

Super resolution: an overview

C Papathanassiou and M Petrou
School of Electronics and Physical Sciences,
University of Surrey,
Guildford, GU2 7XH, United Kingdom
email: c.papathanassiou@surrey.ac.uk

Abstract—Super-resolution algorithms produce a single high-resolution image from a set of several, low-resolution images of the desired scene. The low-resolution frames are shifted differently with respect to the high resolution frame with subpixel increments. This paper presents first a theoretical overview of super-resolution algorithms. The most important methods, namely, the iterative back-projection, projection onto convex sets, and maximum a posteriori estimation are then compared within the same framework of implementation.

I. INTRODUCTION

In many applications of remote sensing, images of high resolution (HR) are required. Super-resolution (SR) achieves this by combining several low-resolution images of the desired scene. These LR images can be acquired either as a sequence over time by the same sensor, or taken at the same time with different sensors. A single HR image can then be constructed only if the LR frames are shifted with respect to the HR grid differently from each other and with sub-pixel increments. Thus each LR frame contains unique information which cannot be obtained from the other LR frames and this is exploited to obtain a HR image.

II. PROBLEM STATEMENT

Consider a high-resolution image of size $L_1 M_1 \times L_2 M_2$ written as a lexicographically ordered vector $\mathbf{z} = [z_1, z_2, \dots, z_N]^T$, where $N = L_1 M_1 L_2 M_2$. Note that $\mathbf{z}$ is considered to be generated by sampling a continuous scene. Let us also define $N_1 \equiv L_1 M_1$ and $N_2 \equiv L_2 M_2$.

If L_1 and L_2 represent the downsampling factors in the x and y directions respectively, then each observed low-resolution frame is of size $M_1 \times M_2$. Let P be the number of the available LR frames. In a similar manner, the k th LR frame may be expressed in lexicographical notation as $\mathbf{y}_k = [\tilde{y}_{k1}, \tilde{y}_{k2}, \dots, \tilde{y}_{kM}]^T$, for $k = 1, 2, \dots, P$, where $M = M_1 M_2$. The complete set of the LR frames is denoted by $\mathbf{Y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_P^T]^T = [y_1, y_2, \dots, y_{PM}]^T$.

First, an observation model relating the LR frames to the HR image should be formulated. The observed LR frames are assumed to have been produced by geometric warping, blurring, and uniform downsampling performed on the HR image $\mathbf{z}$. Moreover, each LR frame is typically corrupted by additive Gaussian noise which is uncorrelated between different LR frames. Thus, the k th LR frame may be written as

$$\mathbf{y}_k = \mathbf{D}\mathbf{B}_k\mathbf{T}_k\mathbf{z} + \mathbf{n}_k, \text{ for } k = 1, 2, \dots, P \quad (1)$$

where $\mathbf{T}_k$ represents a warping matrix of size $N \times N$ which models the motion that occurs during image acquisition, $\mathbf{B}_k$ is an $N \times N$ blurring matrix which can be either linear space invariant or linear space variant, and is caused by optical system problems (such as imaging systems out of focus, operating beyond the diffraction limit, suffering from aberration, etc.), relative motion between the imaging system and the imaged scene, and the point spread function (PSF) of the LR sensor, $\mathbf{D}$ is an $M \times N$ downsampling matrix which is employed to generate aliased LR frames from the warped and blurred HR image, and $\mathbf{n}_k$ denotes a lexicographically ordered M -dimensional noise field.

The observation model may be simplified as follows:

$$\mathbf{y}_k = \mathbf{W}_k\mathbf{z} + \mathbf{n}_k, \text{ for } k = 1, 2, \dots, P \quad (2)$$

where $\mathbf{W}_k \equiv \mathbf{D}\mathbf{B}_k\mathbf{T}_k$ is an $M \times N$ matrix.

