SUPER RESOLUTION IMAGING

SUPER RESOLUTION IMAGING
This is the
Subtitle

Edited by
DAVID REISMAN
University of Surrey

SECOND EDITOR
Second Affiliation

Kluwer Academic Publishers
Boston/Dordrecht/London

This book is dedicated to the memory of Adam Smith, who gave the study of economics its impetus.

Contributing Authors

Terrance E. Boult is the Weisman Chair profess and chairman of the newly formed Computer Science and Engineering Department at Lehigh University. From 1994 to 2001 he was with the Ele. Eng. and Comp. Science department at Lehigh. He served on the faculty of Computer Science at Columbia University faculty from 1986 until 1994. and he earned his Ph.D. in Computer Science from Columbia University in 1986,

Ross J. Micheals is a doctoral canidate at Lehigh University in the newly formed Computer Science and Engineering Department.

Ming Chao Chiang recieved his doctoral degree from the Computer Science department of Columbia University in 1998.

Chapter 1

SUPER-RESOLUTION VIA IMAGE WARPING

Theory, Implementation and Evalutation

Terrance E. Boult
[1]

Computer Science and Engineering Deparartment
Lehigh University, Bethlehem, PA 18015, USA
tboult@lehigh.edu

Ming-Chao Chiang

Ross J. Micheals
Computer Science and Engineering Deparartment
Lehigh University

Abstract

This chapter focuses on three issues: supporting image warping algorithms for super-resulution, examples of how image warping algorithms impact super-resolution image quality, and the development of quantitative techniques for super-resolution algorithm evaluation.

The warping approach proposed in this chapter is based on the integrating resampler (Chiang and Boult, 1996) which warps the image while both enforcing the underlying image reconstruction and satisfing the *imaging consistent constraint* (Boult and Wolberg, 1993). The imaging consistent constraint requires that the image reconstruction yields a function which, when convolved with the imaging system's point-spread function (PSF), is consistent with the input image. Many popular reconstruction techniques, including bilinear and natural cubic splines, do not satisfy the imaging consistent constraint. In this chapter, we review imaging consistent warping algorithms, how they form the core of the integrating resampler, and their implementation.

Although imaging consistent warping techniques can be used in other super-resolution implementations, such as those discussed in Chapter **??**, we present its use in a simpler direct approach: warping followed by a straightfoward fusion.

Examples are provided on grayscale images of simple patterns, text, and human faces. The use of priors in the fusion, such as those used in Chapter **??**, further enhance the results, but we analysize the simpler approach to isolate the impact of the warping algorithm.

The chapter then discusses the important problem of quantitative evaluation and presents a summary of two different quantitative experiments: using OCR and face recognition as metrics. These experiments clearly show the importance of high-quality reconstruction and warping to super-resolution. Perhaps more importantly, these experiments show that even when images are qualitatively similar, quantitative differences appear in machine processing. As the super-resolution field is pushed towards its boundardies, the ability to measure progress, even if it is small, becomes increasingly important.

Keywords: Super-Resolution, Imaging-Consistent Restoration/Reconstruction, Integrating Resampling, Integrating Resampler, Quantitative Evaluation, OCR, Bi-linear Resampling, Image Reconstruction, Image Restoration, Image Warping, Balanced Repeated Replicates, Replicate Statistics, Face Recognition.

1. BACKGROUND AND INTRODUCTION

The fundamental purpose of image warping is to allow the reshaping of image geometry for a variety of applications. Inherent in any super-resolution algorithm that uses multiple images is the alignment, or "matching," of data among the images—the computation of a mapping from each to a pixel in the super-resolution image. Except in specalized devices that intentially cause precise sub-pixel shifts,(**?**), alignment is almost always to a regular grid, and hence can be viewed as a general warp of the input image. General image warping, as is needed for super-resolution, requires the underlying image to be resampled at non-integer, and generally spatially-varying locations. Hence, super-resolution requires sub-pixel image reconstruction, but is not necessarily amenable to efficient image reconsturction via convolution. When the goal of warping is to produce output for human viewing, only moderately accurate image intensities are needed. In these cases, techniques using bilinear interpolation have been found sufficient. However, as a step for applications such as super-resolution, the precision of the warped intensity values is often important. As we shall show in this chapter, super-resolution based on bilinear image reconstruction may not be sufficient.

One of the first explicit uses of image warping for super-resolution was in (Peleg et al., 1987; Keren et al., 1988). Peleg and Keren estimated an initial guess of the high-resolution image, and simulated the imaging process via warping so that the difference between the observed and simulated low-resolution images was minimized. Irani and Peleg (Irani and Peleg, 1991; Irani and Peleg, 1993) used a back-projection method similar to that used in tomog-

raphy to minimize the same difference. Bascle et al. (Bascle et al., 1996) extended this back-projection method to include a simple motion blur model. We note, however, that all previous work has ignored the impact of image warping techniques.

Not all prior image-based work has used image warping — for example (Elad and Feuer, 1999), used constraints formulated from both the multiple input images and the super-resolution image. Algebraic approaches do have some significant advantages, e.g. (Elad and Feuer, 1999; Bak,); analysis of the underlying linear systems may constrain the blur kernel enough to permit the computation of new information. Also, algebraic approaches are more naturally extended to allow for Baysian estimation and the use of priors, e.g. (Schultz96, ; **?**; Bak,). Note, however, that algebraic constraints still require sub-pixel evaluation of the input for each pixel in the super-resolution image, which is tantamount to warping. One can view warping as a pre-processing that takes the spatial alignment and matching information and generates reconstructed images that would make solution of the algebraic equations more efficient. The lack of high-quality reconstruction for warping may be the unstated reason that algebraic techniques have not embraced warping.

This chapter is structured as follows. In Section 2, the image formation process and the relationships between restoration, reconstruction, and superresolutoin are briefly reviewed. The integrating resampler—an efficient method for warping using imaging-consistent reconstruction & restoration algorithms—is given in Section 3. In Section 4, we introduce the super-resolution algorithms considered in our analysis. Quantitative measurement of super-resolution imaging using three different applications is shown in Section 5.

2. IMAGE FORMATION, IMAGE RESTORATION AND SUPER-RESOLUTION

To address the problem of super-resolution, we need to first understand the process of image formation, reconstruction, and restoration. Although previous chapters provide most of the necessary background, to better describe our warping and super-resolution techniques, we briefly review the image formation process and sensor model as proposed in (Boult and Wolberg, 1993).

Generally, image formation can be described as a cascade of filtering operations. There is an overall blur applied at each pixel, $h(x, y)$, that can be decomposed as the sequence of operations as shown figure Fig. 1.1. Let $f(x, y)$ be the intensity distribution of a scene in front of a lens aperture. That distribution is acted upon by the blurring component of the lens, $h_1(x_y)$, yielding $f_1(x, y)$. The application of a geometric distortion function, $h_2(x, y)$, produces image $f_2(u, v)$. At this point, $f_2(u, v)$ strikes the image sensor where it undergoes ad-

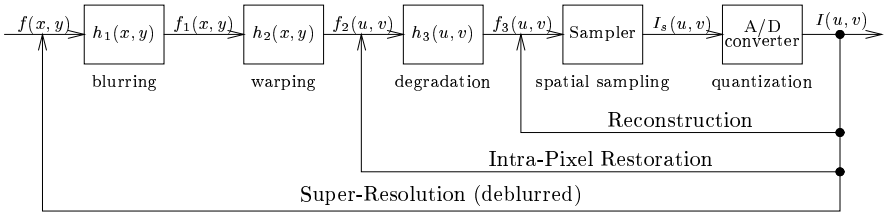


Figure 1.1. The image formation process and the relationship between restoration, reconstruction, and super-resolution.

ditional blurring by a point spread function, $h_3(u,v)$, which generates image $f_3(u,v)$. This blurring reflects the limitation of the sensor to accurately resolve each point without the influence of neighboring points. We choose to use a simple model wherein this blurring takes place within one pixel because for CCD and CID cameras since the physical boundaries between photosites generally allow only insignificant charge transfer between pixels. Image $f_3(u,v)$ undergoes spatial sampling as it hits the discrete CCD or CID photosites. The combination of convolution with the photosite blur h_3 and sampling is known as area sampling and reflects the finite size of a discrete photosite. If h_3 was assumed to be an impulse, then we have point sampling. While point sampling is often assumed for theoretical considerations, is not true in practice. In either case, intensities in the sampled image I_s are now defined only for integer values of u and v. The digital image $I(u,v)$ is obtained via an analog-to-digital converter that quantizes the samples of I_s. Note that parts of this decomposition are more conceptual than physical since, for instance, he geometric and bluring components occur simultaneously.

Reconstruction and restoration start with I (and models for one or more blur kernels $h_i(x,y)$), and seek to solve for one or more $f_j(x,y)$. Recovering an approximation of $f(x,y)$ is known as image *restoration* and is of considerable interest in image processing. The most common formulation of that problem, however, is actually recovering a descretized, rather than continuous, form of f. Recovering $f_2(x,y)$ might be called intra-pixel restoration, though it is is not commonly discussed in the literature.

Given this image formation model we might define super-resolution as the use of multiple images and/or prior model information to recover an approximation to $f(x,y)$ better than what would be obtained by image reconstruction followed by debluring using knowledge of $h_j(x,y)$. This definition includes approximating the image at a larger size with reasonable approximations for frequencies higher than those representable at the original size. While it may seem non-traditional it also includes improving the SNR while keeping the im-

age the size fixed. Given such an SNR improved image, one could simply perform a finer resampling and deblur to obtain a super-resolution with increased spatial resolution. Note that since debluring amplifies noise, the increased SNR can have a more significant result on the super-resolution image than might be initially expected. In practice, however, one would want to improve the SNR at the higher spatial resolution to reduce the impact of reconstruction artificats when increasing the resolution.

Because of the multiple, and different degradations in this imaging model, we will define two different types of super-resolution that will be considered in this chapter. Recovering a discrete approximation, with resolution higher than I_s, to f_1 is called (plain) super-resolution and appoximation to f is called super-resolution with deblurring. Note that super-resolution with deblurring requires knowledge of the primary bluring kernel — a reasonable assumption for simple lens blur but tenuous for atomospheric or depth of field effects. Because super-resolution increases the signal-to-noise ratio in the approximation to f_1, it significantly ameliorates the ill-conditioned nature of deblurring.

3. IMAGING-CONSISTENCY AND THE INTEGRATING RESAMPLER

Image reconstruction plays a key role in all super-resolution algorithms. Given the finite set of samples, there is an uncountably infinitely number of functions satisfy that data, and hence, image interpolation involves adding regularization constraints to allow a unique function to be defined given the image data. Often there is a need to balance computational complexity against the sophisticated nature of the assumptions and constraints. The many constraints been developed in the design of image reconstruction filters have been extensively discussed: in books (Andrews and Hunt, 1977; Pratt, 1978; Gonzalez and Wintz, 1987; Pavlidis, 1982; Wolberg, 1990; Pratt, 1990), articles (Simon, 1975; Andrews and Patterson, 1977; Hou and Andrews, 1987; Park and Schowengerdt, 1982; Reichenbach and Park, 1989; Jain, 1989; Oakley and Cunningham, 1990), and comparison papers (Parker and and D.E. Troxel, 1983; Mitchell and Netravali, 1988; Maeland, 1988). Many of these constraints are related to how well the underlying filter approximates the ideal sinc filter. Even the "ideal" sinc interpolation is based on the assumption that the image is an infinite signal sampled at or above its Nyquist rate. While it is true that optics limit image bandwidth, it need not result in images that are Nyquist sampled. If the underlying function f was Nyquist sampled, then except for noise removal, there is no need for super-resolution.

