Linear models for multi-frame super-resolution restoration under non-affine registration and spatially varying PSF.

Sean Borman[a] and Robert L. Stevenson

Laboratory for Image and Signal Analysis
Department of Electrical Engineering
University of Notre Dame, Notre Dame, IN 46556, U.S.A.

ABSTRACT

Multi-frame super-resolution restoration refers to techniques for still-image and video restoration which utilize multiple observed images of an underlying scene to achieve the restoration of super-resolved imagery. An *observation model* which relates the measured data to the unknowns to be estimated is formulated to account for the registration of the multiple observations to a fixed reference frame as well as for spatial and temporal degradations resulting from characteristics of the optical system, sensor system and scene motion. Linear observation models, in which the observation process is described by a linear transformation, have been widely adopted. In this paper we consider the application of the linear observation model to multi-frame super-resolution restoration under conditions of *non-affine image registration* and *spatially varying PSF*. Reviewing earlier results,[1] we show how these conditions relate to the technique of image warping from the computer graphics literature and how these ideas may be applied to multi-frame restoration. We illustrate the application of these methods to multi-frame super-resolution restoration using a Bayesian inference framework to solve the ill-posed restoration inverse problem.

Keywords: Super-resolution, multi-frame restoration, image warping, image resampling, image restoration, inverse problem, Bayesian restoration.

1. INTRODUCTION

It is hardly surprising that multi-frame super-resolution restoration techniques, which offer the prospect of enhanced resolution seemingly for "free," have been the focus of considerable attention in recent years. In a wide variety of imaging modalities, a sequence of related images derived from an underlying scene may be easily obtained. By utilizing the totality of the information comprising the observed image sequence simultaneously, multi-frame restoration methods can exceed the resolving ability of classical single-frame restoration techniques. The mechanism whereby multi-frame restoration methods can reverse the degrading effects of aliasing in imaging systems which undersample the observed scene is well understood.[2] Although these ideas are no longer new, their application to general imaging environments remains challenging.

At the heart of all multi-frame restoration methods is an *observation model* which relates the observed image sequence measurements to the unknown restoration which is to be estimated. The observation model often incorporates a model for noise which may corrupt the observation process. Additionally, in modern restoration methods, it is common to incorporate *a-priori* knowledge regarding the characteristics of the solution to be estimated. The goal of the multi-frame restoration process is the estimation of a super-resolved image or sequence given the observed data, the observation model and whatever *a-priori* knowledge may be available.

As the restoration process seeks to "invert" the effects of the observation process, multi-frame restoration is an example in the broad category of mathematical *inverse problems*. We show that in practice the multi-frame restoration problem is typically *ill-posed*. In order to address this problem, the technique of mathematical *regularization* is applied in order to achieve acceptable restoration.

[a]Author to whom correspondence should be addressed.
Robert L. Stevenson rls@nd.edu
Sean Borman Sean.Borman.1@nd.edu

In this paper we demonstrate how to generalize the widely used linear observation model to accommodate non-affine geometric image registration and spatially varying imaging system degradations. This effort represents a contribution in the application of multi-frame restoration techniques to more general imaging systems and environments.

We begin, in Section (2), with a discussion on *image warping*, a technique commonly used in computer graphics in which digital images are resampled under geometric transformations. In the computer graphics literature, image warping is often considered synonymous with *image resampling*, despite the fact that the concept of resampling is more general than the (typically spatial) geometric transformations considered in computer graphics applications. Despite this mild abuse of terminology, we shall nevertheless utilize the term *resampling* interchangeably with *warping*. The basic theory of image resampling is presented and after considering an idealized image resampling process, we focus on a realizable resampling system which is applicable to both affine and non-affine geometric transformations.

In Section (3) we develop a multi-frame observation model which is appropriate to the super-resolution restoration problem. We show that there are close similarities between image resampling theory and the model we propose for multi-frame restoration, with the result that we may utilize ideas from image resampling in application to multi-frame restoration. We develop a linear but spatially varying observation filter which relates the multiple observed images to the unknown super-resolution restoration.

A practical algorithm for computing the linear observation filter is presented in Section (4) and in Section (5) we show how this filter may be represented in a classical linear inverse problem formulation of the restoration problem. In Section (6) we propose using a Bayesian maximum *a-posteriori* (MAP) estimation procedure to solve this inverse problem, using a Markov random field prior model. The optimization problem that results from the requirement to maximize the *a-posteriori* probability is convex and thus computationally quite feasible despite the typically large number of unknowns.

In Section (7) we present an example of the application of these methods to the restoration of a super-resolved frame from a short image sequence. We summarize the findings of this paper in Section (8).

2. IMAGE RESAMPLING

Image resampling refers to methods for sampling discrete images under coordinate transformation. Using the terminology of the computer graphics community, image resampling or *image warping* is the process of geometrically transforming a sampled source image or *texture image* to a sampled output image or *warped image* using a coordinate transformation called a *warp*. Given the source image $f(\boldsymbol{u})$, defined only at samples $\boldsymbol{u} \in \mathbb{Z}^2$ and the continuous, invertible coordinate transformation or warp, $H : \boldsymbol{u} \mapsto \boldsymbol{x}$ we wish to find the sampled output image $g(\boldsymbol{x})$ on the lattice $\boldsymbol{x} \in \mathbb{Z}^2$.

A naïve algorithm for image resampling might go as follows: for each pixel $\boldsymbol{x} \in \mathbb{Z}^2$ in the warped image g, let $g(\boldsymbol{x}) = f(H^{-1}(\boldsymbol{x}))$. Inherent to this approach are at least two problems; firstly, $H^{-1}(\boldsymbol{x})$ need not fall on sample points, thus necessitating some form of interpolation of the source image f; secondly, and more seriously, $H^{-1}(\boldsymbol{x})$ may undersample $f(\boldsymbol{u})$ resulting in aliasing which is manifested as objectionable artifacts in warped images in areas where the mapping results in minification. In the quest for realism in computer graphics rendering, addressing these artifacts spurred the development of image resampling theory as well as the development of many optimized techniques for fast, high-quality, texture rendering.