Note that the above-mentioned imaging model assumes that the HR image $\mathbf{z}$ remains constant during the acquisition of the LR frames, except for any motion described by the warping matrix $\mathbf{T}_k$. Therefore, w_{kmr} are functions of the motion parameters of each LR pixel relative to the fixed HR grid. In consequence, (2) may be rewritten as

$$\tilde{y}_{km} = \sum_{r=1}^N w_{kmr}(\mathbf{s}_k) z_r + n_{km} \quad (3)$$

where $\tilde{y}_{km}$ is the value of the m th pixel of the k th LR frame, and vector $\mathbf{s}_k \equiv [s_{k1}, s_{k2}, \dots, s_{kQ}]^T$ encompasses the Q registration parameters for the k th LR frame with the coordinate system of the HR image. In practice, the motion parameters are not often known *a priori* and should be estimated along with the HR image $\mathbf{z}$.

III. SUPER-RESOLUTION METHODS

SR reconstruction techniques can be employed in the spatial or frequency domain. Spatial domain algorithms allow more flexibility in incorporating a priori constraints, noise models, and spatially varying degradation models. We compare here 3 spatial domain methods.

A typical SR image reconstruction algorithm consists of three stages, namely, registration, interpolation and restoration. These steps can be performed independently or simultaneously depending on the approach taken. Registration refers to the estimation of the relative shifts of each LR frame with respect to a reference LR image with subpixel accuracy. Since the shifts

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 25 JUL 2005		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Super resolution: an overview				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) School of Electronics and Physical Sciences, University of Surrey, Guildford, GU2 7XH, United Kingdom				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM001850, 2005 IEEE International Geoscience and Remote Sensing Symposium Proceedings (25th) (IGARSS 2005) Held in Seoul, Korea on 25-29 July 2005.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 4	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

between LR images are arbitrary, non-uniform interpolation is required to obtain a uniformly spaced HR image. Finally, image restoration is applied to the upsampled image to remove blurring and noise.

A. Iterative Backprojection (IBP)

Given an estimate of the HR image and a model of the imaging process, a set of simulated LR frames is generated. These simulated LR frames are then compared with the observed LR frames. The difference (error) between the two sets is used to correct the estimate of the HR image. The process is repeated iteratively until some stopping criterion is met, such as the minimization of the energy of the error, or until the maximum number of allowed iterations is reached. This method was formulated by Irani and Peleg [1], based on the ideas presented in [2] and [3].

Note that the LR frames are registered with each other to subpixel accuracy and an HR grid is defined in a certain relative shift from all LR frames. The shifts are not updated during the subsequent process which is aimed at defining values for the pixels of the HR grid.

The method can be used to incorporate constraints, such as smoothness or any other additional constraint which represents a desired property of the solution. The method employs a back-projection operator $\mathbf{H}^{BP}$. This can potentially affect the solution to which the iterative algorithm converges. The main practical weakness of the method lies exactly in the choice of $\mathbf{H}^{BP}$, especially when *a priori* constraints should be included. In addition, only linear constraints can be considered. Thus, although the IBP based methods provide a mechanism for constraining the SR reconstruction to conform with the observation data, they cannot resolve the problem of non-uniqueness. It can be proven that in the common case of a real and symmetric point spread function (PSF) $\mathbf{H}$, a possible choice of $\mathbf{H}^{BP}$ is $\mathbf{H}^{BP} = \mathbf{H}$ [2]. Irani and Peleg extend their work in [4] by considering a more general motion model.

B. Projection Onto Convex Sets (POCS)

The most prominent feature of the POCS formulation is the ease with which prior knowledge about the solution can be incorporated into the reconstruction process. However, for complex constraints the projection operators may be difficult to compute.