In (Boult and Wolberg, 1993), a new constraint was added to the mix of potential assumptions for image reconstruction: requiring the algorithm to be *image-consistent*. An algorithm is called imaging-consistent if it is the ex-

act solution for some input function, which, according to the imaging model, would have generated the measured input. This constraint is particularly important for super-resolution because it means each resampling would, when subjected to the imaging model, actually be consistent with the measured image.

For image reconstruction, we can achieve a image-consistent reconstruction by first restoring the image to yield an approximation to f_2, then performing an additional blur by the pixel's PSF. Although restoration is ill-posed, blurring produces an image reconstruction that is totally consistent with the input data, regardless of the resampling rate. The use of image restoration technique permits the work presented in this chapter to achieve image reconstruction in a fundamentally different way than traditional approaches. Our approach is in the spirit of the work of (Huck et al., 1991), where it is argued that sampling and image formation should be considered together. Imaging-consistent algorithms directly combine knowledge of image formation and sampling into the reconstruction & restoration process. The way that knowledge is used, however, is quite different from (Huck et al., 1991).

Imaging-consistent algorithms follow quite naturally from a general approach to algorithm development known as information-based complexity (IBC) (see (Traub et al., 1988)). From IBC, it can be shown that the imaging-consistent algorithms enjoy very good error properties for many definitions of error. In particular, imaging-consistent algorithms have, within the presribed space of functions, an error at most twice that of any algorithm for *any* error measure defined as a weighted norm on the space of solutions (e.g., L^2, or even a weighted least-squares measure). Note that most image-quality measures yielding a scalar are error measures of this type — e.g., the QSF measure of (Drago and Granger, 1985; Granger, 1974), QSF extensions that include contrast effects, any weighted integral of the modulation transfer function (MTF), and the measure of (Park and Schowengerdt, 1982) when weighted and integrated over frequency v. For the algorithms discussed here we presume the space of functions are continuous and piecewise analytic with a bounded first derivative in each piece. More discussion of these error properties, and alternative spaces of functions, can be found in (Chiang, 1998).

Of course, an algorithm that performed a full restoration followed by blurring could be computationally expensive. Fortunately, with some effort, the imaging consistency constraint can be applied in a functional way, and encorporated into a very efficient algoirthm. In this chapter, only an overview is provided. For more details, and the derivation of four other imaging consistent algoirthms, see (Boult and Wolberg, 1993; Chiang, 1998). One-dimensional image models are presented herein, but since higher dimensions may be treated separably, the process is easily extended.

The simplest imaging-consistent method to consider is based on a piecewise quadratic model for the image. If we assume each photosite PSF (h_3) is a Rect filter (1.0 inside the pixel, zero otherwise), an imaging consistent algorithm is easy to derive. To ensure that the function is continuous and local, we define the value of the reconstruction at the pixel boundaries k_i and k_{i+1} to be equal to E_i and E_{i+1}. Any method of approximation could be used to compute E_i, though our examples will only include cubic convolution or linear interpolation. See (Chiang, 1998, Section 2.3) for a more details derivation and more examples.

Given the values E_i at the pixel edges, an imaging consistent constraint is that the integral across the pixel must equal V_i. This results in exactly three constraints:

$$g_i(-1/2) = E_i; \quad g_i(1/2) = E_{i+1}; \quad \int_{-1/2}^{1/2} g_i(x)\,dx = V_i. \tag{1}$$

From Eq. (1), one can derive the following quadratic polynomial

$$g_i(x) \;=\; 3(E_i + E_{i+1} - 2V_i)x^2 - (E_i - E_{i+1})x - (E_i + E_{i+1} - 6V_i)/4, \tag{2}$$

where $-1/2 \le x \le 1/2$. Using cubic convolution with parameter a to derive E_i and E_{i+1} yields

$$\begin{aligned} E_i &= \frac{1}{8}(aV_{i-2} + (4-a)V_{i-1} + (4-a)V_i + aV_{i+1}), \\ E_{i+1} &= \frac{1}{8}(aV_{i-1} + (4-a)V_i + (4-a)V_{i+1} + aV_{i+2}). \end{aligned} \tag{3}$$

So that the cubic convolution kernel resembles the sinc funtion, the parameter a is generally in the range [-3, 0], with the values -0.5, -0.75, and -1.0 having special significance (see (Simon, 1975; Keys, 1981; Park and Schowengerdt, 1983)). Note that with $a = 0$, we have

$$E_i = (V_{i-1} + V_i)/2 \quad \text{and} \quad E_{i+1} = (V_i + V_{i+1})/2. \tag{4}$$

In other words, for $a = 0$, cubic convolution interpolation of the edge values (i.e. midpoints between pixels) is equal to the value given by bilinear interpolation.

When applied over the entire image, Eq. (2) yields f_2, an intra-pixel restoration. If an imaging-consistent reconstruction is desired, it may be obtained from the intra-pixel restoration via covolution with the pixel PSF. Assuming a Rect PSF for the pixel, one can integrate Eq. (2) to derive a functional form for reconstruction. The result is the per-pixel cubic polynomial

$$G_i(x) \;=\; \int_{x-1/2}^{1/2} g_i(z)\,dz + \int_{-1/2}^{x-1/2} g_{i+1}(z)\,dz \tag{5}$$

$$
\begin{aligned}
= \quad & (E_{i+2} - E_i - 2(V_{i+1} - V_i))x^3 \\
& + (2E_i - E_{i+1} - E_{i+2} + 3(V_{i+1} - V_i))x^2 + (E_{i+1} - E_i)x + V_i,
\end{aligned}
$$

where $0 \leq x \leq 1$ spans from the center of one input pixel to the next.

It is interesting to note that if $a = 0$ as in Eq. (4) (i.e. linear interpolation) is used to determine E_i, the resulting imaging-consistent reconstruction (Eq. (5)) is tantamount to cubic convolution with the "optimal" value of $a = -0.5$ — proof can be found in (Chiang, 1998, Section 2.4). No other value of a yields a reconstruction that satisfies the imaging-consistent constraint with a simple PSF. That is, if we use cubic convolution with $a \neq 0$ to estimate E_i, the resulting imaging consistent polynomial is not is equivalent to any cubic convolution. We have found using cubic convolution with $a = -0.5$ to estimate E_i is one of the best imaging consistent algorithms and is the value used for most of the examples in this chapter.

This section presented a model that is globally continuous and analytic except on the pixel boundaries, which results in a per pixel model which is quatratic after restorition (cubic after reconstruction). In (Boult and Wolberg, 1993; Chiang, 1998), we also present/analysize alternatives that are globally differential or smoother, and also models that have multiple polynomials per pixel.

3.1. IMAGING CONSISTENT WARPING: THE INTEGRATING RESAMPLER

To define an imaging-consistent warping, we generalize the idea of the imaging-consistent reconstruction/restoration. Whereas imaging-consistent reconstruction assumes that the degradation models are identical for both input and output; imaging-consistent warping allows both the input and output to have their own degradation model, and the degradation model to vary its size for each output pixel.

The imaging-consistent algorithms described above and in (Boult and Wolberg, 1993) are linear filters. We designed them for use in what we call the *integrating resampling* approach. For the super-resolution results described herein, we consider only the integrating resampler assuming a Rect PSF filter as described in (Chiang and Boult, 1996), which we refer to as QRW.

As described before, our model of image formation requires the image to be spatially sampled with a finite area sampler. This is tantamount to a weighted integral being computed on the input function. Because we have a functional form for the restoration, we can simply integrate this function with the PSF for the output area sampler. In this section, although we assume that the output sampler has a Rect PSF, it should be noted that there is no limitation on other potential degradation models. However, Rect is used not only to simplify the

algorithms, but because it is a good model for super-resolution where each photosite is represented with a pixel.

When resampling the image and warping its geometry, this new approach allows for efficient pre-filtering and post-filtering. Additionally, because a functional form of the input has already determined, no spatially-varying filtering is needed, unlike a case using a direct inverse mapping.

Computing the exact value of the imaging-consistent warped value (the integrated restored function weighted by the PSF) can representedq in functional form if the mapping function has a functional inverse and the PSF is simple. In general, however, super-resolution algorithms may have complex maps requiring numerical integration, since such maps cannot be represented in closed form. To reduce the computational complexity, we propose a scheme where for within each input pixel, we use a linear approximation to the spatial warp, but use the full non-linear warp to determine the location of pixel boundaries. This integrating resampler, first used in (Boult and Wolberg, 1992) and formally described in (Chiang and Boult, 1996), also handles antialiasing of partial pixels in a straightforward manner.

Assume n input pixels are being mapped into k output pixels according to the mapping function $m(t)$. Let m_i be the mapped location of pixel i, for i=0, . . . , n. Compute δ_j, j=0, . . . , k, as the linear approximation to the location of $m^{-1}(j)$, as shown in Fig. 1.2. To avoid fold-over problems, we assume that the mapping function is strictly increasing. For an approach to modeling fold-over, see (Wolberg and Boult, 1989).

```
for (i = j = 0; j ≤ k; j++) {
  while (i < n − 1 && m_{i+1} < j) i++;
  δ_j = i + (j − m_i)/(m_{i+1} − m_i); }
```

Figure 1.2. Linear approximation to the location of $m^{-1}(j)$.

For efficient computation of the integral as well as the ability to perform proper antialiasing, the integrating resampler—given as pseudo code in Fig. 1.3 — "runs" along the input and output determining in which image the next pixel boundary will be crossed. To do this, there are two variables: *inseg* $\in [0, 1]$, which represents the fraction of the current input pixel left to be consumed, and *outseg*, which specifies the amount of input pixels required to fill the current output pixel. In the integrating resampler, the function $R(t; g)$ is obtained from the definite integral of an imaging-consistent restoration $g(x)$ as

$$R(t; g) = \int_{-0.5}^{t} \text{PSF}(x)\ g(x)\ dx \tag{6}$$

which, natrually, changes according to pixel. An example showing image values V_i, edge values E_i, imaging consistent intra-pixel restoration g_i, and imaging consistent reconstuction R is presented in table 1.1. Note this is *not* the

i	V_i	E_i	g_i	R
0	0	0	0	0
1	0	0	0	0
2	0	0	$-47.8x^2 - 15.9x + 3.9$	$-47.8t^3 - 15.9t^2 + 3.9t$
3	0	-15.94	$334.6x^2 - 143.4x - 27.8$	$334.6t^3 - 143.4t^2 - 27.8t - 91.6$
4	255	127.50	$-334.6x^2 + 143.4x + 282.8$	$334.6t^3 - 143.4t^2 - 27.8t - 219.1$
5	255	270.94	$47.8x^2 - 15.9x + 251.0$	$47.8t^3 - 15.9t^2 + 251.0t - 135.4$
6	255	255.0	255	255t -127.5
7	255	255.0	255	255t -127.5

Table 1.1 . A simply step edge, and the resulting values and polynomials that serve as input to the integrating resampler. In computing the edge values, we presume pixel replication outside the image.

same cubic as Eq. (5) — an integral over a full pixel size by our prevous definitions, implies a combination of two different quadratics. The table shows values only within individual pixels.

Assuming proper update to the algorithm's state, whenever *inseg* $<$ *outseg*, we know that the input pixel will finish first, so it may be can consumed. If, on the other hand, it happens that *inseg* $\geq$ *outseg*, the output pixel will finish first, so an output is produced. Thus, in each iteration of the loop we either consume one input pixel or produce one output pixel. Therefore, the algorithm requires at most $k + n$ iterations.

