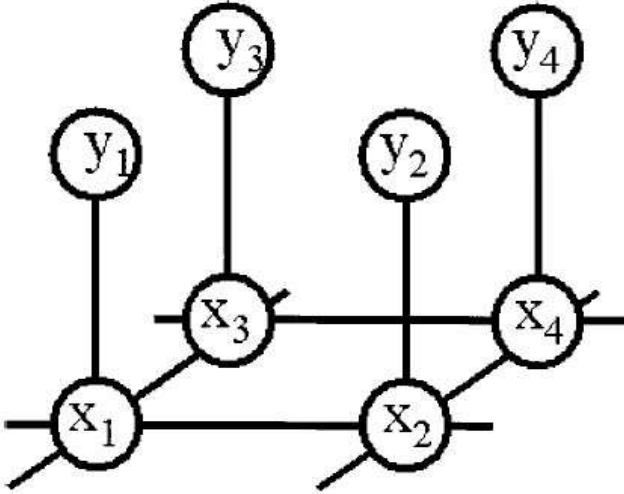
Figure 5: MRF network used by Freeman et al.

In this model, $Pr(g)$ is defined as

$$Pr(g) = \frac{1}{Z} \exp\{-U(g)\} = \frac{1}{Z} \exp\left\{ -\sum_{c \in C} \phi_c(g) \right\} \quad (34)$$

where Z is a normalizing constant, $U(g)$ is an "energy" function, $\phi_c(g)$ is a potential function that depends on pixel values within a clique (aka patch) c, and C is the set of cliques.

Freeman et al. train the network on a set of real-world photographs and achieve fairly good results (see figures 6 and 7). It should be noted that their paper only considers a single observed image and therefore should be considered an *image restoration* algorithm rather than a super-resolution algorithm. However, it highlights the power of using a more flexible prior model combined with a learning framework.

A similar patch based technique recently developed in the computer vision community is the epitome[N. Jojic and A.Kannan 2003; Cheung et al. 2005]. The most recent work enables super-resolution in videos using priors learned from previous frames in the scene. The results related to super-resolution look good, but are applied to somewhat benign scenes, so it remains to be seen whether the technique truly works well or not for super-resolution.

Capel and Zisserman[Capel and Zisserman 2001] employ a "face space" prior learned from images of human faces. Figure 8 shows

Figure 6: A restored image of a tiger from Freeman et al. (Top) Observed low-resolution image. (Bottom) Restored image.

the effect of changing the weight on their prior term. Baker and Kanade[Baker and Kanade 2002] also learn priors from images of human faces, but instead of learning entire faces, they learn a set of typical face patches. The prior term then indicates how close the images are to being formed from a set of patches. Results are shown in figure 9.

While papers that use learned constraints claim to be less ad-hoc than algorithms that use "artificial" constraints like smoothness, this is completely dependent on the training data. There could be unintended, and subtle, consequences to using training sets that are not well thought out. The main issue is that the training sets need to reflect the statistics of the images being restored without introducing bias; however, it is no simple task to evaluate whether or not a given training set satisfies this criterion. Since most learning based approaches use somewhat ad-hoc training sets, it is difficult to evaluate the algorithms and also to compare different algorithms to one another.

4.4 Other Restoration Techniques

4.4.1 Projection Onto Convex Sets

Another method for reducing the space of possible reconstructions is projection onto convex sets (POCS)[Youla 1978; Elad and Feuer 1997]. This is a set-theoretic approach where each piece of *a priori* knowledge is formulated as a constraining convex set. Once the group of convex sets is formed, an iterative algorithm is employed to recover a point on the intersection of the convex sets,