Consider the following constraints which are imposed by the complete set of the observed LR frames:

$$C_{m_1 m_2 k} = \{z(n_1, n_2) | r_k^{(z)}(m_1, m_2) \leq \delta_0\} \quad (4)$$

$$1 \leq m_1 \leq M_1, 1 \leq m_2 \leq M_2, k = 1, 2, \dots, P$$

where

$$r_k^{(z)}(m_1, m_2) = y_k(m_1, m_2) - \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} z(n_1, n_2) h(m_1, m_2; n_1, n_2) \quad (5)$$

$z(n_1, n_2)$ denotes the (n_1, n_2) th HR pixel, $y_k(m_1, m_2)$ the (m_1, m_2) th pixel of the k th LR frame, h the PSF function and δ_0 represents the uncertainty that we have in the observation.

Therefore, the residual $r_k^{(z)}(m_1, m_2)$, which is actually the observation noise $\mathbf{n}_k$, should be below the uncertainty level δ_0 which can be set according to the noise statistics. So, $C_{m_1 m_2 k}$ is the set of possible solutions that make the value of pixel (m_1, m_2) in the k th LR frame to differ by less than δ_0 from the observed value.

These constraints, which are also known as *data consistency constraints*, are closed and convex, and define a feasible solution $\mathbf{z}$ in the field of HR images for which the sensor outputs are the observed LR frames. Each constraint is defined for a single LR pixel, and thus there are $M = M_1 M_2$ such constraints for a single LR frame, and PM constraints for the complete LR dataset.

According to [5], the projection of an arbitrary image $z(n_1, n_2)$ onto $C_{m_1 m_2 k}$ is given in (6). If we enumerate pixels (m_1, m_2) by a single index, the above projection may be denoted as $\mathcal{P}_i \mathbf{z}$ where i takes values $1, 2, \dots, PM$, with PM being the total number of pixels in all LR available frames. Then these projection operators have to be applied in turn for all LR pixels. According to the fundamental theorem of POCS, having PM data consistency constraints $C_{m_1 m_2 k}$ for the complete set of LR images, the sequence

$$\mathbf{z}^{(n+1)} = \mathcal{P}_{PM} \mathcal{P}_{PM-1} \dots \mathcal{P}_2 \mathcal{P}_1 \mathbf{z}^{(n)} \quad (7)$$

converges to the desired HR image $\mathbf{z}$ for an initial estimate $\mathbf{z}^{(0)}$. Note that the reached solution is in general non-unique and depends on the initial guess. However, additional constraints may be imposed from prior knowledge to favor a particular HR image. Moreover, they enable robust performance in the presence of inconsistent or missing data. These constraints may be amplitude constraints, energy constraints, a reference image constraint, or a bounded support constraint. The projection operators of these constraints are relatively easy to formulate. More constraints may be defined in a similar manner, and incorporated in the iterative sequence of (7).

As in the IBP method, the registration parameters of all LR frames with respect to the HR grid and with respect to each other are computed once and assumed known throughout the whole process. The PSF which relates the HR pixels with the LR pixels is also assumed to be known. Thus the purpose of POCS is to assign values to the pixels of the HR grid.

C. Probabilistic Methods

According to this family of methods, the HR image $\mathbf{z}$ and the registration parameters $\mathbf{s}$ are considered as random fields described by the joint prior probability density function $P(\mathbf{z}, \mathbf{s})$. The steps of a typical maximum *a posteriori* (MAP) reconstruction method are described below. The task is to form a joint MAP estimate of $\mathbf{z}$ and $\mathbf{s}$ given the observed LR frames $\mathbf{y}$. Then the estimates can be computed as

$$(\hat{\mathbf{z}}, \hat{\mathbf{s}}) = \arg \max_{\mathbf{z}, \mathbf{s}} P(\mathbf{z}, \mathbf{s} | \mathbf{y}) \quad (8)$$

where $P(\mathbf{z}, \mathbf{s} | \mathbf{y})$ is the joint posterior probability of $\mathbf{z}$ and $\mathbf{s}$, conditioned on the observed $\mathbf{y}$.