The underlying idea of this integrating resampler can be found in the work of Fant (Fant, 1986) who proposed an efficient bilinear warping algorithm. With some effort, one can see that by setting $R(t) = v_i t - 0.5$, the integrating resampler implements a bilinear warp.

In summary, the contribution discussed in this section is twofold:

1 the generalization of Fant's orginal algorithm into the integrating resampler which supports the use advanced imaging-consistent reconstruction algorithms, and

2 the provision for a modeling real lens effects by using a real warps that affect the image radiance. In Fant's original work, the goal was to warp images for graphic effects, and hence to affect geometry without disturbing the intensities. To do this, the algorithm maintaines knowledge of the input size and normalizes the integral to account for this size, giving a normalized intensity. Thus, if a constant image was stretched to twice its normal width, it would change shape but retain the same intensities. If a lens was placed into an imaging system so as to double the width of the image on the sensor plane, then the value measured would be halved. The algorithm is flexible enough to support both "graphics" and "lens" modeling. If the super-resolution is over images that vary because of,

Pad the input; compute k_l, k_r, and i_l, the indices to the leftmost and rightmost output pixels and the index to the leftmost input pixel that contributes to the output; and compute the linear approximation to the location of $\delta_j = m^{-1}(j)$, for $j = k_l, \ldots, k_r + 1$.

```
normalizingfactor = δ_{k_l+1} − δ_{k_l};          //  set up for normalization
δ_{k_l} = MAX(δ_{k_l}, 0);                        //  ensure that δ_{k_l} is nonnegative
inseg = 1.0 − FRACTION(δ_{k_l});                  //  fraction of input pixel left to be consumed
outseg = δ_{k_l+1} − δ_{k_l};                     //  #input pixels mapped onto one output pixel
acc = 0.0;                                        //  reset accumulator for next output pixel
for (j = 0; j < δ_{k_l}; j++) out[j++] = 0;       //  zero out the garbage at left end
for (i = i_l, j = k_l; j <= k_r; ) {              //  while there is output to produce
  Use the current pixel (in[i]) and neighbors to
  update R(), the integral of the restoration g().
  leftpos = 1.0 − inseg;                          //  get left endpoint for integration
  if (inseg < outseg) {                           //  if we will consume input pixel first
    acc += R(1) − R(leftpos);                     //  add integral to end of output pixel
    i++;                                          //  index into next input pixel
    if (i == n) {                                 //  check end condition
      if (normalize) acc /= normalizingfactor;    //  normalize the output, if appropriate
      out[j] = acc;                               //  init output
      break;                                      //  exit from the loop
    }
    outseg −= inseg;                              //  inseg portion has been filled
    inseg = 1.0;                                  //  new input pixel will be available
  }
  else {                                          //  Else we will produce output pixel first
    acc += R(leftpos + outseg) − R(leftpos);      //  add integral to end of output pixel
    if (normalize) acc /= normalizingfactor;      //  normalize the output, if appropriate
    out[j] = acc;                                 //  init output
    j++;                                          //  index into next output pixel
    acc = 0.0;                                    //  reset accumulator for next output pixel
    inseg −= outseg;                              //  outseg portion of input has been used
    outseg = δ_{j+1} − δ_j;                       //  new output size
    normalizingfactor = outseg;                   //  need for normalization
  }
}
for (j = k_r + 1; j < k; j++) out[j++] = 0;       //  zero out the garbage at right end
```

Figure 1.3. The integrating resampler assuming a output model of an Rect PSF filter. See text for discussion.

say, atmospheric variations, or if we are correcting for lens distortions, an unnormalized warp should used.

These contributions are at the foundations of our fusion of image-consistent warping and super-resolution.

4. WARPING-BASED SUPER-RESOLUTION

We turn now to the use of warping for super-resolution. As described in earlier chapters, super-resolution refers to the process of constructing high-resolution images from low-resolution image sequences. Given the image sequence X_L, our warping-based super-resolution algorithm is formulated, as follows:

Define Reference and Mapping Choose one of the images, say X_p, as the reference image, and compute the motion field between all the images and the reference image.

Warp Scale up the reference image using QRW, and then use QRW to warp all the images to the reference image based on the motion field and scale computed in the previous step.

Fusion Obtain a super-resolution image by fusing all the images together.

Deblur (If desired) Deblur the resulting super-resolution image using $h_1(x, y)$

This method presumes that lens blur, h_1, is approximately the same for all images. Otherwise, the debluring step must be performed before the fusion stage. However, because of the noise amplification caused by debluring in each image, debluring before fusion is not as effective and should be used only when necessary.

We presume that a dense motion-field is computed between each image in the sequence — a straightfoward calculation for motion that locally approximates a rigid transform. The motion field computations for the examples presented in this chapter are based on a sum-of-square difference matching algorithm, with, 11x11 and 7x7 template windows for the first and second experiments respectively. In each case, the matching is a dense disparity surface. Sub-pixel estimates are obtained by fitting a quadratic to each point at which the match is unique and its two neighbors. When off-the-shelf lenses and cameras are used, pre-warping can be used to remove the distortions. In (Chiang and Boult, 1996), we showed that the pre-warping with integrating resampler can improve the match quality given significant lens distortion.

In the face-based experiments, we did not have access to a face image database with a large number of views of the same subject. (We used the FERET database; more on this later). Therefore, we used a set of synthetic

downsamplings from with a single image to generate a sequence of perspective warps. Since the experiment called for warping so many images we directly used the matching information defined by the synthetic mappings. The mappings for each face were randomly generated, but the same set were used for both the bilinear and the QRW super-resolution warpings.

The fusion step is not the focus of this chapter, nor of our past work. We have tried several different approaches to fuse the images together, including simply averaging or a median filter. Our experiments show that the median filter is better, though often not much better than the averaging filter. Median filtering is used for the OCR experiments and simple averages for the others. More advanced techniques using priors, e.g. (Schultz96, ; **?**; Bak,), could probably produce better results but would still be expected to benefit from the increased quality in fusion input. Again, the experiments in this chapter sought to isolate the effects of warping.

The test data shown in this section was taken using two different Sony cameras, models XC-77 and XC-999, captured by a Datacube MV200 System. Fig. 1.4 show our an experimental result with Fig. 1.5 showing the same results except that all the images are normalized so that the dynamic ranges are identical. All the resulting super-resolution images are 256x256, and were scaled-up by a factor of approximately four (4). We note that the previous works (Gross, 1986; Peleg et al., 1987; Keren et al., 1988; Irani and Peleg, 1991; Irani and Peleg, 1993; Bascle et al., 1996) report results only scaling by a factor of two (2).

Fig. 1.4 shows the super-resolution results of our first example. Fig. 1.4a shows an input image blown up by a factor of 4 using pixel replication so that the value of each pixel can easily be seen. Fig. 1.4b shows super-resolution by our implementation of the back-projection method described in (Irani and Peleg, 1991) (not the orginal authors, see (Chiang, 1998; **?**) for details); Fig. 1.4c shows super-resolution using bilinear resampling followed by deblurring; Fig. 1.4d, super-resolution using QRW followed by deblurring. Also, for the purpose of comparison, we are assume that Figs. 1.4c and 1.4d have undergone the same degradation before sampling. Fig. 1.4e and f shows the super-resolution results without deblurring.

Fig. 1.5 shows the results after all the images are normalized so that the dynamic ranges are identical, as follows:

$$I_n = \frac{255}{\max_u - \min_u}(I_u - \min_u)$$

where I_n and I_u are, respectively, the normalized image and the image to be normalized, $\max_u$ and $\min_u$ are, respectively, the minimum and maximum intensity values of the image to be normalized.

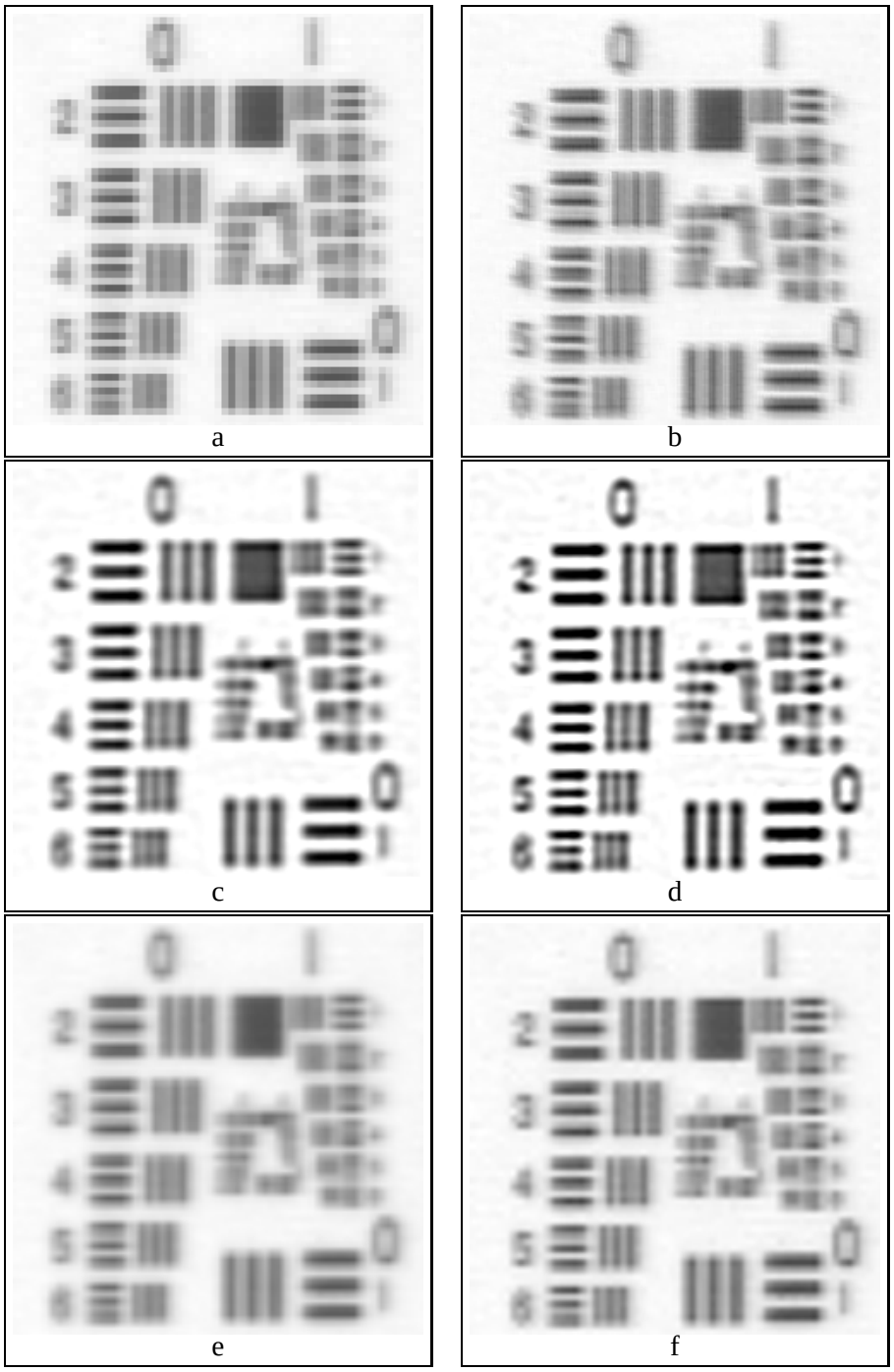


Figure 1.4. Final results from an 8 image sequence (64x64) taken by XC-77. (a) An orginal image blown up by a factor of 4 using pixel replication; (b) super-resolution by back-projection using bilinear resampling to simulate the image formation process and (e) as the initial guess; (c) super-resolution using bilinear warping followed by deblurring; (d) super-resolution using QRW followed by deblurring. Image (e) shows (c) (bilinear warping) without deblurring and (f) shows (d) (QRW) without deblurring.