$$g_{i+1} = P_M P_{M-1} \cdots P_2 P_1 \{g_i\} \quad (35)$$
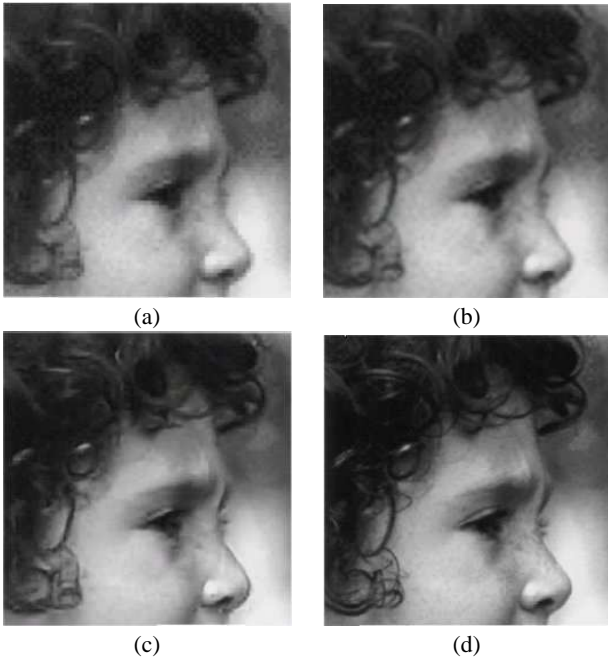
(a)   (b)

(c)   (d)

Figure 7: A restored image of a boy from Freeman et al. (a) Observed low-resolution image. (b) Bi-cubic interpolation. (c) Method of Freeman et al. (d) Actual high-res image.

where P_j is the projection of a given point onto the jth convex set and M is the number of convex sets. In essence, we are restricting the final restored image to lie on the intersection of the constraining sets, $\{P_j\}_{j=1}^M$. The reason we require *convex* sets is that convergence is guaranteed for the case where each set is convex.

One potential group of convex sets is based on the l_2 distance measure,

$$G_k = \left\{ g \mid ||W_k g - y_k||^2 \le 1 \right\}, 1 \le k \le K. \qquad (36)$$

This defines a set of ellipsoids (one for each input image) and restricts the final solution to lie inside the ellipsoids. Other possible convex sets include ones based on the l_∞ norm, those imposing smoothness, and those constraining the image intensity to be positive. Two problems with the POCS approach are that uniqueness is not guaranteed for the final recovered image and that defining the projections P_j can be difficult.

4.4.2 Iterative Back Projection

Irani and Peleg[Irani and Peleg 1991] proposed a super-resolution algorithm based on iterative back projection (IBP). The key idea is that the error between the observed low-resolution images and the corresponding low-resolution images formed using an estimate of the SR image can be used to iteratively refine the estimated SR image. The update equation is

$$g^{n+1}(p) = g^n(p) + \sum_m \left(y_k(m) - \tilde{y}_k(m) \right) \frac{(h_{mp}^{BP})^2}{c \sum_{m'} h_{m'p}^{BP}} \qquad (37)$$

where g^n is the estimate of the SR image at the nth iteration, y_k is the kth low-resolution image, $\tilde{y}_k$ is the kth low-resolution image as approximated from g^n, and h^{BP} is a back projection kernel. While

$\lambda = 0.001$ $\lambda = 0.005$ $\lambda = 0.01$

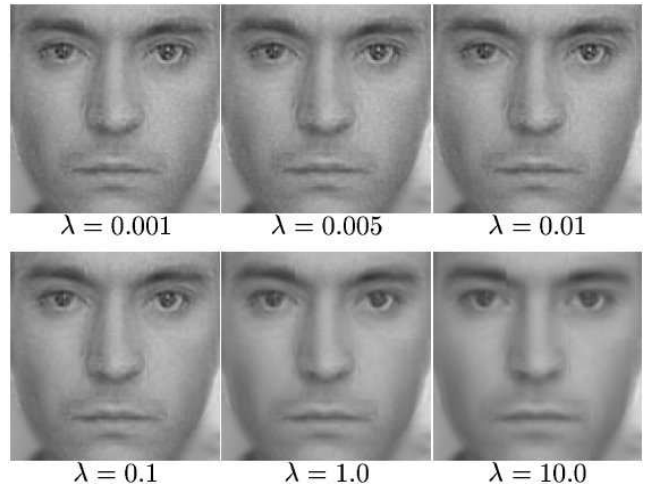$\lambda = 0.1$ $\lambda = 1.0$ $\lambda = 10.0$