$$\mathcal{P}_{m_1 m_2 k}[z(n_1, n_2)] = \begin{cases} z(n_1, n_2) + \frac{r_k^{(z)}(m_1, m_2) - \delta_0}{\sum_{l_1} \sum_{l_2} h^2(m_1, m_2; l_1, l_2)} h(m_1, m_2; n_1, n_2), & \text{if } r_k^{(z)}(m_1, m_2) > \delta_0 \\ z(n_1, n_2), & \text{if } -\delta_0 < r_k^{(z)}(m_1, m_2) < \delta_0 \\ z(n_1, n_2) + \frac{r_k^{(z)}(m_1, m_2) + \delta_0}{\sum_{l_1} \sum_{l_2} h^2(m_1, m_2; l_1, l_2)} h(m_1, m_2; n_1, n_2), & \text{if } r_k^{(z)}(m_1, m_2) < -\delta_0 \end{cases} \quad (6)$$

Applying Bayes theorem, (8) may be rewritten as

$$(\hat{\mathbf{z}}, \hat{\mathbf{s}}) = \arg \max_{\mathbf{z}, \mathbf{s}} \frac{P(\mathbf{y} | \mathbf{z}, \mathbf{s}) P(\mathbf{z}, \mathbf{s})}{P(\mathbf{y})} \quad (9)$$

Since $\mathbf{z}$ and $\mathbf{s}$ are statistically independent, $P(\mathbf{z}, \mathbf{s}) = P(\mathbf{z})P(\mathbf{s})$. Moreover, $P(\mathbf{y})$ is not a function of $\mathbf{z}$ or $\mathbf{s}$, so it may be omitted from the optimization process with respect to $\mathbf{z}$ and $\mathbf{s}$. Therefore, the estimates are given by

$$(\hat{\mathbf{z}}, \hat{\mathbf{s}}) = \arg \max_{\mathbf{z}, \mathbf{s}} P(\mathbf{y} | \mathbf{z}, \mathbf{s}) P(\mathbf{z}) P(\mathbf{s}) \quad (10)$$

Equivalently one aims to minimize the negative logarithm of the function in (10). Thus

$$(\hat{\mathbf{z}}, \hat{\mathbf{s}}) = \arg \min_{\mathbf{z}, \mathbf{s}} \{-\log[P(\mathbf{y} | \mathbf{z}, \mathbf{s})] - \log[P(\mathbf{z})] - \log[P(\mathbf{s})]\} \quad (11)$$

Now we should determine the prior image probability density function (pdf) $P(\mathbf{z})$, the prior motion probability density function $P(\mathbf{s})$, and the conditional probability density function $P(\mathbf{y} | \mathbf{z}, \mathbf{s})$.

The prior image pdf which preserves convexity, may be expressed as

$$P(\mathbf{z}) = \frac{1}{A} \exp \{-\lambda f(\mathbf{z})\} \quad (12)$$

where A is some normalizing constant which ensures that $P(\mathbf{z})$ is a probability, λ is some positive constant, and $f(\mathbf{z})$ is a function of the values of the HR image $\mathbf{z}$. This function may be chosen to encourage neighboring pixels to have similar values so that the first derivative of the HR image function is continuous (the so called “membrane model”), or so that the second derivative of the data is continuous (the so called “thin plate model”).

The prior model for the registration parameters $\mathbf{s}$ is highly application specific. In general, $P(\mathbf{s})$ may be dropped from the cost function, implying that we do not have any prior knowledge about the registration parameters.