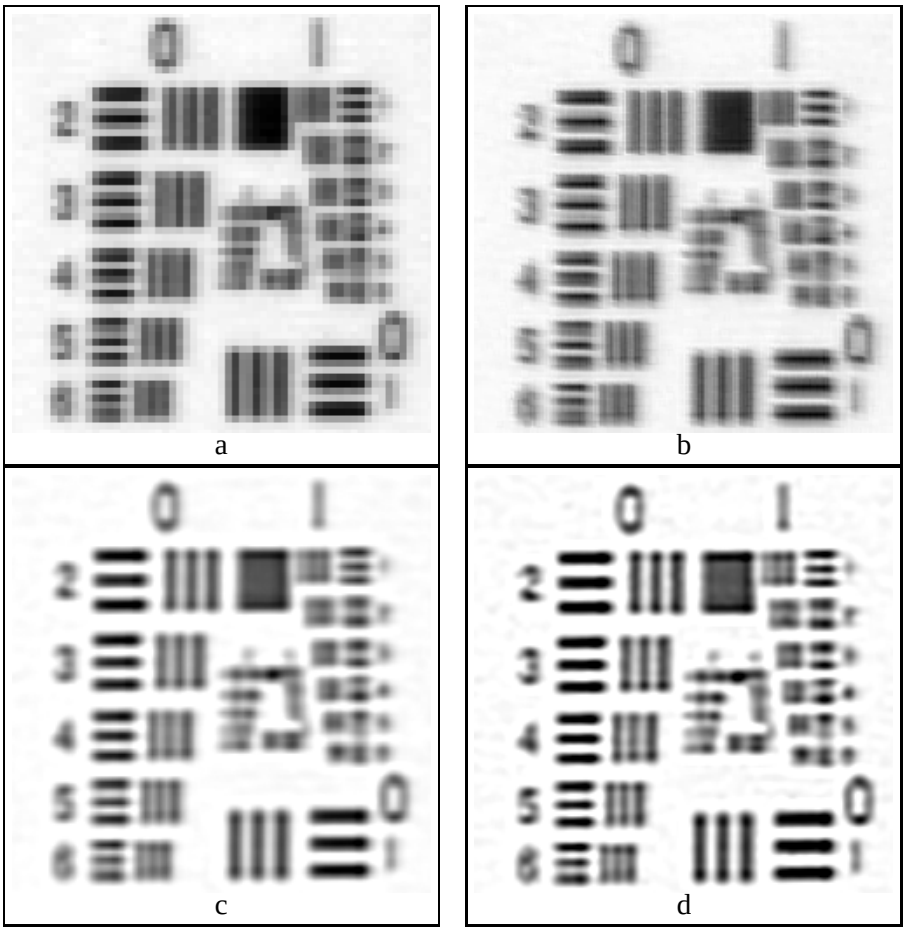


Figure 1.5. Results of Fig. 1.4 after being normalized to have the same dynamic range. (a) An orginal image blown up by a factor of 4 using pixel replication; (b) super-resolution by back-projection (c) super-resolution using bilinear warping followed by deblurring; (d) super-resolution using QRW followed by deblurring.

Fig. 1.6 shows an example captured with a Sony XC999, which is a one-chip color camera. Note the target is similar to (yet different) from that in the first example. Fig. 1.6a shows one of the original images blown up by a factor of 4; it can be easily seen that inter-frame motion is involved in this case. Fig. 1.6b shows super-resolution using QRW followed by deblurring. Obviously, our super-resolution method removes most of the interframe motion and significantly improves the sharpnes of the image.

We implemented the back-projection method proposed in (Irani and Peleg, 1991) and found it somewhat difficult to work with since it is sensitive to the choice of its parameters called normalizing factors. For the comparisons,

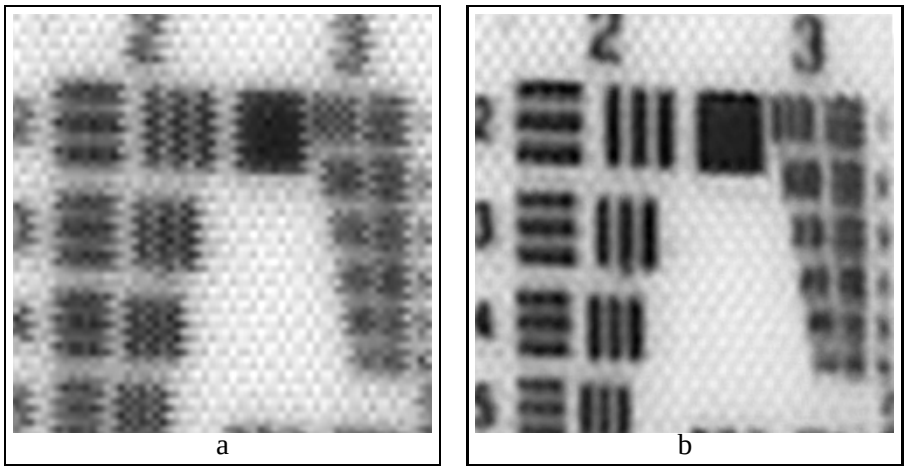


Figure 1.6. Super-resolution results from a very noisy image sequence of 32 images (64x64) taken by XC-999. (a) one of the original images blown up by a factor of 4; (b) super-resolution with QRW followed by deblurring.

we tried many normalizing factors and chose one that resulted in the back-projected images (Fig. 1.4d) with minmal sum-of-square difference (SSD) between the observed and simulated images. It is worth pointing out that in this particular case, SSD is not necessarily a good error measure because it is not robust. Furthermore, the same normalizing factor does not always give the best result in terms of the error measure when different resampling algorithms are used or when the input set is changed.

Results from our experiments show that the direct method we propose herein is not only computationally cheaper, but it also gives results comparable to or better than those using back-projection. Moreover, it is easily seen from Fig. 1.4 that the integrating resampler outperforms traditional bilinear resampling. Not surprisingly, our experiments show that most of the additional information carried by each image is concentrated on the high frequency part of the image. This observation also explains why the integrating resampler outperforms bilinear resampling. As was shown in (Boult and Wolberg, 1993), when viewed as a reconstruction filter, bilinear causes more blurring than the imaging-consistent reconstruction of equation 5.

Table 1.2 gives the running times of our first example, as measured in 1996, using a Sun 143MHz Ultra SPARC running Solaris 2.5 and a 120MHz Pentium running Linux 2.0.12. Note that for both the maintenance of precision and ease of implementation, all operations were performed in double-precision floating point, and neither algorithm was explicitly optimized. Also, note that the motion field computation, required by both methods, is included in the timings. This required the warping and fusion of 8 images with the final size 256×256.

Time in seconds	QRW		Back-Projection	
	SPARC	Pentium	SPARC	Pentium
Warping	1.94	3.52	NA	NA
Fusion	0.04	0.25	NA	NA
Deblurring	0.81	1.31	NA	NA
Total	2.79	5.08	12.33	24.80

Table 1.2 . Running times for our first examples (8 images) assuming the same degradation model. See text for details.

As shown in Table 1.2, for this example, our method is more than four times faster than our implementation of Irani's back-projection method. In general, Irani's back-projection method takes an amount of time roughly proportional to both the number of iterations and the degradation model. Our experiments show that although each iteration of Irani's back-projection method takes approximately 65% of the running time of our method, the algorithm performs a minimum of two iterations. Thus, even in its best case, Irani's back-projection method is about 30% slower than our method. Our experiments also suggest that more than three iterations are often required to minimize the sum-of-square difference — implicating that the direct warping approach is often more than 100% to 200% faster than Irani's back-projection method.

5. QUANTITATIVE EVALUATION

In this section, we turn our discussion to the quantitative measurement of super-resolution. Historically, as new super-resolution techniques have been developed, the standard practice has been to present a few examples of the technique to allow the reader to reach their own qualitative conclusions. It is difficult to justify comparisons between super-resolution algorithms that make different fundamental assumptions — for instance regular sub-pixel shifts vs. aperture changes vs. object motion. However, in order to make progress, we need to be able to quantatively measure the improvements of algorithms.

Some seemingly natural measures would be a blind measure of image quality — (**?**) some measure of difference between a high-resolution image and the recovered super-resolution image or some type of spectral analysis seeing how well high-requencies are recovered. Blind image-quality metrics are, however, fraught with problems as an overall measure of super-resolution algorithms because they are inherently task independent and disregard the underlying signal. Image differences from ground truth have been used in many areas as a measure of quality for comparison, but it remains difficult to decide how the differences should be weighted. In spectral analysis, we can look at how well the resulting super-resolution spectum compares with the orginal. In

section 5.1 we briefly review the quantitative and qualitative spectral analysis from (Chiang, 1998, Chapter 4).

While we have explored image differencences and spectral measures, it is difficult to reduce them to a simple quantitative measure to allow comparison of two algorithms. The major difficulty with either difference or spectral analysis is how to combine the different spatial or spectral differences to a comparison metric. Simple approaches such as RMS of the difference is not very meaningful, just as RMS error is not a very good measure of image quality for compression. Also note that for super-resolution magnifications of more than double, the original images contain frequencies so far above the Nyquist rate of the small images that the reconstruction techniques have no hope of recovering them. The intensity variations measured in these regions are a mixture of blurring and unmodeled aliasing. Neither super-resolution nor any resampling/reconstruction algorithm can recover all the lost information. While we are aware of these differences (these existence cannot be avoided), their significance is unknown. Rather than attempting to define what is important in some generic image sense, we believe that task oriented measures are more appropriate.

We present two realistic problems for which super-resolution is a natural augmenting tool. Since the metrics are computed using commercial products, their implementation and ease of which the evaluation may be reproduced are straightforward. The first problem, optical character recognition, or OCR, is considered in section 5.2. The second, face-based human identification, is presented in section 6. In both domains, image detail is critical to the systems' final performance.

In our OCR experiment, input is obtained from a hand-held camera, as opposed to the more traditionally used flat-bed scanner. Limited by NTSC resolution, we will show not only that super-resolution can significantly increase the recognition rate, but also the importance of warp quality. The experiment, described in section 5.2 and (**?**), is significant, however, has drawbacks. The quantitative analysis used only a small number of samples. In addition, binary nature of the input may allow over-enhancements to cause increased recognition rates. Because of these limitations, we sought an additional domain.

Another approach, based on appearance matching and pose estimation, can be found in (Chiang, 1998, Chapter 7). That analysis used grayscale images captured at two different resolutions and compared the results of running SLAM (Nene et al., 1994), an appearance-based recognition & pose algorithm, over various super-resolution images. The results were, in general, consistent with the OCR problem, except that there were instances where blurring the image actually increased accuracy of the recognition & pose estimation. In these unexpected cases, super-resolution did not help. Again, the sample size was small, and the targets with significant numbers of high-contrast edges may

have dominated the results. Finally, in retrospect, the problem of pose computation from low resolution images was slightly artificial, and hence is not presented here.