2.1. Conceptual image resampling pipeline

Heckbert was one of the first researchers to address these complications in his now classic work.[3] Following Heckbert's exposition, an ideal image resampler may be considered to be composed of a four stage process, enumerated below and illustrated in Fig. (1). In these equations, functions of a continuous variable are denoted with a subscript "c" and $\circledast$ denotes convolution:

1. Reconstruct a continuous version $f_c(\boldsymbol{u})$ of the sampled texture image $f(\boldsymbol{u})$ with $\boldsymbol{u} \in \mathbb{Z}^2$ using the reconstruction kernel $r(\boldsymbol{u})$.

$$f_c(\boldsymbol{u}) = f(\boldsymbol{u}) \circledast r(\boldsymbol{u}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{k}) \cdot r(\boldsymbol{u} - \boldsymbol{k}). \quad (1)$$

2. Use the inverse mapping $H^{-1}(\boldsymbol{x})$ which associates locations $\boldsymbol{x}$ in the warped output image with locations $\boldsymbol{u}$ in the texture image to find the continuous warped image $g_c(\boldsymbol{x})$.

$$g_c(\boldsymbol{x}) = f_c\left(H^{-1}(\boldsymbol{x})\right). \tag{2}$$

3. Apply the anti-alias prefilter $p(\boldsymbol{x})$ to the warped output image $g_c(\boldsymbol{x})$ to produce a prefiltered image $g'_c(\boldsymbol{x})$ suitable for sampling.

$$g'_c(\boldsymbol{x}) = g_c(\boldsymbol{x}) \circledast p(\boldsymbol{x}) = \int g_c(\boldsymbol{\alpha}) \cdot p(\boldsymbol{x} - \boldsymbol{\alpha})\, \mathrm{d}\boldsymbol{\alpha}. \tag{3}$$

4. Sample the prefiltered image $g'_c(\boldsymbol{x})$ to produce the discrete output image $g(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathbb{Z}^2$.

$$g(\boldsymbol{x}) = g'_c(\boldsymbol{x}) \quad \text{for} \quad \boldsymbol{x} \in \mathbb{Z}^2. \tag{4}$$

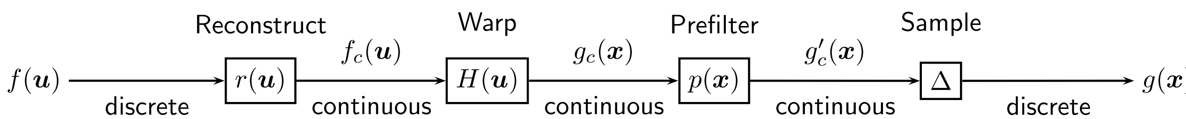

Figure 1. Heckbert's conceptual image resampling pipeline.

2.2. Realizable image resampling pipeline

Although the four-stage resampling pipeline described above is conceptually useful, in practice construction of the intermediate functions f_c, g_c and g'_c can be avoided. Starting with Eqn. (4) and working backward to Eqn. (1) it is possible to derive an expression for the warped output which does not require the intermediate steps.

$$g(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = g'_c(\boldsymbol{x}) = \int f_c\left(H^{-1}(\boldsymbol{\alpha})\right) \cdot p(\boldsymbol{x} - \boldsymbol{\alpha})\, \mathrm{d}\boldsymbol{\alpha} = \int p(\boldsymbol{x} - \boldsymbol{\alpha}) \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{k}) \cdot r\left(H^{-1}(\boldsymbol{\alpha}) - \boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}. \tag{5}$$

By defining

$$\rho(\boldsymbol{x}, \boldsymbol{k}) = \int p(\boldsymbol{x} - \boldsymbol{\alpha}) \cdot r\left(H^{-1}(\boldsymbol{\alpha}) - \boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}, \tag{6}$$

we may write Eqn. (5) in the succinct form,

$$g(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{k}) \rho(\boldsymbol{x}, \boldsymbol{k}), \tag{7}$$

where $\rho(\boldsymbol{x}, \boldsymbol{k})$ is a *spatially varying* resampling filter. Thus the warped image $g(\boldsymbol{x})$ may be computed using only a discrete filtering operation. The resampling filter of Eqn. (6) is described in terms of the warped reconstruction filter r with the integration performed over $\boldsymbol{x}$-space. With a change of variables $\boldsymbol{\alpha} = H(\boldsymbol{u})$, and integrating in $\boldsymbol{u}$-space the resampling filter can be expressed in terms of the warped prefilter p as,

$$\rho(\boldsymbol{x}, \boldsymbol{k}) = \int p(\boldsymbol{x} - H(\boldsymbol{u})) \cdot r\left(\boldsymbol{u} - \boldsymbol{k}\right) \left|\frac{\partial H}{\partial \boldsymbol{u}}\right| \mathrm{d}\boldsymbol{u}, \tag{8}$$

where $|\partial H / \partial \boldsymbol{u}|$ is the determinant of the Jacobian of the transformation H. This form for the resampling filter will be useful in the sections ahead.

2.3. Affine and non-affine geometric transformations

In the case of an affine geometric transformation H, the Jacobian $\left|\frac{\partial H}{\partial \boldsymbol{u}}\right|$ is a constant. The resulting resampling filter $\rho(\boldsymbol{x}, \boldsymbol{k})$ can be shown to be *space invariant*,[3] thus radically simplifying the resampling process as the output image may be computed by a simple convolution with the resampling filter which may be accomplished efficiently using Fourier-domain methods. For non-affine warps, however, the Jacobian is not constant and the associated resampling filter is spatially varying. Computing the warped output image then involves the application of a spatially varying resampling filter which is computationally far more costly than the affine case.

3. A MULTI-FRAME OBSERVATION MODEL

In this section we develop the multi-frame observation model which is used in formulating the super-resolution restoration problem. Comparing this observation model with the theory of image resampling developed above, we note that these two problems share common characteristics. We begin by modeling a continuous-discrete process where P sampled, low-resolution, images are captured from an underlying continuous scene. The observed imagery may constitute a temporal image sequence obtained from a single (typically moving) camera, or may derive from multiple cameras observing the scene.

Our objective is the estimation of the underlying scene, computed on a discrete, super-resolution sampling lattice. This is achieved using a discrete-discrete observation model derived from the continuous-discrete model. *Super-resolution* describes a process of bandwidth extrapolation; not merely an increase in the number of image samples. Super-resolution restoration is possible since the totality of information available from the image sequence as a whole is typically greater than that available from any single image of the scene.