Figure 8: Face space MAP reconstruction results with varying weight on the prior term. As λ increases the image moves closer to the average face. (Figure from [Capel and Zisserman 2001])

iterative back projection is relatively easy to understand, the method does not directly address the ill-conditioning of the problem and incorporation of *a priori* constraints is difficult.

5 Limits on Super-Resolution

Recently there has been growing interest in determining what the limits of super-resolution algorithms are. The most prominent work in this direction is that of Baker and Kanade[Baker and Kanade 2002]. Another analysis was later done by Lin and Shum[Lin and Shum 2004], following up on the results of Baker and Kanade. Both analyses address a somewhat simplified, but well justified, form of the super-resolution problem.

5.1 Baker and Kanade

Baker and Kanade derive three results that each show that the super-resolution problem becomes much more difficult as the magnification factor increases. The first is that for certain classes of point spread functions, the reconstruction constraints are not invertible and the null space of the linear system grows quadratically with the magnification term. Second, they show that for other more general classes of point spread function, the condition number grows at or greater than quadratic. The third result is that the solution space for potential reconstructions grows "extremely fast" with increases in magnification factor – meaning regularization terms will dominate the solution at some point.

The following assumptions are made throughout Baker and Kanade's analysis,

- The point spread function takes the form,

$$
\begin{aligned}
PSF_k(x) &= (w_k * a_k)(x) \text{ where} & (38)\\
a_k(x) &= \left\{ \begin{array}{ll} \frac{1}{S^2} & \text{if } |x| \le S/2 \text{ and } |y| \le S/2 \\ 0 & \text{otherwise} \end{array} \right\} & (39)
\end{aligned}
$$

with the width of the photoreceptor S the same for all input images.

- The optical blur portion of the PSF, $w_k(z)$, is a Dirac delta function, $\delta(z)$, implying no optical blurring (this is later relaxed).

- The registration function $r_k(z)$ corresponds to global translation. Thus, registration takes the form

$$r_k(z) = \frac{1}{L}z + c_k \qquad (40)$$

where L is the magnification factor and $c_k \in \mathbb{R}^2$ is a constant (although different for each image).

The performance of any resulting super-resolution algorithm based on these assumptions will depend on the number of input images K and the values of the local translations c_k. The analysis assumes the best possible scenario where there is an arbitrary number of input images and the registration is perfectly estimated. Thus, any derived upper bounds will be for best case scenarios.

5.1.1 Invertibility for Square Point Spread Functions

If we use a square point spread function $PSF_k(z) = a_k(z)$ and the registration is a translation, we can simplify the image formation equation (eqn 5) to,

$$y_k(m) = \sum_p W_k(m, p) \cdot g(p) \qquad (41)$$

$$W_k(m, p) = \frac{1}{L^2} \int_p a_k \left(\frac{1}{L}z + c_k - m \right) dz \qquad (42)$$

where the integration is over the pixel p. Using our earlier assumed form for a_k, we can further simplify $W_k(m, p)$ to