The elements of $\mathbf{n}$ in (2) are assumed to be independent and identically distributed (iid) Gaussian samples with variance σ_n^2 . Therefore, the multivariate pdf of $\mathbf{n}$ is given by:

$$P(\mathbf{n}) = \frac{1}{(2\pi)^{\frac{PM}{2}} \sigma_n^{PM}} \exp \left\{ -\frac{1}{2\sigma_n^2} \sum_{k=1}^{PM} n_k^2 \right\} \quad (13)$$

Given the observation model in (2) and the noise pdf in (13), the conditional pdf $P(\mathbf{y} | \mathbf{z}, \mathbf{s})$ may be written as

$$P(\mathbf{y} | \mathbf{z}, \mathbf{s}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma_n^N} \exp \left\{ -\frac{1}{2\sigma_n^2} (\mathbf{y} - \mathbf{W}_s \mathbf{z})^T (\mathbf{y} - \mathbf{W}_s \mathbf{z}) \right\} \quad (14)$$

The aim is to minimize the cost function (11) with respect to $\mathbf{z}$ and $\mathbf{s}$. Note that (11) is not readily differentiable with

respect to $\mathbf{s}$ for many motion models. In addition, depending on the choice of $f(\mathbf{z})$, (11) usually is a quadratic function in $\mathbf{z}$ and can be minimized with respect to $\mathbf{z}$ if $\mathbf{s}$ is fixed. The authors in [6] apply a cyclic coordinate-descent optimization approach in which $\mathbf{z}$ is fixed and $\mathbf{s}$ estimated, and then $\mathbf{s}$ is fixed and $\mathbf{z}$ is estimated. The iterative algorithm terminates when it adequately converges or a pre-set number of iterations is reached. The initial estimate $\hat{\mathbf{z}}^1$ may be an interpolation of the first LR image.

The estimated LR images are compared with the HR estimate $\hat{\mathbf{z}}^n$ after it is projected through the observation model. The estimation of the registration parameters $\hat{\mathbf{s}}_k^n$ effectively results in the estimation of the weights $w_{km_1 m_2 n_1 n_2}^n$ for the n th iteration. Once we have updated all $\hat{\mathbf{s}}_k^n$, we aim in updating the HR image $\hat{\mathbf{z}}$.

IV. EXPERIMENTS

A simulated sequence of random translational shifts is used to generate a series of translated LR images (each of size 50×50 pixels) from the 16 bit grayscale image shown in Fig 1(a). The camera is assumed to move so that all LR frames it captures share the same image plane. In other words, no zooming, panning, or tilting of the camera motion is allowed. To simulate in practice such data, from each location we capture a HR frame which we blur with PSF

$$h = \begin{bmatrix} 0.05 & 0.1 & 0.05 \\ 0.1 & 0.4 & 0.05 \\ 0.05 & 0.1 & 0.05 \end{bmatrix}$$

and subsample by a factor of 4 in each direction. No noise is added to each LR frame. The first LR frame in the generated sequence is presented in Fig. 1(b).

The first LR frame is considered as the reference LR frame. All LR frames are bilinearly interpolated to the desired high resolution. The interpolated frames are now registered to the SR grid, which has been set up by the interpolated version of the LR reference frame, using normalized cross-correlation. Thus we can accurately identify the translational displacements.

The initial HR image estimate is the bilinearly interpolated version of the first LR frame as shown in Fig. 1(c). For each of the IBP, POCS, and MAP estimation algorithms, 20 iterations are performed. The estimated SR images are presented respectively in Fig. 1(d)-1(f).

Fig. 2(a) depicts the mean absolute error (MAE) per pixel for each method versus the number of iterations used to estimate the SR image. POCS and MAP estimation are shown to be consistently more effective in SR reconstruction than IBP. For comparison, the MAE's of the image formed by