3.1. Continuous-discrete multi-frame observation model

Let the observed sampled low-resolution image sequence consist of P images $g^{(i)}(\boldsymbol{x}),\ i \in \{1, 2, \ldots, P\}$ derived from a continuous scene $f_c(\boldsymbol{u})$. The observed images are geometrically related to $f_c(\boldsymbol{u})$ via the coordinate transformations $H^{(i)} : \boldsymbol{u} \mapsto \boldsymbol{x}$ which account for the relative motion between the scene and the camera. Associated with each observation is a (possibly spatially varying) point spread function (PSF) $h^{(i)}$, which may differ from observation to observation. The PSFs model the lens and sensor response, defocus, motion blur and so on. Finally the images are sampled. The continuous-discrete multi-frame observation process is illustrated in Fig. (2) and is described by,

$$g^{(i)}(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \int h^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha}) \cdot f_c\left(H^{(i)^{-1}}(\boldsymbol{\alpha})\right) \mathrm{d}\boldsymbol{\alpha}. \tag{9}$$

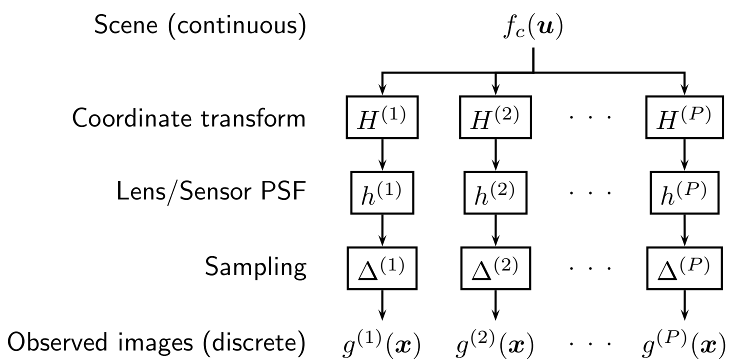

Figure 2. Continuous-discrete multi-frame observation model.

3.2. Discrete-discrete multi-frame observation model

We seek a discretized approximation of the continuous scene $f_c(\boldsymbol{u})$ on a high-resolution sampling lattice given the observed images $g^{(i)}$ and knowledge of the observation process characterized by $H^{(i)}$, $h^{(i)}$ and $\Delta^{(i)}$ for $i \in \{1, 2, \ldots, P\}$. The continuous scene $f_c(\boldsymbol{u})$ may be approximated by interpolating the discrete samples $f(\boldsymbol{k})$, $\boldsymbol{k} \in \mathbb{Z}^2$ using an interpolation kernel h_r as,

$$f_c(\boldsymbol{u}) \approx \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{V}\boldsymbol{k}) \cdot h_r\left(\boldsymbol{u} - \boldsymbol{V}\boldsymbol{k}\right) \tag{10}$$

where $\boldsymbol{V}$ is the sampling matrix[4]

$$\boldsymbol{V} = \begin{bmatrix} 1/Q_x & 0 \\ 0 & 1/Q_y \end{bmatrix},$$

and Q_x and Q_y are respectively the horizontal and vertical sampling densities of the super-resolution image estimate. Combining Eqn. (9) with Eqn. (10) then yields the relationship between the observed frames and the discrete approximation of the scene as,

$$g^{(i)}(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \int h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha}) \cdot f_c\left(H^{(i)^{-1}}(\boldsymbol{\alpha})\right) \mathrm{d}\boldsymbol{\alpha} = \int h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha}) \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{V}\boldsymbol{k}) \cdot h_r\left(H^{(i)^{-1}}(\boldsymbol{\alpha}) - \boldsymbol{V}\boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}. \quad (11)$$

3.3. Image resampling and the multi-frame observation model

Comparing Eqn. (11) with the image resampling filter in Eqn. (5), which is reproduced below for convenience,

$$g(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \int p(\boldsymbol{x}-\boldsymbol{\alpha}) \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{k}) \cdot r\left(H^{-1}(\boldsymbol{\alpha}) - \boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha},$$

it is apparent that there is a parallel between the image resampling equation and the multi-frame observation model. The spatially varying PSF $h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})$ associated with the observation $g^{(i)}(\boldsymbol{x})$ replaces the space-invariant image resampling prefilter $p(\boldsymbol{x}-\boldsymbol{\alpha})$ in the resampling equation, and the reconstruction filter $h_r(H^{(i)^{-1}}(\boldsymbol{\alpha})-\boldsymbol{V}\boldsymbol{k})$ has a counterpart in the reconstruction filter $r(H^{-1}(\boldsymbol{\alpha})-\boldsymbol{k})$ in the resampling equation. Noting these parallels, we are encouraged to proceed as we did in Section (2.2) and define the spatially varying filter,

$$\rho^{(i)}(\boldsymbol{x},\boldsymbol{k}) = \int h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha}) \cdot h_r\left(H^{(i)^{-1}}(\boldsymbol{\alpha}) - \boldsymbol{V}\boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}, \quad (12)$$

allowing us to write Eqn. (11) as,

$$g^{(i)}(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{V}\boldsymbol{k}) \cdot \rho^{(i)}(\boldsymbol{x},\boldsymbol{k}). \quad (13)$$

As in the resampling derivation, we make the change of variables $\boldsymbol{\alpha} = H^{(i)}(\boldsymbol{u})$, and integrate in $\boldsymbol{u}$-space so the observation filter $\rho^{(i)}(\boldsymbol{x},\boldsymbol{k})$ can be expressed in terms of the warped PSF $h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})$ as,

$$\rho^{(i)}(\boldsymbol{x},\boldsymbol{k}) = \int h^{(i)}\left(\boldsymbol{x}, H^{(i)}(\boldsymbol{u})\right) \cdot h_r\left(\boldsymbol{u}-\boldsymbol{V}\boldsymbol{k}\right) \left|\frac{\partial H^{(i)}}{\partial \boldsymbol{u}}\right| \mathrm{d}\boldsymbol{u}. \quad (14)$$

The parallels between image resampling and the multi-frame observation model are summarized in Table (1).