In the case of face recognition we are addressing a very real problem — the recognition or verification of a human's identity from low-resolution facial images. In the human identification problem, it is common for a wide-field of view camera to be used and for subjects to be at varying distances. Increasing the working range of existing systems is an ongoing research topic, and super-resolution is one potential way of achieving this. We synthetically "project" the images to produce the low resolution data. For true super-resolution, we would need multiple views of hundreds of heads and a robust facial matching techniques (since the head may rotate between frames). A experiment using multiple camera resolutions and facial matching is being planned. In this experiment, there is a large enough data space to infer confidence in the results. This allows us to quantitatively compare super-resolution using bilinear resampling with super-resolution using QRW and have statistical confidence in the hyphothesis that improved image warping (QWR) may improve super-resolution results.

5.1. SPECTRAL ANALYSIS OF SUPER-RESOLUTION FROM SYNTHETICALLY DOWN-SAMPLED IMAGES

For a spectral comparison, a sequence of five synthetically down-sampled images were used. The original high-resolution image provides the necessary ground truth for comparing the super-resolution results.

Fig. 1.7 shows the experimental results. The test images are generated from the high-resolution image shown in Fig. 1.7 by translation followed by down sampling. Fig. 1.7a shows the original high-resolution image; Fig. 1.7b the scale-up of the down-sampled version of Fig. 1.7a by a factor of 4 using bilinear resampling (no super-resolution); Fig. 1.7c the super-resolution result from the synthetically down-sampled image sequence using bilinear resampling with deblurring; Fig. 1.7d the super-resolution result from the synthetically down-sampled image sequence using QRW with deblurring.

Table 1.3 shows the powers of the Fourier transform of the images shown in Fig. 1.7. For summary analysis, the images are divided into regions based on their relation to the sampling rate of the low-resolution (64x64) image. The regions are shown graphically in Table 1.3a, with the power within them shown in Table 1.3b. The column marked P_3 shows the power of the whole region (i.e., the whole spectrum). The columns marked P_2, P_1, and P_0 show, respectively, the power of the central 192x192 region, the power of the central

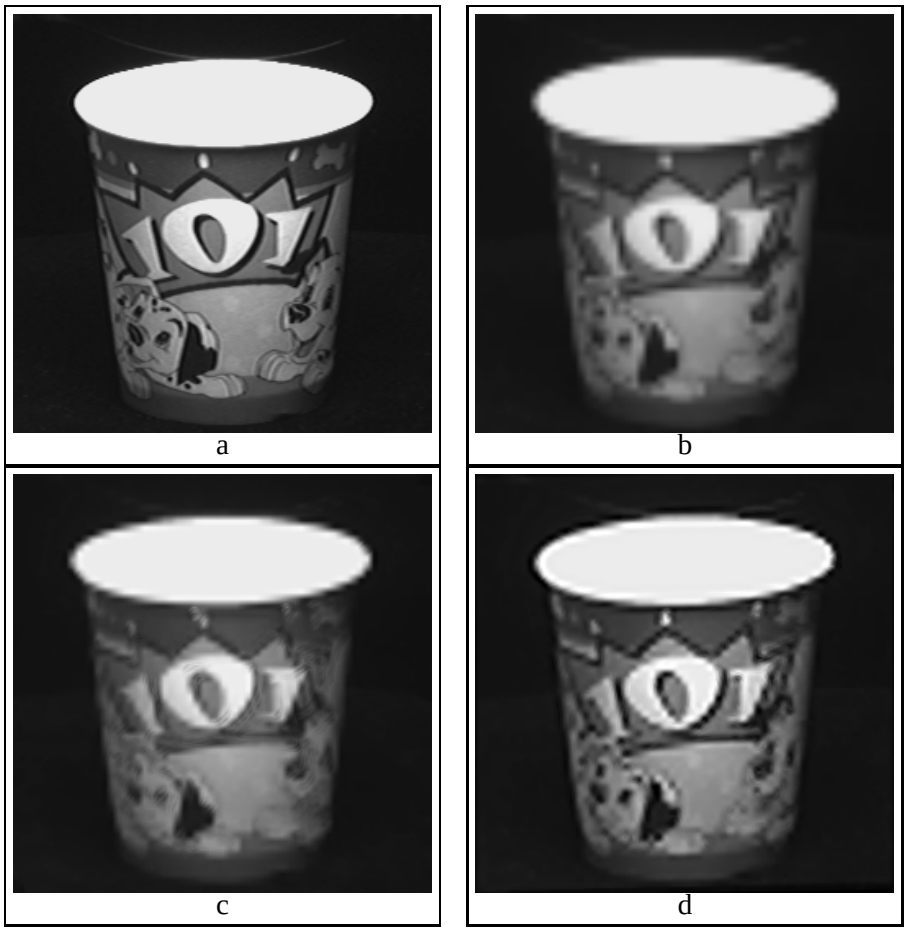
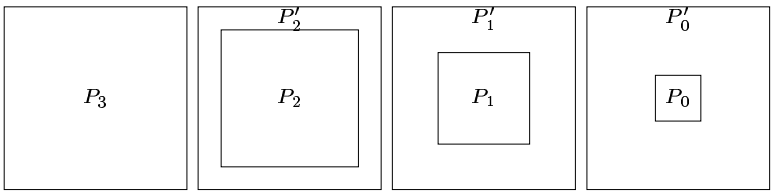


Figure 1.7. Results from a sequence of five synthetically down-sampled images. (a) the original image; (b) the scale-up of the down-sampled version of (a) by a factor of four using bilinear resampling (no super-resolution); (c)–(d) super-resolution from the synthetically down-sampled image sequence using, respectively, bilinear resampling with deblurring and QRW with deblurring.

128x128 region, and the power of the central 64x64 region of the image spectrum. The column marked P_2', P_1', and P_0' show, respectively, the power of the whole region minus the power of the central 192x192 region, the power of the whole region minus the power of the central 128x128 region, and the power of the whole region minus the power of the central 64x64 region. As is to be expected, most of the power concentrates in the central 64x64 region which is the frequencies that can be directly represented in the low-resolution images. Outside this region, the power is relatively small. Obviously, super-resolution using QRW with deblurring does considerably better.



a

Method	P_3	P_2	P_1	P_0	P_2'	P_1'	P_0'
Original	11058.09	11052.15	11036.98	10928.85	5.93	21.11	129.24
Bi-linear	10722.40	10722.14	10721.59	10714.03	0.25	0.80	8.37
SRBD	10712.98	10712.14	10711.22	10698.52	0.83	1.76	14.46
SRQRWD	11060.74	11059.61	11058.35	11030.00	1.13	2.39	30.74

b

Table 1.3 . Power of the Fourier transform of the images shown in Fig. 1.7. (a) Regions in the computation of the powers of the 2D Fourier transform. (b) the power within those regions with P_3 being the whole region (i.e., the whole spectrum); P_2, P_1 and P_0 being the regions inside the inner squares (192x192, 128x128, and 64x64, respectively); Primed labels indicated the complementary area of a region. Recall the original images were 64x64, so P_0 is associated with the power representable in those images, and P_0' what was gained by processing. Bi-linear is simple warping with bi-linear without super-resolution or debluring (i.e reconstruction of a single input image). SRBD is super-resolution using bi-linear warping followed by debluring, and SRQRWD is super-resolution using QRW followed by debluring.

Figs. 1.8 shows the difference between the a slice of the 2D Fourier spectra (scanline 128 i.e. DC component in y) of the super-resolution reconstructions and the ground truth high-resolution image shown in Fig. 1.7a. Clearly this spectral analysis example shows the superiority, both quantitatively and qualitatively, of QRW warping over simple bilinear warping or bilinear super-resolution. However, the quantitative analysis was very crude as it was based only on the power in the various regions. The next two sections shows the superiority of QRW using task-based quantitative metrics.

5.2. OCR-BASED EVALUATION

In this section, we discuss our evaluation using OCR, which we refer to as *OCR-based measurement.* The fundamental idea of this approach is to use OCR as our fundamental metric (as the name suggests). The evaluation consists of three basic steps:

1 Obtain the super-resolution images using the super-resolution algorithm described in Section 4.

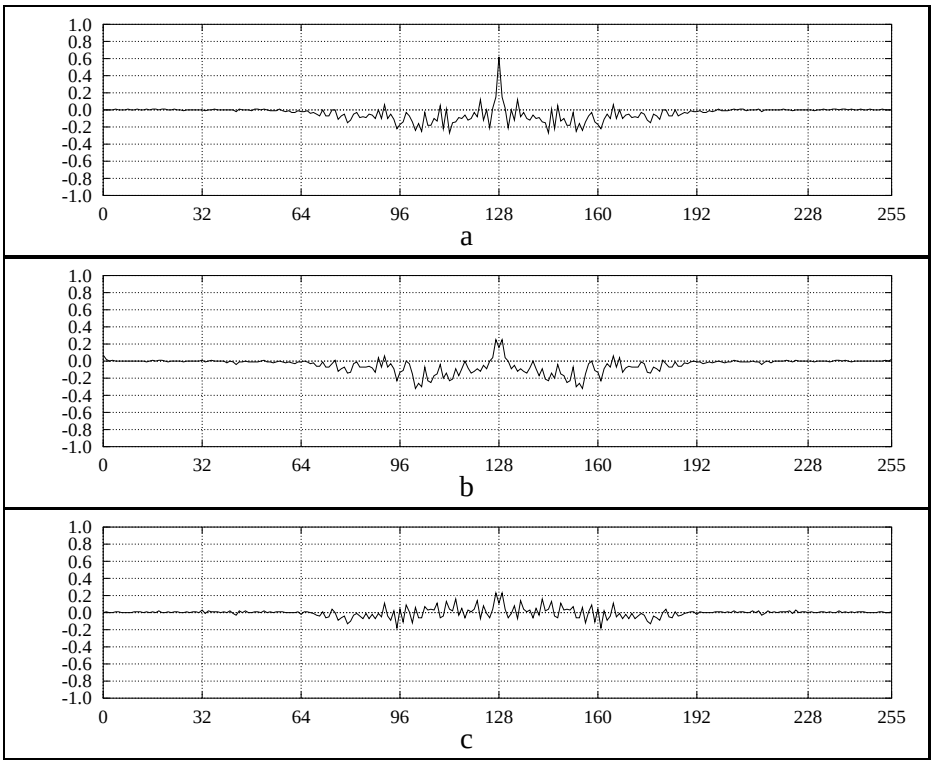


Figure 1.8. Display of $|F^S_{128}(u)| - |F^O_{128}(u)|$ where $|F^S_{128}(u)|$ and $|F^O_{128}(u)|$ are the Fourier spectra in x, sampled at the the DC component (row 128) in y) for the various super resolution approximations and the original image Fig. 1.7. (a) shows the difference using bilinear resampling (no super-resolution) and the original; (b) shows the difference between super-resolution using bilinear resampling with deblurring and the original and (c) showing the difference between super-resolution using QRW with deblurring and the original.

2 Pass the super-resolution results obtained in the previous step through a "character-oriented" OCR system.

3 Determine the number of characters recognized, i.e., the rate of recognition.

The goals of this evaluation is to quantify the effectiveness of super-resolution. Evaluations herein are made by comparing the super-resolution results and those using bilinear warping.

While most "state-of-the-art" OCR programs can use dictionary lookup to aid in their recognition, we chose to use a pure character based system. This ensures that the system's behavior is driven only by the image input and not significantly impacted by the grammatical context of examples. We have also chosen to use a font-independent system, i.e., one that is not trained on the font

and resolution being used. Training the OCR system might allow the training to compensate for poor, but consistent, behavior in resampling or super-resolution.