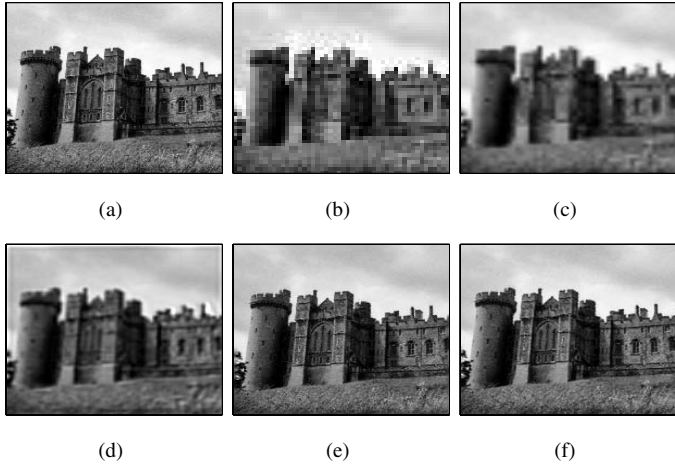


Fig. 1. (a) Original image. (b) Simulated low-resolution frame 1 ($L_1 = L_2 = 4$). (c) Bilinear interpolation of frame 1. (d) IBP: Super-resolution image. (e) POCS: Super-resolution image. (f) MAP: Super-resolution image.

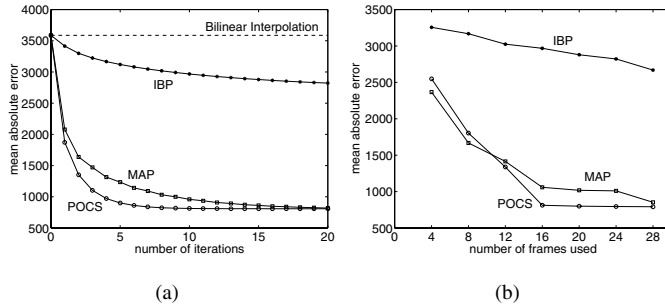


Fig. 2. (a) Mean absolute error per pixel vs number of iterations using 16 frames. (b) Mean absolute error per pixel vs number of frames used for 10 iterations.

bilinear interpolation of a single LR frame is also shown. Fig. 2(b) shows the MAE per pixel for all three methods using various numbers of LR frames. We can see that with only four frames, the performance is comparable with that of the bicubic interpolator. Increasing the number of the available LR frames results in improved restoration. Note that 16 frames is the minimum theoretical limit in order to solve the SR problem with our data. The improvement is dramatic until this limit is reached when using POCS and MAP estimation methods. Additional frames improve only slightly the SR image.

The computational cost of the algorithms is rather high for real time implementation. For a reconstruction of an image with 200×200 pixels from 16 frames of 50×50 pixels each, approximately 10 iterations are needed. The Matlab code used required 3.85s, 4.61s and 6.67s per iteration for the POCS, MAP, and IBP algorithms respectively.

Next we examine the tolerance to registration errors. We compute the HR image using each of the three methods for various levels of misregistration (see Fig. 3). Of the three methods compared, the IBP approach showed the most resilience to misregistration errors. POCS and MAP are quite sensitive to registration errors, with POCS being a little more

sensitive than MAP. However, both techniques are clearly superior compared with IBP when registration is successful.

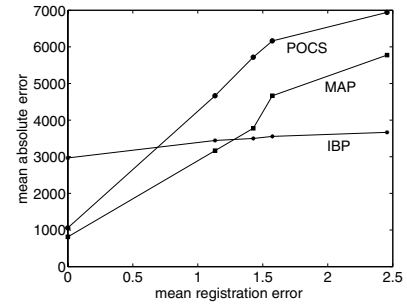


Fig. 3. Mean absolute error per pixel vs mean registration error in units of HR pixels.

In practice, POCS and MAP are comparable in terms of speed and performance. They result in sufficient reconstruction of the desired HR image with a small number of iterations.

ACKNOWLEDGMENT

This work has been carried out with the support of the UK MOD Data and Information Fusion Defence Technology Centre under project DIF DTC 9.3.

REFERENCES

- [1] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231–239, May 1991.
- [2] —, "Super resolution from image sequences," in *Proc. 10th International Conference on Pattern Recognition*, vol. 2, June 16–21 1990, pp. 115–120.
- [3] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'88)*, June 5–9 1988, pp. 742–746.
- [4] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion, and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, Dec. 1993.
- [5] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 201–212, Apr. 1984.
- [6] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.