Image resampling		Multi-frame Observation model	
Discrete texture	$f(\boldsymbol{u})$	Discrete scene estimate	$f(\boldsymbol{u})$
Reconstruction filter	$r(\boldsymbol{u})$	Interpolation kernel	$h_r(\boldsymbol{u})$
Geometric transform	$H(\boldsymbol{u})$	Scene/camera motion	$H^{(i)}(\boldsymbol{u})$
Anti-alias pre-filter	$p(\boldsymbol{x})$	Observation SVPSF	$h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})$
Warped output image	$g(\boldsymbol{x})$	Observed images	$g^{(i)}(\boldsymbol{x})$
Resampling filter		Observation filter	
$\rho(\boldsymbol{x},\boldsymbol{k}) = \int p(\boldsymbol{x}-H(\boldsymbol{u})) \cdot r\left(\boldsymbol{u}-\boldsymbol{k}\right) \left\lvert\frac{\partial H}{\partial \boldsymbol{u}}\right\rvert \mathrm{d}\boldsymbol{u}$		$\rho^{(i)}(\boldsymbol{x},\boldsymbol{k}) = \int h^{(i)}\left(\boldsymbol{x},H^{(i)}(\boldsymbol{u})\right) \cdot h_r\left(\boldsymbol{u}-\boldsymbol{V}\boldsymbol{k}\right) \left\lvert\frac{\partial H^{(i)}}{\partial \boldsymbol{u}}\right\rvert \mathrm{d}\boldsymbol{u}$	

Table 1. Comparing image resampling with the multi-frame observation model.

4. DETERMINING THE OBSERVATION FILTER

The linear multi-frame observation process of Eqn. (13) relates the unknown scene f to the observed frames $g^{(i)}$ via the spatially varying reconstruction filters $\rho^{(i)}(\boldsymbol{x},\boldsymbol{k})$. The coefficients of the reconstruction filters must be determined according to the coordinate transformations $H^{(i)}$, the lens and sensor characteristics $h^{(i)}$ and the sampling characteristics $\Delta^{(i)}$ for $i \in \{1,2,\ldots,P\}$.

4.1. An algorithm for computing the warped pixel response

In Fig. (3) we present an algorithm for computing the coefficients of the spatially varying observation filter of Eqn. (14). In order to obtain a reasonably efficient algorithm implementation we make the simplifying assumption that the size of the restoration pixels is sufficiently small that the integral may be approximated by a weighted point sample. Also, we assume that the interpolation kernel is constant over the support of each restoration pixel (a box function) so that we may omit its evaluation and instead ensure that the sum of the filter coefficients is unity. Note that the function of the backprojection step in the algorithm is to reduce the range of $\boldsymbol{k}$ over which we need to evaluate the warped filter.

for each observed image $\boldsymbol{g}^{(i)}$ {
  for each pixel $\boldsymbol{x}$ {
    back-project the boundary of $h^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha})$ from $g^{(i)}(\boldsymbol{x})$ to the restored image space using ${H^{(i)}}^{-1}$
    determine a bounding region for the image of $h(\boldsymbol{x}, \boldsymbol{\alpha})$ under ${H^{(i)}}^{-1}$
    for each pixel indexed by $\boldsymbol{k}$ in the bounding region {
      set $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k}) = h^{(i)}\left(\boldsymbol{x}, H^{(i)}(\boldsymbol{u})\right) \cdot \left|\frac{\partial H^{(i)}}{\partial \boldsymbol{u}}\right|$ with $\boldsymbol{u} = \boldsymbol{V}\boldsymbol{k}$
    }
    normalize $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ so that $\sum_k \rho^{(i)}(\boldsymbol{x}, \boldsymbol{k}) = 1$
  }
}

Figure 3. Algorithm for computing the observation filter coefficients $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$.

4.2. Graphical illustration of the computation of the warped pixel response

The backprojection and projection process used to determine the filter response $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ in the algorithm discussed above is graphically illustrated in Fig. (4). First, an idealized scenario is considered, followed by an example showing actual filter coefficients computed for a realistic pixel model.

In Fig. (4)(a), the support of the response of an idealized observed pixel is illustrated as a darkly-shaded region. The image of this region under H^{-1} is shown as the darkly-shaded region in the restored image space in Fig. (4)(b). Also shown in Fig. (4)(b), as a lightly-shaded region, is the smallest integer-pixel bounding region containing the image of the observed pixel response under H^{-1}. A single high-resolution pixel contained in the lightly-shaded bounding region in Fig. (4)(b) is shown in white. The image of this high resolution pixel under H is illustrated by the small, white, warped region contained within the darkly-shaded region representing the observed pixel response in Fig. (4)(a). The ratio of the areas of the high resolution pixel and its image under H is given by the inverse of the determinant of the Jacobian of the transformation H.

In Fig. (4)(c) and Fig. (4)(d) we illustrate an example of the computation of the filter coefficients $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ for a realistic pixel model and geometric transformation. Fig. (4)(c) shows the spatial response of a sensor pixel characterized by a uniform spatial response and 100% fill-factor in a camera with a diffraction limited optical system. Under a non-affine projective transformation, the corresponding observation filter response is shown in Fig. (4)(d). Notice the warping of the pixel response function which results from the projective transformation relating the views. Also note that the regular sampling lattice in the super-resolution restoration space in Fig. (4)(d) maps under H to the irregular lattice in the space of the observed pixel response in Fig. (4)(c). This is most easily seen by examining the samples adjacent to the left and right edges of Fig. (4)(c). The gray-level samples in Fig. (4)(d) represent the values of the discrete observation filter associated with the observed pixel under the geometric transformation H.

5. THE LINEAR MULTI-FRAME OBSERVATION MODEL

In Eqn. (13), samples $g^{(i)}(\boldsymbol{x})$ in the observed images are related to the sampled approximation of the scene we seek to estimate $f(\boldsymbol{V}\boldsymbol{k})$ via a linear, but spatially varying, discrete filter kernel $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$. In practice the recorded images $g^{(i)}(\boldsymbol{x}),\ i \in \{1, 2, \ldots, P\}$ in Eqn. (9) are known only for a finite subset of $\boldsymbol{x} \in \mathbb{Z}^2$, namely