Since the OCR program is not trained on the particular font, we break our analysis of errors up into two categories. The first error measure compares the OCR output to the ground-truth input characters. We use C_r^i to indicate the number of characters correctly recognized. We consider multiple types of errors, including C_i, C_m, C_e, and C_s which give, respectively, the number of incorrectly recognized characters, the number of missing characters, the number of extra characters, and the number of characters that are split into two or more characters, with $C_{i+m+e+s}$being the sum of these. %Correct indicates the percentage of characters correctly recognized (C_r^i divided by $C_{i+m+e+s}$). For many fonts, some characters are so visually similar that without the use of training or context, distinguishing pairs of characters is quite difficult, e.g., `0` vs `O`, `1` vs `l` vs `!` vs `|`, and in some fonts, / vs *l* vs *t* and *h* vs *b* (see Fig. 1.13). In the context of our experiments, Fig. 1.9 contains three ambiguous characters; Fig. 1.11 four ambiguous characters; Fig. 1.13 seventeen ambiguous characters.

For brevity, we also use the abbreviations as shown in Table 1.4 to describe the algorithms discussed.

Abbreviation	Meaning
BRX	bilinear resampling without distortion correction and deblurring
BRDCD	bilinear warping with distortion correction and deblurring
BR	bilinear resampling without deblurring
BRD	bilinear resampling with deblurring
SR	super-resolution using QRW without deblurring
SRD	super-resolution using QRW with deblurring
SRDCD	super-resolution using QRW with distortion correction and deblurring

Table 1.4 . Abbreviations of algorithms considered. Distortion correction is needed to remove radial lens distortions common in inexpensive lenses.

The OCR program used for the experiments described herein is "Direct for Logitech, Version 1.3." The images used are normalized with respect to the super-resolution image with deblurring, as follows:

$$I_n = \frac{\bar{I}_s}{\bar{I}_u} I_u$$

where I_n is the normalized image, $\bar{I}_s$ is the average of the intensity values of the super-resolution image with deblurring, and $\bar{I}_u$ is the average of the intensity values of the image to be normalized. Within each dataset, the same threshold is used to binarize all the images.

The test data shown in this section was taken using laboratory quality imaging systems, a Sony XC-77 camera, attached to either a Datacube MV200 System or a Matrox Meteor Capture Card. As is to be expected, better imaging reduces the need for super-resolution; lower quality cameras increase the significance of super-resolution imaging.

All examples herein are scale-up by a factor of four, with the distance between camera and sample being changed so that the scaled images would yield an image with character sizes within the range accepted by the OCR system. We qualitatively evaluated the approach on a wider range of fonts and imaging conditions. Note that fonts with thinned letters, such as the "v" in Fig. 1.11, tend to be broken into multiple letters. Characters in slanted serif fonts tend to connect and thus, fail to be recognized. Inter-word spacing is not handled well (and multiple spaces are ignored in our measures). The ease of finding a good threshold depends on the uniformity of the lighting, image contrast, and lens quality. Better OCR algorithms may remove most or all of these difficulties. The quantitative examples show a few of these features, but in general, we choose examples that are not dominated by these artifacts.

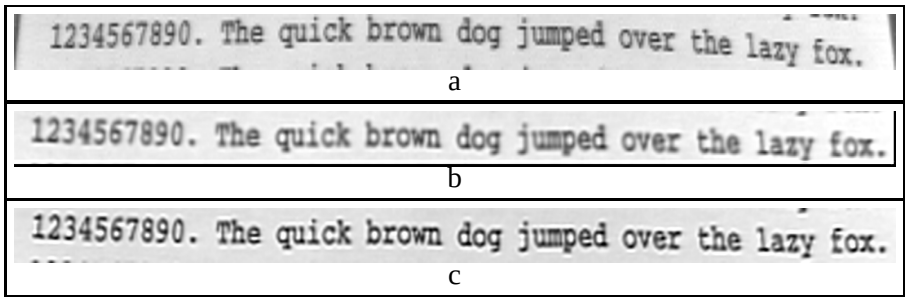


Figure 1.9. Super-resolution results from a sequence of 32 391x19 images taken by a Sony XC-77 Camera. (a) one of the original images scaled up using bilinear resampling and without distortion correction; (b) and (c) the results after distortion correction, with (b) showing bilinear warping with deblurring and (c) showing super-resolution using QRW with deblurring.

Fig. 1.9 shows the super-resolution results from a sequence of 32 391x19 images taken by a Sony XC-77 camera attached to a Datacube MV200 System before they are passed through OCR for recognition. Fig. 1.9a shows one of the original images scaled up using bilinear resampling and without distortion correction. Fig. 1.9b and Fig. 1.9c show the results after distortion correction, with Figs. 1.9b showing bilinear warping with deblurring and Figs. 1.9c showing super-resolution using QRW with deblurring.

Figs. 1.10 summarizes the results of passing the super-resolution results shown in Fig. 1.9 through OCR for recognition. This example did not contain any font-related ambiguities. The original text (see Fig. 1.9) consists of total

Method	Output of OCR
BRX	1234561990. te gu<clc bow dog jumped over te sazs
BRDCD	:23*;56?990. X:e quick brown dog jt:Wed over the Lazf for
SRDCD	2234567890. The quick brown dog jumped over the lazes 'ox.

Method	%Correct	C_r^i	C_i	C_m	C_e	C_s	$C_{i+m+e+s}$	
BRX	67	32	7	8	0	1	16	
BRDCD	71	35	11	1	0	1	14	
SRDCD	94	45		2	0	0	1	3

Figure 1.10. Output of OCR for the first example, the text shown in Fig. 1.9. The smaller size of text and more significant distortions make the impact of super-resolution using QRW very dramatic.

48 characters, including the two periods but excluding whitespace. Columns marked C_i, C_m, C_e, and C_s give, respectively, the number of incorrectly recognized characters, the number of missing characters, the number of extra characters, and the number of characters that are split into two or more characters.

Because of the nonuniformity of the lighting in this example, each image had its own threshold which was chosen to maximize its recognition rate. Using bilinear resampling without distortion, correction, and deblurring (BRX), 32 out of the 48 characters (67%) are recognized. Using bilinear warping with distortion correction and deblurring (BRDCD), 35 out of the 49 characters it found (71%) are recognized. Using the super-resolution algorithm given in Section 4 with deblurring (SRDCD), 45 out of the 48 characters (94%) are recognized. Compared to bilinear resampling without distortion correction, super-resolution using QRW recognizes 27% more characters. Compared to bilinear with distortion correction and deblurring, super-resolution using QRW recognizes 21% more of characters. With text consisting of thousands of characters, this is definitely a significant improvement.

Qualitatively, one might note the errors are concentrated on the outer edges of the example where there was the most significant distortions and the worst lighting.

Fig. 1.11 shows the super-resolution results for an example with larger characters (almost twice the size) taken with a better lens (much less distortion and less blur). The input was a sequence of 8 430x75 images taken by a Sony XC-77 camera attached to a Matrox Meteor Capture Card before they are passed through OCR for recognition. The original text consists of total 66 characters including three periods and one exclamation mark.

Fig. 1.12 shows the experimental results of passing the super-resolution results shown in Fig. 1.11 through OCR. While the bilinear and QRW super-resolution (SRD) images look nearly identical, the quantitative OCRanalysis

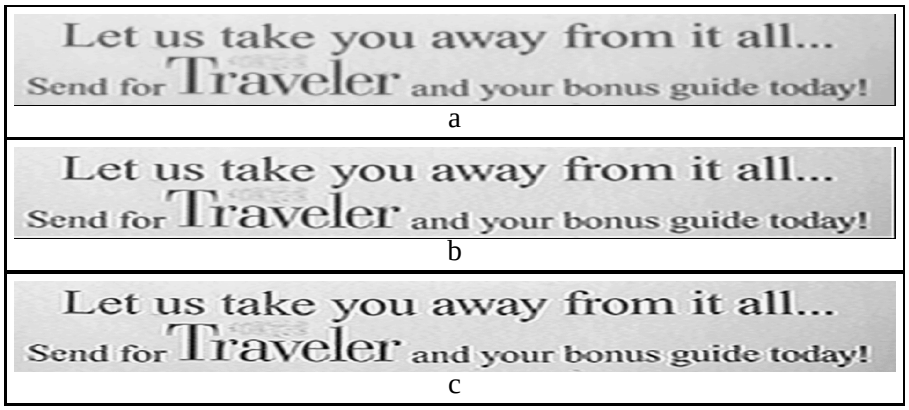


Figure 1.11. Super-resolution results from a sequence of 8 430x75 images taken by XC-77. (a) one of the original images scaled up using bilinear warping; (b) (a) deblurred; (c) super-resolution using QRW with deblurring.

Method	Output of OCR	C_i	C_m	C_e	C_s
BR	Let us take you assay from it all...	0	0	0	1
	Send fior TI aveler and your bonus gtude todyl	1	2	2	1
BRD	Let us take you assay from it all...	0	0	0	1
	Send for Tl aveler and your bonus guide today!	0	0	0	1
SR	Let us take you away from it all...	0	0	0	0
	Send fior Tl alreler and your bonus guide today	0	1	1	2
SRD	Let us take you away from it all...	0	0	0	0
	Send for Tra@&eler and your bonus guide today!	0	0	0	1

Method	%Correct	C_r^i	C_i	C_m	C_e	C_s	$C_{i+m+e+s}$
BR	89.7	61	1	2	2	2	7
BRD	97.0	64	0	0	0	2	2
SR	94.0	63	0	1	1	2	4
SRD	98.5	65	0	0	0	1	1

Figure 1.12. Results of the OCR test for the second example, shown in Fig. 1.11. Again debluring helped in both cases and super-resolution using QRW with debluring was the best algorithm.

shows a difference. Compared to bilinear resampling without deblurring (BR), 7% more of the characters are recognized with QRW. Compared to bilinear warping with deblurring (BRD), 2% more of the characters are recognized. 2% may mean a lot depending on applications. Compared to bilinear resampling without deblurring (BR), super-resolution with deblurring (SRD) reduces the number of incorrectly recognized, missing, extra, and split characters reduces from 7 to 1.

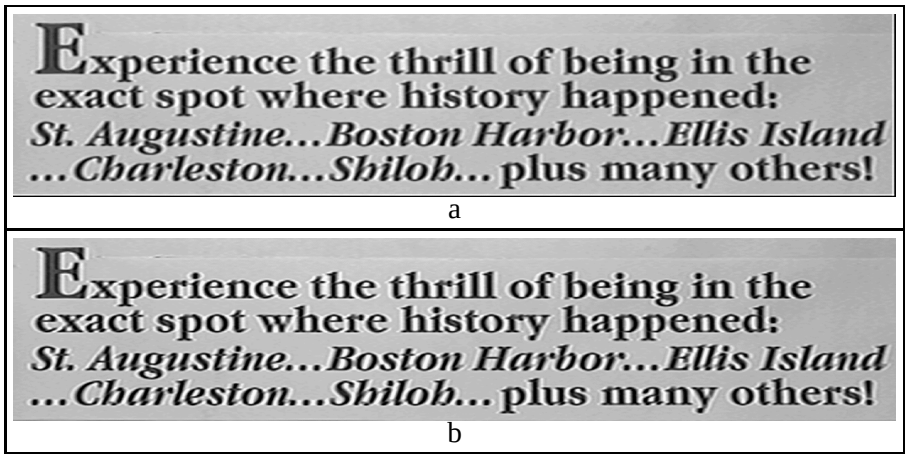


Figure 1.13. The third example sequence of 8 490x140 images taken by XC-77. The top (a)shows the results using bilinear warping with debluring and the bottom (b) shows the results of super-resolution using QRW with deblurring. While the images look nearly identical, the quantitative OCRanalysis shows a difference.