$$W_k(m, p) = \frac{1}{(L \cdot S)^2} \cdot A \qquad (43)$$

where A is the area of intersection between SR pixel p and the translated low-resolution pixel m (see figure 10. Without loss of generality, we can assume each pixel in the SR image occupies an area of 1 (which implies each LR image pixel occupies an area of $(LS)^2$). If LS is an integer larger than 1, then equation 41 is not invertible for any set of translations $\{c_k\}_{k=1}^K$ and the smallest dimension of the null-space is $(LS-1)^2$. If LS is not an integer, then some $\{c_k\}_{k=1}^K$ will enable inversion of equation 41.
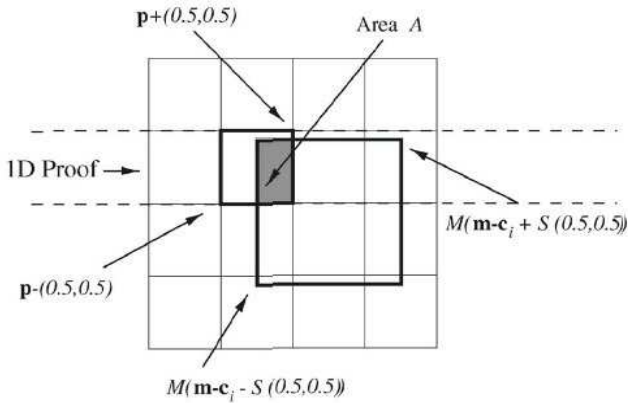
Figure 10: The area of intersection between SR image pixel p and translated low-resolution pixel m. (Figure from [Baker and Kanade 2002])

We omit the proof, but it is somewhat intuitive that if the size of each observed pixel is an integer multiple of the size of an SR image pixel then the resulting constraints will have less variety than if this were not so. Figure 11 compares magnifications of $LS = 1.5$ and $LS = 2.0$. With a magnification of 2.0 and no regularization, the problem is ill-conditioned as can be seen by the high-frequency artifacts in the reconstruction; whereas the magnification of 1.5 is invertible without regularization. Note the results agree with the theory. This result is important because it proves that there exist non-invertible scenarios regardless of the number of input images using fairly standard assumptions regarding the PSF and registration.

Figure 11: (a) Super-resolution with $LS = 2.0$, no regularization. (b) Super-resolution with $LS = 1.5$, no regularization. (c) Super-resolution with $LS = 2.0$, with regularization. (Figure from [Baker and Kanade 2002])

5.1.2 Conditioning for Arbitrary Point Spread Functions

The next result shown by Baker and Kanade is that the condition number for the linear system in equation 5 grows at least as fast as $(LS)^2$ for any optical blur w_k. This can be proved by employing the property of singular values of matrix A that for any vector x,

$$\sigma_1 \geq \frac{||Ax||_2}{||x||_2} \geq \sigma_n \qquad (44)$$

where σ_1 is the largest singular value, σ_n is the largest, and $|| \cdot ||_2$ is the L2 norm. Since $Cond(A) = \frac{\sigma_1}{\sigma_n} \geq \frac{||x||_2 ||Ay||_2}{||y||_2 ||Ax||_2}$ for any x and y, we can bound the condition number by finding appropriate x and y. If we choose x to be the SR image of all ones and y to be a checkerboard pattern and use $PSF_k = a_k$, it follows that the corresponding low-resolution images are $y_k(m) = 1$ and $|y_k(m)| \leq \frac{1}{(LS)^2}$ respectively, for all m. This implies a condition number greater than or equal to $(LS)^2$. For the extension to arbitrary PSFs the reader is directed to the original paper for the proof.

5.1.3 Volume of Solutions for Arbitrary Point Spread Functions

The third major result states that in the presence of quantization, the volume of the set of solutions grows asymptotically as $|LS|^{2n}$ where L is the magnification factor, S is the width of the photoreceptor for a single pixel, and n is the number of pixels in the SR image we are trying to recover. This is probably the most relevant result as it shows that the number of possible solutions grows very rapidly with the magnification factor. For standard super-resolution algorithms incorporating a smoothness prior, this means that at some point the number of potential smooth solutions grows so high that the algorithm no longer works well. Figure 12 shows super-resolution results for increasing magnification factor.