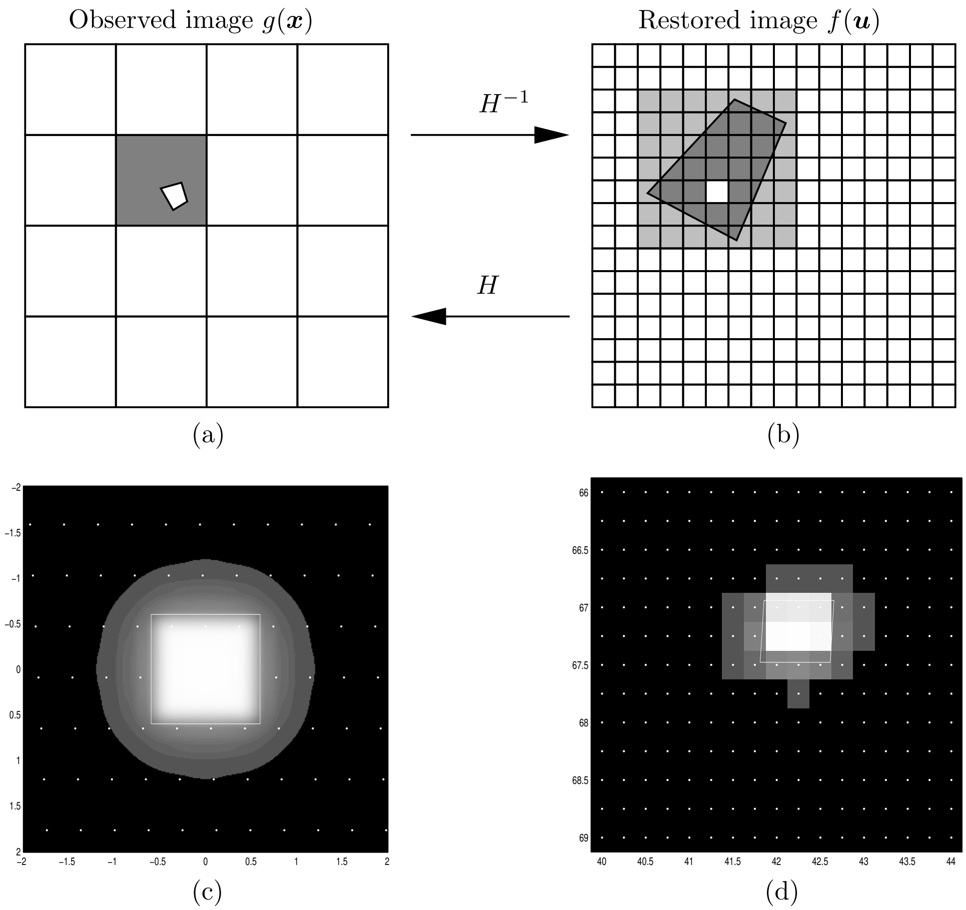

Figure 4. Determining the observation filter response (see text): (a) Support of the response of an idealized observed pixel. (b) Image of the idealized pixel response under the transformation H^{-1} and the associated bounding region. (c) Observed pixel response for realistic observed pixel (assumes diffraction limited optics, spatially uniform pixel response with 100% fill factor). (d) Observation filter $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ associated with the observed pixel under under the geometric transformation H.

on a regular lattice $\boldsymbol{x} \in \{1, 2, \ldots, N_r\} \times \{1, 2, \ldots, N_c\}$ where N_r and N_c are the number of image rows and columns respectively. Similarly the reconstruction of the original scene f in Eqn. (10) will be computed on a lattice consisting of $Q_y N_r$ rows and $Q_x N_c$ columns. Each observed image consisting of $N_r N_c$ pixels may be conveniently represented as a $N_r N_c$-element column vector $\boldsymbol{g}^{(i)}$ using a row-major ordering of the observed data and likewise, the reconstructed image may be represented by a $Q_y N_r Q_x N_c$-element column vector $\boldsymbol{f}$.

Referring to Eqn. (13), the value of the j^{th} pixel in the observed image $\boldsymbol{g}^{(i)}$ is related to the unknown reconstruction $\boldsymbol{f}$ via a linear combination of the unknown elements of $\boldsymbol{f}$ as,

$$\boldsymbol{g}_j^{(i)} = \sum_{k=1}^{Q_y N_r Q_x N_c} A_{jk}^{(i)} \boldsymbol{f}_k \tag{15}$$

where the matrix coefficients $A_{jk}^{(i)}$ are determined from the discrete kernel $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$. Note that the matrix containing the observation filter coefficients will be very sparse. Using matrix-vector notation we have,

$$\boldsymbol{g}^{(i)} = \boldsymbol{A}^{(i)} \boldsymbol{f}. \tag{16}$$

Since P images are observed, it is convenient to stack the observation vectors $\boldsymbol{g}^{(i)}$ and matrices $\boldsymbol{A}^{(i)}$ defining,

$$\boldsymbol{g} \doteq \left[\boldsymbol{g}^{(1)^T} \boldsymbol{g}^{(2)^T} \cdots \boldsymbol{g}^{(P)^T} \right]^T \quad \text{and} \quad \boldsymbol{A} \doteq \left[\boldsymbol{A}^{(1)^T} \boldsymbol{A}^{(2)^T} \cdots \boldsymbol{A}^{(P)^T} \right]^T$$

so that we may express the observation process in the classic form,

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}. \tag{17}$$

6. A BAYESIAN FRAMEWORK FOR RESTORATION

Eqn. (17) represents the linear observation model relating the unknown image $\boldsymbol{f}$ to be estimated, to the observed images $\boldsymbol{g}$. Solving for $\boldsymbol{f}$ is a classic example of an inverse problem. It is well known[5] that the multi-frame super-resolution restoration problem is ill-posed, so special care must be taken when estimating a value for $\boldsymbol{f}$. In this, and earlier work, we have employed a Bayesian framework for computing a meaningful solution to Eqn. (17).

We assume that the observation process is corrupted by an additive noise process. While arguably not the most realistic choice, for reasons of convenience it is commonly assumed that the noise process is additive, zero-mean Gaussian. The linear observation process of Eqn. (17) is thus augmented to include a noise term,

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{n}. \tag{18}$$

The probability density function of the zero-mean Gaussian noise term $\boldsymbol{n}$ is given by,

$$\mathcal{P}_{\boldsymbol{N}}\left(\boldsymbol{n}\right) = \frac{1}{(2\pi)^{\frac{PN_rN_c}{2}} |\boldsymbol{K}|} \exp\left\{ -\frac{1}{2} \boldsymbol{n}^T \boldsymbol{K}^{-1} \boldsymbol{n} \right\} \tag{19}$$

where $\boldsymbol{K}$ is the positive-definite noise covariance matrix.