Fig. 1.14 shows the analysis for a third quantitative experiment. The original text, Fig. 1.13 consists of total 142 characters with mixed fonts and a large number of ambiguous characters. Compared to bilinear resampling without deblurring (BR), 5% more of the characters are recognized. Compared to bilinear resampling with deblurring (BRD), 3% more of the characters are recognized. Again, 3% could mean a lot depending on applications – if there were 2000 characters on a page, it is difference of 60 characters. Discounting the ambiguous characters will increase the rate of recognition by 2% for all methods except bilinear (BR) which increases only 1%. Compared to bilinear resampling with or without deblurring (BR or BRD), super-resolution with deblurring (SRD) reduces the number of incorrectly recognized, missing, extra, and split characters reduces from 18 to 12. Looking at the details one can see that the italics was handled poorly. While a multi-font OCRsystem might do much better overall, the example does show that super-resolution improves the results.

5.3. OCREXPERIMENTS SUMMARY

The qualitative aspects of our experimental results can be summarized as follows:

- Naturally, the better the quality of the original images and the larger the input characters, the smaller the impact of super-resolution on OCR. But even for large clear text, it did have a measurable impact.

Method	Output of OCR	C_i	C_m	C_e	C_s
BR	Experlence the thrill of belag in the	3	0	0	0
	exact spot where history happened	0	1	0	0
	St. Augustfne...Boston Hahbor...gWs Istand	5	2	0	0
	...Cbarleston@..SbSlob... plus maay othersl	0	7	0	0
BRD	Experience the thrill of being in the	0	0	0	0
	exact spot where history happened:	0	0	0	0
	St. Augustfne. . .Boston Hahbor...E s Island	3	2	2	0
	... CbarZeston...Sbilob... plus Expeothersl	9	1	1	0
SR	Experience the thrill of belag in the	2	0	0	0
	exact spot where history happened	0	1	0	0
	Stg Augustine...Boston Hahbor...SWs Island	4	2	0	0
	...Cbarlesto@t...Sbf/ob.. plus many othersl	6	1	0	1
SRD	Experience the thrill of being in the	0	0	0	0
	exact spot where history happened:	0	0	0	0
	Sf. Augustxne...Boston Harbor...Sis Istand	4	2	0	0
	...Cbar/eston...Sbi/oh.. plus many othersl	5	1	0	0

Method	%Correct	C_r^a	C_i	C_m	C_e	C_s	$C_{i+m+e+s}$
BR	87.3	124	8	10	0	0	18
BRD	87.7	128	12	3	3	0	18
SR	88.1	126	12	4	0	1	17
SRD	91.5	130	9	3	0	0	12

Figure 1.14. Performance of OCR testing for the third experiment. As you can see by looking at the recovery detail, the italics was poorly handled. Overall super-resolution using QRW and deburing performed the best.

- If there is no motion, minimal warping and good large text, super-resolution will not help much more than simple temporal averaging with bilinear warping.

- Type style has a strong impact on the rate of recognition.

This section has shown how OCRcan be used used for super-resolution evaluation. The advantages of this approach is that it is very straight forward; there are a large number of both commercial and free OCRpackages, and data collection is also straightforward. The use of OCRis well suited for evaluation of super-resolution tasks which will be similar in nature to OCR, e.g. license plates, 2D pattern matching, handwritting recongition.

In general, the difficulties of using OCR as a measurement for super-resolution can be summarized as follows:

- The rate of recognition depends, to a large extent, on the quality of the OCR programs. Better OCRprograms, especially those that use dictio-

nary lookup, would reduce the impact of low level processing. But in general, better low level processing would provide better results.

- If binary images are required for recognition, as most of the OCR programs do even if implicitly converted internally, then the rate of recognition is sensitive to the thresholding process used to convert gray-scale or color images to binary images. This evaluation used external thresholding and different thresholds may give different results. While localized thresholding would help increase the rate of recognition, we have not used them here.

- Many OCR programs treat their inputs as "binary", thus as an evaluation for super-resolution techniques, it may seem to down-play the importance of accurate grayscale production, especially at middle intensity levels. On the other hand, these intermediate levels do occur on character boundaries and may, in fact, be the driving factor in the superiority of super-resolutions. However the nearly binary nature of the data may suggest that over enhancement might do better.

6. FACE-BASED EVALUATION

In this section, we evaluate how SR can be used to improve the performance of a face recognition (FR) system. A formal description of the FR problem is given first, followed by an evaluation of a simulated SR-enhanced FR sytem.

Face Recognition Systems

One view of an FR system is a facility that provides a mapping from facial images to labels that uniquely identify the subject of the image. We assume that given an FR system, there exists some image set of known subjects, also known as a *gallery*, which we denote as G. In addition, there exists some *probe set* $P = \{p_1, p_2, \ldots, p_{|P|}\}$ where each image, $p_i \in P$ and $p_i \notin G$, is an image of some subject the FR system to recognize. In the system considered here, assume that the gallery G and the probe set P are non-identical, but have been created from the same population.

Most FR systems, presented with a gallery image g and probe image p_i, have the capability of computing some bounded similarity measure $s(g, p_i)$ representing the "strength" of the match between the images. Without loss of generality, assume that a score $s(g, p)$ of 10.0 indicates the highest possible system confidence that the subject in image g is the same subject in image p, and that a score $s(g, p)$ of 0.0 indicates the lowest possible system confidence that the subjects are the same (or highest confidence that they are different).

Let $\mathrm{id}(x)$ represent the true identity of the subject in image x, and g_h represent the gallery image of subject h. Given a probe p, a vector similarity scores

Algorithm	Description
s_0	FaceIt *Image to Scan Template*
s_1	FaceIt *Image to Full Tempalte* (Normal mode)

Figure 1.15. **Comparison operators. The precise comparison operators used in our evaluation and their names as described by the FaceIt documentation.**

$s(g, p)$ can be calculated from all images $g \in G$. Sorting the similarity vector and finding the subject's relative position along it, determines the probe's *rank*. Specifically, a probe has a rank of n over gallery set G if in the similarly vector, there exist exactly n scores greater than or equal to $s(g_{\mathrm{id}(p)}, p)$.

The algorithm used in the evaluation was Visionics' FaceIt. FaceIt is a commercial, high-quality, face detection and recognition system based on Local Feature Analysis. FaceIt requires no explicit "training" stage; i.e. no internal component of the system encorpores information across the entire gallery. This is different from many linear subspace classifiers (such as Eigenface approaches) which must be trained with respect to the entire gallery. With FaceIt, adding an image to the gallery does not require a retraining, and has no side effects on the probe to gallery image comparison. In other words, given a particular gallery and probe image g and p, the addition of a new gallery image g' has no effect on the FaceIt similarity measure $s(g, p)$. Naturally, the rank could be effected. For our evaluation, we selected two different similarity measures, described in Figure 1.15.

Experimentation

The context of the evaluation consisted of a subset of the Essex database packaged with FaceIt. Due to constraints in the available data, we selected 352 image pairs. Each pair consists of two images of the same subject with approximately the same pose, but slightly different expression. These images, in FERET nomenclature, are referred to as A and B images. All subjects selected were front facing, imaged in front of simple (essentially monochromatic) backgrounds, and diffusely illuminated. Since we were more interested in the effects of the image preprocessing stages, we presented favorable data to the system.

More formally, let $q^i = \{q_{(i,A)}, \ldots, q_{(i,D)}\}$ represent the set of four images of subject i and $Q^c = \{q_{(1,c)}, \ldots, q_{(L,c)}\}$ represent the set of all c images — in our case, $c \in \{A, B, C, D\}$. Let $f(q)$ represent some image processing function on image q (this will soon be replaced with a super-resolution algorithm), s represent a similarly measure, and $\mathrm{rank}_s(G, p)$ represent the rank of probe p over gallery G via similarity measure s. Then, using G, P, s, and f, an evaluation that obtains a set of ranks (one set per probe image) can be denoted

as

$$\mathit{eval}(G, P, s, f) = \bigcup_{p \in P} \mathit{rank}_s(G, f(p)). \tag{7}$$

So that confidence intervals could be obtained, the evaluation framework of (**?**) was followed. This framework is based on population stratification and methods of replicate statistics (*balanced repeated replicates* or BRR, specifically). Letting each subject correspond to a stratum, three samples per stratum (or PSUs) are obtained by fixing one set of images as the gallery and probing the gallery with the remaining data sets. That is, one set of samples was obtained with $\mathit{eval}(Q^A, Q^B, s, f)$, another with $\mathit{eval}(Q^A, Q^C, s, f)$ and so on.

Dataset Generation In this section we describe the probe and gallery sets used in our evaluation. The notation used here will also be used to describe the results of the evaluation.

A series of low-resolution images to serve as input to the SR algorithm is generated first. Based on previous results (Chiang, 1998), it was decided that four low-resolution input images would be used. To simluate a low-resoultion face from slightly different views, a perspective projection was used. Let m represent a random, but known, perspective and scalar pixel-to-pixel mapping, where the image width and height are reduced to 25 percent of their original size. In this evaluation, the perspecive projection was limited so that only a pixels horizontal coordinate would be displaced by at most 10% of the original image width. Let M represent the inverse operation — a 4 times dilation and perspective "correction." Since four low-resolution images were generated, four such m mappings needed. A set of four (distinct) mappings generated from a random seed k is denoted as $rvec(m)_k = \{m_k^1, m_k^2, m_k^3, m_k^4\}$. It follows that the set of complimentary mappings is denoted as $rvec(M)_k$.

Given a map m, image p, and warping algorithm a, a new low-resolution image p' can be denoted as a function of m and p, or $p' = a(m, p)$. Let r and b represent QRR and bilinear warping algorithms respectively. Then, for each probe image $p_i \in P$, a set of four low-resolution images generated from warping algorithm a is

$$P_i^a = \bigcup_{m_i^k \in rvec(m)_i} a(m_i^k, p_i) = \{a(m_k^1, p_i), \ldots, a(m_i^4, p_i)\}. \tag{8}$$

If we denote a SR_a as a SR algorithm using warp a, then we may now we modify Equation (7), our previous definition of *eval*, to reflect SR_a

$$\mathit{eval}(G, P, s, a) = \bigcup_{p \in P} \mathit{rank}_s(G, \mathit{SR}_a(P_i^a)). \tag{9}$$

In this ideal case, the SR algorithm generates the corrective maps by directly inverting each distortion m. Naturally, this is only possible when m is known

file=low-res-images.eps,width=3.5in,height=1.2in

Figure 1.16. **Low-resolution images of Subject 202**. Example of a set of four, perspective projection distored, low resolution images which serve as SR algorithm input. All images were generated from the higher-resolution A image of subject 202. Closer inspection reveals subtle difference between the images: the first face appears narrower and slighly elongated with respect to the other three faces, the fourth appears slighly smaller, lower, and wide than the other three, etc. The black stripes on the left and right side of the images are artifacts resulting from the projection. Because each subject has their own set of unique maps, some low-resolution image sets will show more or less variation between images

a priory. Note that other phenomenon, such as sensor noise, deformations in expression, and changes in illumination, are not encorporated into the evaluation. By using ideal conditions the evaluation, such as nearly ideal inverse mappings, our evaluation is not only significantly simplied, but better reflects a more "upper bound"-like measure of the effects of SR.