5.2 Lin and Shum

While Baker and Kanade show asymptotic trends in super-resolution algorithms based on the magnification factor and width of the photoreceptor for each pixel, they don't show limits on the magnification L, the number of input images needed to obtain acceptable results, or at what point additional images fail to provide new information. Lin and Shum[Lin and Shum 2004] extend the results of Baker and Kanade by analyzing these problems. Basing their analysis on perturbation theory of linear systems, they derive a practical limit of 1.6X magnification in the presence of registration and image noise and a theoretical limit of 5.7X for perfect registration. Further, they show a sharp cutoff in reconstruction performance once a certain set of low-resolution pixels have been captured and also show limits on the number of low-resolution images needed to form such a set.

6 Future Research Directions

While the super-resolution problem has been heavily studied, two recent trends have emerged that seem ripe for further advancement. The first trend is that of analyzing the limitations of super-resolution algorithms. The work of Baker and Kanade and that of Lin and Shum are fairly complete for the case of translated input images with standard reconstruction based algorithms[6]. However, it would be interesting to examine the effects of non-translational motion and further explore the effects of registration error and other forms of noise. Another thing that should be more carefully researched are the effects of quantization. Specifically, it would be good to have an analysis of how increased quantization granularity affects conditioning of the image formation equation.

The second trend in super-resolution is the use of learned priors to constrain the solution space. While much research has been focused in this direction recently, there is certainly still room for improvement. The primary issues here are finding good models for representing *a priori* information and finding efficient learning techniques for these models. The modeling and learning aspects go hand in hand because as model flexibility increases, the ability to effectively learn parameters of the model decreases.

In addition to these two trends, we have identified some new directions for future research. The first asks the question "Can we design camera sensors that are optimal for super-resolution algorithms?". Current CCDs act like box functions, spatially integrating over a uniform rectangular region for each pixel. It is likely that other types of sampling, such as integration over a Gaussian function, would yield better results[7].

Another direction is to combine super-resolution with dense stereo reconstruction. Dense stereo reconstruction is the process of estimating the depth of every point in the scene from two or more input images taken from different poses. Current techniques typically estimate depths for every pixel in one of the images, but it should be possible to obtain depths on a finer grid by combining super-resolution techniques with dense stereo.

A third potential problem is to combine high dynamic range imaging with super-resolution. High dynamic range imaging involves taking multiple images from the same viewpoint at different *exposures* and then combining the images to obtain an image of higher dynamic range[8]. By taking multiple images at different exposures and with slight translational offsets it might be possible to obtain a super-resolved high dynamic range image more efficiently than if the two stages are done independently.

7 Conclusions

We have provided a review of the current state of super-resolution research, covering the past, present, and future of the problem. A standard image formation model was first introduced, followed by summaries of the major classes of algorithms including frequency domain approaches, deterministic regularization, stochastic techniques, and projection onto convex sets. We then covered recent analyses on the limits of super-resolution algorithms and finally discussed potential future directions for research.

Acknowledgments

References

BAKER, S., AND KANADE, T. 2002. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell. 24*, 9, 1167–1183.

BORMAN, S., AND STEVENSON, R. 1998. Super-resolution from image sequences – A review. In *Proceedings of the 1998 Midwest Symposium on Circuits and Systems*, IEEE, Notre Dame, IN, USA, 374–378.

CAPEL, D. P., AND ZISSERMAN, A. 2001. Super-resolution from multiple views using learnt image models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 627–634.

CHEUNG, V., FREY, B. J., AND JOJIC, N. 2005. Video epitomes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

ELAD, M., AND FEUER, A. 1997. Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images. *IP 6*, 12 (December), 1646–1658.

ELAD, M., AND FEUER, A. 1999. Super-resolution reconstruction of image sequences. *IEEE Trans. Pattern Anal. Mach. Intell. 21*, 9, 817–834.

FREEMAN, W. T., PASZTOR, E. C., AND CARMICHAEL, O. T. 2000. Learning low-level vision. *Int. J. Comput. Vision 40*, 1, 25–47.

---

[6]Reconstruction based algorithms are those that use an image formation model to relate the high-resolution image to low-resolution inputs. Inversion of the formation model then is the crux of such algorithms.