Bayesian maximum *a-posteriori* (MAP) estimation is used to estimate the super-resolved image $\boldsymbol{f}$ given the observed data $\boldsymbol{g}$. The MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$ maximizes the *a-posteriori* probability $\mathcal{P}\left(\boldsymbol{f}|\boldsymbol{g}\right)$ as,

$$\hat{\boldsymbol{f}}_{\text{MAP}} = \arg\max_{\boldsymbol{f}} \left\{ \mathcal{P}\left(\boldsymbol{f}|\boldsymbol{g}\right) \right\} = \arg\max_{\boldsymbol{f}} \left\{ \frac{\mathcal{P}\left(\boldsymbol{g}|\boldsymbol{f}\right) \mathcal{P}\left(\boldsymbol{f}\right)}{\mathcal{P}\left(\boldsymbol{g}\right)} \right\} \tag{20}$$

where the expression in Eqn. (20) results from the application of Bayes' theorem. Taking logarithms and noting that the maximum is independent of $\mathcal{P}\left(\boldsymbol{g}\right)$, yields

$$\hat{\boldsymbol{f}}_{\text{MAP}} = \arg\max_{\boldsymbol{f}} \left\{ \log \mathcal{P}\left(\boldsymbol{g}|\boldsymbol{f}\right) + \log \mathcal{P}\left(\boldsymbol{f}\right) \right\}. \tag{21}$$

6.1. The likelihood function

Given the observation expression in Eqn. (18) the conditional probability $\mathcal{P}\left(\boldsymbol{g}|\boldsymbol{f}\right)$ is determined by the noise pdf in Eqn. (19) as

$$\mathcal{P}\left(\boldsymbol{g}|\boldsymbol{f}\right) = \mathcal{P}_{\boldsymbol{N}}(\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}) = \frac{1}{(2\pi)^{\frac{PN_rN_c}{2}} |\boldsymbol{K}|} \exp\left\{ -\frac{1}{2} (\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f})^T \boldsymbol{K}^{-1} (\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}) \right\}. \tag{22}$$

This expression is called the *likelihood function* given its interpretation as the likelihood of observing $\boldsymbol{g}$ given $\boldsymbol{f}$.

6.2. The prior probability density

In the Bayesian framework it is assumed that we have prior knowledge regarding the probability density of the super-resolution image random vector $\mathcal{P}(\boldsymbol{f})$. We utilize a Markov random field (MRF) model to impose a probability density function on the space of possible super-resolution restorations. Markov random fields are often used for image modeling applications as they provide a convenient framework for specifying a global probability density function by specifying only spatially local energy interactions.[6–9] By the Hammersley-Clifford theorem[10,11] the pdf of a MRF is given by the Gibbs distribution $\mathcal{P}(\boldsymbol{f}) = \frac{1}{k_p} \exp\{-\frac{1}{\beta}\mathcal{E}(\boldsymbol{f})\}$ where $\mathcal{E}(\boldsymbol{f})$

measures the energy of the configuration $\boldsymbol{f}$, k_p is a normalizing constant called the *partition function* and β is the MRF "temperature" parameter (terminology carried over from the statistical physics literature). In the case of a MRF model, the energy $\mathcal{E}(\boldsymbol{f})$ is specified in terms of local interactions between sets of spatially neighboring lattice locations. In particular we choose,

$$\mathcal{P}(\boldsymbol{f}) = \frac{1}{k_p} \exp\left\{ -\frac{1}{\beta} \sum_{c \in \mathcal{C}} \rho_T(\boldsymbol{\partial}_c \boldsymbol{f}) \right\} \tag{23}$$

where the summation is over the set of all "cliques" $\mathcal{C}$ which define the sets of interacting sites, with $\boldsymbol{\partial}_c$ computing the local spatial activity for each clique. The non-linear penalizing function $\rho_T(x)$ is taken to be the convex Huber penalty function,[12]

$$\rho_T(x) = \begin{cases} x^2 & |x| \le T \\ 2T|x| - T^2 & |x| > T \end{cases}. \tag{24}$$

Local spatial activity is computed using finite difference approximations to the second directional derivatives (vertical, horizontal and two diagonal directions) in the image $\boldsymbol{f}$, specifically,

$$\begin{aligned}
\boldsymbol{\partial}_1 \boldsymbol{f} &= f[n_1-1,n_2] - 2f[n_1,n_2] + f[n_1+1,n_2], \\
\boldsymbol{\partial}_2 \boldsymbol{f} &= f[n_1,n_2-1] - 2f[n_1,n_2] + f[n_1,n_2+1], \\
\boldsymbol{\partial}_3 \boldsymbol{f} &= \tfrac{1}{2} f[n_1+1,n_2-1] - f[n_1,n_2] + \tfrac{1}{2} f[n_1-1,n_2+1], \\
\boldsymbol{\partial}_4 \boldsymbol{f} &= \tfrac{1}{2} f[n_1-1,n_2-1] - f[n_1,n_2] + \tfrac{1}{2} f[n_1+1,n_2+1].
\end{aligned} \tag{25}$$

For pixels on the image boundaries, where it is not possible to compute these finite differences, appropriate first order differences are substituted.

This non-linear MRF image model imposes local image smoothness but nevertheless does allow for the formation of sharp edge features in the restored image.[5]

6.3. Objective function

Substituting the likelihood term in Eqn. (22) and the prior in Eqn. (23) into Eqn. (21) and removing constants independent of $\boldsymbol{f}$ gives the objective function

$$\begin{aligned}
\hat{\boldsymbol{f}}_{\text{MAP}} &= \arg\max_{\boldsymbol{f}} \left\{ -\frac{1}{2}(\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f})^T \boldsymbol{K}^{-1} (\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}) - \frac{1}{\beta} \sum_{c \in \mathcal{C}} \rho_T(\boldsymbol{\partial}_c \boldsymbol{f}) \right\} \\
&= \arg\min_{\boldsymbol{f}} \left\{ \frac{1}{2}(\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f})^T \boldsymbol{K}^{-1} (\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}) + \gamma \sum_{c \in \mathcal{C}} \rho_T(\boldsymbol{\partial}_c \boldsymbol{f}) \right\}
\end{aligned} \tag{26}$$

where $\gamma \doteq \frac{1}{\beta} > 0$ is a user-adjustable *regularization parameter* reflecting the degree to which the prior is favored over the observed data. Eqn. (26) shows that the objective function consists of two competing terms; the first measures the extent to which the estimate $\boldsymbol{f}$ is compatible with the observed data $\boldsymbol{g}$ while the second measures the conformance of the estimate to our prior expectations regarding the characteristics of the solution. The adjustable regularization parameter γ allows this balance to be controlled by the user. Methods for automatic selection of the regularization parameter have been studied,[13] but in this work γ is chosen experimentally.