As reflected in Equation (9) changing a SR algorithm has no effect on the particular maps (m and M) used. This critical constancy ensures that differences in the resulting ranking can be attributed only to the change in warping algorithm.

6.1. EXPERIMENTATION AND RESULTS

To establish a peformance gain (or loss) for SR, the raw low-resolution images will be used as the baseline evaluation. In a system without SR, these raw images are what would be used as probe images. The ideal SR images give an indication of the "best" possible performance gain due to SR.

Unfortunately, *rank*, as defined previously, can be a non-robust measure. The penalty incurred by a misdetection (defined as a rank greater than t, where t is some predefined threshold) is linearly related. For example, suppose a probe has a very high (poor) rank — 400, for instance. The penalty incurred in the mean rank due to this 400-rank misdetect is much greater than a probe with a rank of, say 50. In both cases, however, if our threshold $t = 10$, then they are both clearly misdetects. Therefore, for our evaluation, we perform the BRR mean and variance over the following statistic, where r represents a probe's rank and α is some threshold rank

$$\theta_n = 1 \text{if } r ¡ \alpha, \theta_n = 0 \text{otherwise} \tag{10}$$

Another view of θ_n is the expected value of the fraction of probes within the top α match candidates. For this particular evaluation, we used $alpha = 0.01$. In other words, a probe scored a 1 if it subject was within the top α of candidate images (just the top image in our case).

The generation of multiple low-resolution images provides a broad baseline. Let $Q^c_{(a,i)}$ represent the set of all ith low-resolution images generated from

Essex databaset set Q^c using warping algorithm a. In other words, the set $Q^D_{(b,3)}$, or

$$\{b(m_1^3, q_{(1,D)}), b(m_1^3, q_{(2,D)}), \ldots, b(m_1^3, q_{(L,D)})\} \tag{11}$$

is the set of all low-resolution images from set Q^C generated from a map of index 3. Note that this partitioning is somewhat arbitrary, and only dependent on particular indexes. This notation will be used again shortly. Similarly, we let

$$Q^A_{(r,\text{SR})}, \ldots, Q^D_{(r,\text{SR})}, Q^A_{(b,\text{SR})}, \ldots, Q^D_{(b,\text{SR})}. \tag{12}$$

denote the super-resolution image sets generated from their respective low-resolution images. Finally, there the BRR mean estimates generated from each ($Q^A_{(a,i)}$, ..., $Q^D_{(a,i)}$) set (see (**?**) for more information about this construction). These are denoted by replacing the set index with the symbol μ. For example

$$Q^\mu_{(b,1)}, Q^\mu_{(r,1)}, \ldots, Q^\mu_{(b,4)}, Q^\mu_{(r,4)}, Q^\mu_{(b,\text{SR})}, Q^\mu_{(r,\text{SR})} \tag{13}$$

These BRR estimators incorporate rank information across all A, ..., D sets (appropriately).

The raw results of the evaluation are shown in Figure 1.17. For each row, the table shows the expected value of the fraction of probes that produce ranks of 0. The BRR estimated standard error of these means is shown in parenthesis. As shown in the figure, both QRW and bilinear based super-resolution improves the fraction of recognized faces with a statistical significance.[1] For all experiments, QRW super-resolution produced better fractions then bilinear. This is not always the case for the component fractions. This indicates that in certain cases, it is possible that a particular low-resolution image may be better than a super-resolution image. In a real face recognition system, however, ground truth is not available. Therefore, it would be impossible to know *which* low-resolution produces the correct result. Nevertheless, this phenomenon is more of a face-recognition system issue than a super-resolution issue. It should be noted that increasing α (for $\alpha < 10$) does not dramatically change the fact that QRW produces statistically significant higher fractions.

The figure shows the results from both of the FaceIt algorithms. In the first, the overall peformance is much better and we see that super-resoltion helps for both bilnear warping and for QRW. It also shows that QRW is statsitically supperior, being at least three standard deviations above bilnear-based super-resolution. The second algorithm, overall, did not perform as well. The behaivor of QRW is consistent — again showing a statistically significant improvement for super-resolution over the individual low-resolution inputs. However, this algorithm had cases where bilinear outperformed QRW on the

[1] This is a statistically sound statement which is dependent on the unique properties of BRR.

probe sets	alg s_0	alg s_1
$Q^{\mu}_{(b,1)}$	0.8277 (0.0083)	0.6591 (0.0114)
$Q^{\mu}_{(r,1)}$	0.8390 (0.0087)	0.6382 (0.0119)
$Q^{\mu}_{(b,2)}$	0.8438 (0.0085)	0.6591 (0.0112)
$Q^{\mu}_{(r,2)}$	0.8570 (0.0086)	0.6705 (0.0125)
$Q^{\mu}_{(b,3)}$	0.8219 (0.0082)	0.6609 (0.0113)
$Q^{\mu}_{(r,3)}$	0.8589 (0.0081)	0.6430 (0.0115)
$Q^{\mu}_{(b,4)}$	0.8286 (0.0088)	0.6866 (0.0112)
$Q^{\mu}_{(r,4)}$	0.8589 (0.0081)	0.6335 (0.0120)
$Q^{\mu}_{(b,\mathrm{SR})}$	0.8494 (0.0079)	0.6647 (0.0115)
$Q^{\mu}_{(r,\mathrm{SR})}$	0.8731 (0.0078)	0.6875 (0.0109)

Figure 1.17. **Mean Fraction of Success Estimates**. The results of the evaluation, showing the expected value of the fraction of the probes that have rank 0 (most likely candidate). Standard devations are shown in parenthesis.

indivisual examples. However, this examples shows that unlike QRW, with bi-linear warping, super-resolution was not better than the individual inputs.

7. CONCLUSION

This chapter discussed techniques for image consistent reconstruction and warping using the integrating resampler. By coupling the degradation model of the imaging system directly into the integrating resampler, we can better approximate reconstruct the image and the warping characteristics of real systems. This, in turn, significantly improves the quality of super-resolution images. Examples of super-resolutions for gray-scale images show the usefulness of the integrating resampler in applications scaling by a factor of upto 4 using 8-32 images. We disussed three quantittive evaluations approaches and in each case saw that super-resolution using QRW was superior to bilinear approaches. Even in the cases where the super-resolution images were visually similar, we had measurable quantitative improvements.

References

Andrews, H. and Hunt, B. (1977). *Digital Image Restoration.* Prentice-Hall, Englewood Cliffs, NY.

Andrews, H. and Patterson, C. (1977). Digital interpolation of discrete images. *IEEE Trans. Computers*, C-25:196–202.

Bascle, B., Blake, A., and Zisserman, A. (1996). Motion deblurring and super-resolution from an image sequence. *Computer Vision—ECCV*, pages 573–581.

Boult, T. and Wolberg, G. (1992). Correcting chromatic abberations using image warping. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.*

Boult, T. and Wolberg, G. (1993). Local image reconstruction and subpixel restoration algorithms. *CVGIP: Graphical Models and Image Processing*, 55(1):63–77.

Chiang, M. (1998). *Imaging-Consistent Warping and Super-Resolution.* PhD thesis, Department of Computer Science, Columbia University.

Chiang, M. and Boult, T. (1996). The integrating resampler and efficient image warping. *Proceedings of the DARPA Image Understanding Workshop*, pages 843–849.

Chiang, M. and Boult, T. (2000). Super-resolution via imaging warping. *Image & Vision Computing Journal Special Issue.*

Drago, F. and Granger, E. (1985). Optics requirements for making color fiche transparencies of maps, charts, and documents. *Image Quality: An Overview Proc. SPIE*, 549:25–33.

Elad, M. and Feuer, A. (1999). Superresolution reconstruction of image sequences.

Fant, K. (1986). A nonaliasing, real-time spatial transform technique. *IEEE Computer Graphics and Applications*, 6(1):71–80. See also "Letters to the Editor" in Vol.6 No.3, pp. 66-67, Mar 1986 and Vol.6 No.7, pp. 3-8, July 1986.

Gonzalez, R. and Wintz, P. (1987). *Digital Image Processing.* Addison-Wesley, Reading, MA.

Granger, E. (1974). Subjective assessment and specification of color image quality. *SPIE Proc. Image Assessment and Specification*, page 86.

Gross, D. (1986). Super-resolution from sub-pixel shifted pictures. Master's thesis, Tel-Aviv University.

Hou, H. and Andrews, H. (1987). Cubic splines for image interpolation and digital filtering. *IEEE Trans. Acoust. Speech, Signal Process.*, ASSP-26:508–517.

Huck, F., Alter-Gartenberg, R., and Z.-U. Rahman (1991). Image gathering and digital restoration for fidelity and visual quality. *CVGIP: Graphical Models and Image Processing*, 53:71–83.

Irani, M. and Peleg, S. (1991). Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3):231–239.

Irani, M. and Peleg, S. (1993). Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, 4(4):324–335.

Jain, A. (1989). *Fundamentals of Digital Image Processing.* Prentice-Hall, Englewood Cliffs, NY.

Keren, D., Peleg, S., and Brada, R. (1988). Image sequence enhancement using sub-pixel displacements. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* pages 742–746.

Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech, Signal Process.,* ASSP-29:1153–1160.

Maeland, E. (1988). On the comparison of interpolation methods. *IEEE Trans. Medical Imaging,* MI-7(3):213–217.

Mitchell, D. and Netravali, A. (1988). Reconstruction filters in computer graphics. *Computer Graphics (SIGGRAPH '88 Proceedings),* 22(4):221–228.

Nene, S., Nayar, S., and Murase, H. (1994). SLAM: Software Library for Appearance Matching. Tech. Rep. CUCS-019-94, Columbia University, Department of Computer Science.

Oakley, J. and Cunningham, M. (1990). A function space model for digital image sampling and its application in image reconstruction. *CVGIP: Graphical Models and Image Processing,* 49:171–197.

Park, S. and Schowengerdt, R. (1982). Image sampling, reconstruction, and the effect of sample-scene phasing. *Applied Optics,* 21(17):3142–3151.

Park, S. and Schowengerdt, R. (1983). Image reconstruction by parametric cubic convolution. *CVGIP: Graphical Models and Image Processing,* 23:258–272.

Parker, J. and and D.E. Troxel, R. K. (1983). Comparison of interpolating methods for image resampling. *IEEE Trans. Medical Imaging,* MI-2(1):31–39.

Pavlidis, T. (1982). *Algorithms for Graphics and Image Procesing.* Computer Science Press, Rockville, MD.

Peleg, S., Keren, D., and Schweitzer, L. (1987). Improve image resolution using subpixel motion. *Pattern Recognition Letter,* pages 223–226.

Pratt, W. (1978). *Digital Image Processing.* John Wiley & Sons, New York, NY.

Pratt, W. (1990). *Digital Image Processing.* John Wiley & Sons, New York, NY.

Reichenbach, S. and Park, S. (1989). Two-parameter cubic convolution for image reconstruction. *Proc. SPIE Visual Communications and Image Processing,* 1199:833–840.

Schultz96.

Simon, K. (1975). Digital image reconstruction and resampling for geometric manipulation. *Proc. IEEE Symp. on Machine Processing of Remotely Sensed Data,* pages 3A–1–3A–11.

Traub, J., Wasilkowski, G., and Wozniakowski, H. (1988). *Information-Based Complexity.* Academic Press, New York.

Wolberg, G. (1990). *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, California.

Wolberg, G. and Boult, T. (1989). Image warping with spatial lookup tables. *Computer Graphics (SIGGRAPH '89 Proceedings)*, 23(3):369–378.