[7]Obviously, the most ideal sampling function from a theoretical standpoint is an impulse function, but this is not feasible; also, the smaller the area of integration on the sensor, the higher the signal-to-noise ratio will be.

---

[8]Dynamic range refers to the range between the minimum and maximum light intensity captured by a camera. As light intensity decreases, there is a point at which a camera no longer senses any light; likewise there is a threshold at which a camera cannot detect additional incoming light.

FREEMAN, W. T., JONES, T. R., AND PASZTOR, E. C. 2002. Example-based super-resolution. *IEEE Comput. Graph. Appl. 22*, 2, 56–65.

HANSEN, P., AND O'LEARY, D. P. 1993. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput. 14*, 6 (November), 1487–1503.

HONG, M. C., KANG, M. G., AND KATSAGGELOS, A. 1997. An iterative weighted regularized algorithm for improving the resolution of video sequences. In *ICIP*, vol. 2, 474–477.

HONG, M. C., KANG, M. G., AND KATSAGGELOS, A. 1997. A regularized multi-channel restoration aproach for globally optimal high resolution video sequence. *SPIE VCIP 3024*, 1306–1317.

IRANI, M., AND PELEG, S. 1991. Improving resolution by image registration. *CVGIP: Graph. Models Image Process. 53*, 3, 231–239.

KANG, M. G. 1998. Generalized multichannel image deconvolution approach and its applications. *Opt. Eng. 37*, 11 (November), 2953–2964.

LIN, Z., AND SHUM, H. 2004. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *PAMI 26*, 1 (January), 83–97.

N. JOJIC, B. F., AND A.KANNAN. 2003. Epitomic analysis of appearance and shape. In *Proc. of International Conference on Computer Vision*.

PARK, S. C., PARK, M. K., AND KANG, M. G. 2003. Super-resolution image reconstruction : A technical overview. *IEEE Signal Processing Magazine 20*, 3 (May), 21–36.

SCHULTZ, R., AND STEVENSON, R. 1996. Extraction of high-resolution frames from video sequences. *IP 5*, 6 (June), 996–1011.

TSAI, R. Y., AND HUANG, T. S. 1984. Multi-frame image restoration and registration. In *in Advances in Computer Vision and Image Processing*, 317–339.

YOULA, D. C. 1978. Generalized image restoration by the method of alternating orthogonal projections. *IEEE Trans. Circuits Sys. CAS-25*, 694–702.
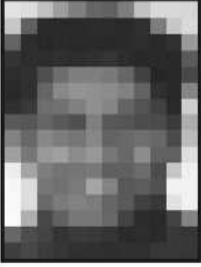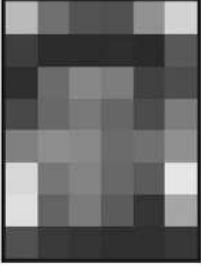
Input	48×64	24×32	12×16	6×8
Output	$\times 2$	$\times 4$	$\times 8$	$\times 16$
Reduction in RMS error vs. cubic B-spline	77% (9.2 vs. 11.9)	56% (12.4 vs. 22.2)	57% (19.5 vs. 33.9)	73 % (33.3 vs. 45.4)

Figure 9: "Hallucination" based reconstruction results using the method from Baker and Kanade. Only four low-resolution images were used for each of the reconstructions. Note that as the magnification increases, the algorithm begins placing patches somewhat randomly. (Figure from [Baker and Kanade 2002])
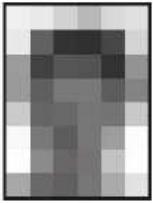
No. of Input Images	1 (Original)	4	16	64	256
Linear Magnification	$\times 1$	$\times 2$	$\times 4$	$\times 8$	$\times 16$

Figure 12: Results from a reconstruction-based super-resolution algorithm for increasing magnification factor. Notice that as the magnification increases, the smoothness constraint is no longer sufficient to recover much of the detail. This can be explained by the large increase in potential solutions as the magnification increases. (Figure from [Baker and Kanade 2002])