6.4. Optimization

Determining the MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$ thus requires the minimization of the objective function in Eqn. (26). The term deriving from the likelihood function is convex in $\boldsymbol{f}$, but need not be *strictly* convex since the nullspace of the observation matrix $\boldsymbol{A}$ may be (and is typically) non-trivial. Thus, considering the likelihood term alone, there is no assurance of the uniqueness of a solution for $\hat{\boldsymbol{f}}_{\text{MAP}}$. Since the prior is described by a sum of convex functions of the operators $\boldsymbol{\partial}_c$ with $c \in \mathcal{C}$, if the intersection of the nullspace of the operators $\boldsymbol{\partial}_c$ and the nullspace of the observation matrix $\boldsymbol{A}$ is trivial, then we are assured that the solution will be unique.[13] In our case, the second derivative penalty functional does not have a trivial nullspace (the family of linear functionals incur identical cost) but in practice this rarely presents any difficulties. If necessary we may impose an additional regularizing term which chooses the minimum norm solution with an appropriate regularizing parameter which may be driven arbitrarily close to zero. In summary therefore, under very mild conditions, the existence and uniqueness of a global minimum, corresponding to the MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$ in Eqn. (26) can be assured and a simple gradient descent optimization[14] may be employed to compute the MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$. Despite the large number of unknowns, the optimization is tractable.

7. EXAMPLES

In this section we demonstrate the application of the super-resolution restoration framework discussed earlier to the restoration of a super-resolved image from a short test image sequence. We have utilized a completely synthetic simulated imaging environment in which the scene, camera and imaging geometry are all under our control.

7.1. Simulated imaging environment

The test "scene" consists of an algebraically-defined planar texture residing within the unit square of the xy-plane given by, $f(x, y) \propto \text{rect}(x)\text{rect}(y)(1 + \cos(40 \arctan(y/x)))$ and consists of 40 radial "spokes" emanating from the origin. This image was deliberately chosen not to be bandlimited.

A short image sequence consisting of seven non-interlaced gray-scale images are generated using ray-tracing methods which compute the images produced by a virtual camera which traverses the three-dimensional simulation space. The ray-tracing camera models the characteristics of a diffraction limited optical system and the spatial integration performed by a focal plane sensor array. The intrinsic characteristics of the simulated camera system are enumerated in Table (2) and selected extrinsic parameters of the seven camera positions and gaze points may be found in Table (3). Fig. (5) illustrates the imaging geometry with the camera positions and gaze points, as well as the seven simulated images corresponding to each camera position. Note the aliasing artifacts due to undersampling by the sensor array. Since the target scene is planar, under the pinhole camera model, the geometric relationship between observed frames is the projective transformation. Since the projective transformation is, in general, non-affine, the observation filter relating the restoration to each of the observed frames is spatially varying.

Projection model	Ideal pinhole
Image array dimensions	128×128
Pixel dimensions	$9\mu\text{m} \times 9\mu\text{m}$
Camera focal length	10 mm
Camera f/number	2.8
Illumination wavelength	550 nm
Diffraction limit cutoff	649.351 cycles/mm
Sampling rate	111.111 samples/mm
Folding frequency	55.5556 cycles/mm

Table 2. Camera intrinsic characteristics.

Camera center			Camera gaze point		
x	y	z	x	y	z
-3.0902	-5.0000	9.5106	0.0100	0.0050	0.0000
-1.0453	-5.0000	9.9452	0.0033	0.0017	0.0000
1.0453	-5.0000	9.9452	-0.0033	-0.0017	0.0000
3.0902	-5.0000	9.5106	-0.0100	-0.0050	0.0000
5.0000	-5.0000	8.6603	-0.0167	-0.0083	0.0000
6.6913	-5.0000	7.4315	-0.0233	-0.0117	0.0000
8.0902	-5.0000	5.8779	-0.0300	-0.0150	0.0000

Table 3. Camera extrinsic parameters.

7.2. Super-resolution restoration

We computed a super-resolution restoration corresponding to the last image in the sequence. The restoration increases the sampling density by a factor of four in each dimension, thus implying a sixteen-fold increase in the number of sample points as compared with a single observed frame. Given that the restoration is computed from only seven observed frames, it is clear that the number of unknowns to be estimated is considerably greater than the number of linear constraints available from the observations. This immediately implies that the restoration process is ill-posed as the solution fails to be unique. This uniqueness problem is addressed through the use of the MRF prior in the Bayesian restoration framework which stabilizes the inverse problem.

In Fig. (6) we show a cubic-spline interpolation of the seventh observed frame, while in Fig. (7) we show the result of a super-resolution restoration with the parameters $T = 1.5$, $\gamma = 0.05$ and the assumption of independent,

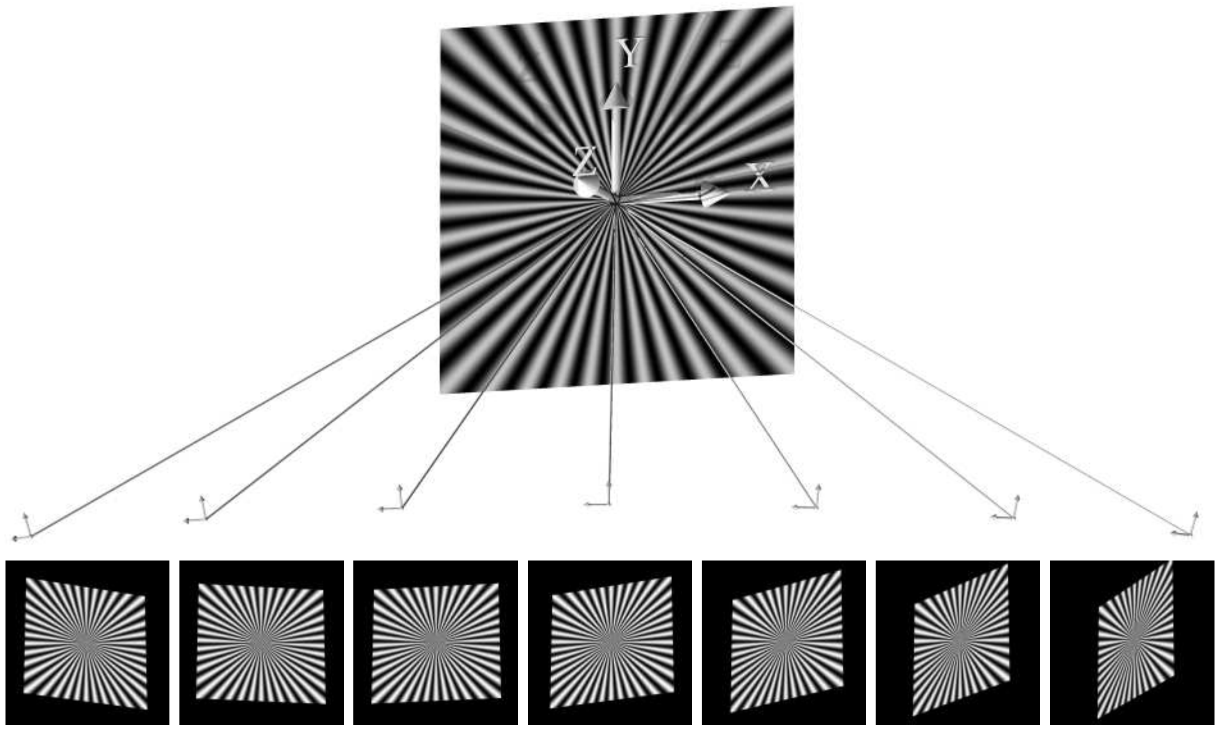

Figure 5. Simulated camera trajectory with corresponding observed low-resolution images.

identically distributed Gaussian noise with unit variance defining $\boldsymbol{K}$. A comparison of the two images clearly illustrates the considerable increase in resolution obtained using the multi-frame restoration method. Details near the center of the original image which were corrupted by aliasing have been restored in the super-resolution image.

8. CONCLUSIONS

In this paper we considered the application of a linear observation model to the multi-frame super-resolution restoration problem under the assumptions of non-affine image registration and spatially varying PSFs. Reviewing earlier work deriving from the computer graphics literature on image resampling theory, we established the relationship between these seemingly disparate areas of research. We showed that image resampling theory could be applied to the formulation of a linear but spatially-varying observation process that could account for non-affine geometric transformations for image registration as well as for spatially varying degradations in the imaging system. The solution to the resulting linear inverse problem relating the degraded observations to the desired unknown restoration was tackled using a Bayesian maximum *a-posteriori* (MAP) estimation formulation which utilizes prior information in the form of a Markov random field image model. The resulting optimization problem to obtain the MAP estimate was shown to be tractable due to the convexity of the objective function. We successfully demonstrated the application of these methods to the restoration of a super-resolved image from a short low-resolution aliased image sequence.

REFERENCES

1. S. Borman and R. Stevenson, "Image resampling and constraint formulation for multi-frame super-resolution restoration," in *Computational Imaging*, C. Bouman and R. Stevenson, eds., *Proceedings of the SPIE* **5016**, (San Jose, CA, USA), Jan. 2003.
2. R. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, R. Y. Tsai and T. S. Huang, eds., **1**, pp. 317–339, JAI Press Inc., 1984.

Figure 6. Cubic spline interpolation.

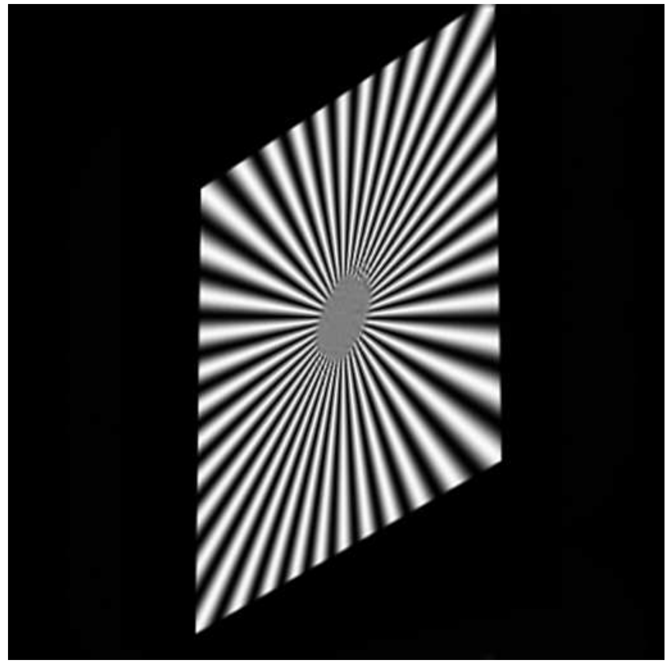

Figure 7. Multi-frame restoration.

3. P. Heckbert, "Fundamentals of texture mapping and image warping," Master's thesis, U.C. Berkeley, June 1989.
4. D. Dudgeon and R. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
5. R. Schultz and R. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing* **5**, pp. 996–1011, June 1996.
6. R. Kindermann and J. L. Snell, *Markov Random Fields and Their Applications*, American Mathematical Society, Providence, Rhode Island, 1980.
7. G. R. Cross and A. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-5**, pp. 25–39, Jan. 1983.
8. H. Derin and H. Elliot, "Modelling and segmentation of noisy and textured images using gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-9**, pp. 39–55, Jan. 1987.
9. R. C. Dubes and A. Jain, "Random field models in image analysis," *Journal of Applied Statistics* **16**(2), pp. 131–164, 1989.
10. P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, Springer-Verlag, 1999.
11. J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society B* **36**, pp. 192–236, 1974.
12. R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity Preserving Regularization of Inverse Visual Problems," *IEEE Transactions on Systems, Man and Cybernetics* **24**, pp. 455–469, Mar. 1994.
13. P. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, SIAM, 1998.
14. M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms, Second Edition*, John Wiley & Sons, 1993.