Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images

2 authors:

Michael Elad
Technion – Israel Institute of Technology
418 PUBLICATIONS   **69,660** CITATIONS

SEE PROFILE

Arie Feuer
Technion – Israel Institute of Technology
155 PUBLICATIONS   **5,582** CITATIONS

SEE PROFILE

APPLE 1048

Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images

Michael Elad and Arie Feuer, *Senior Member, IEEE*

Abstract— The three main tools in the single image restoration theory are the maximum likelihood (ML) estimator, the maximum a posteriori probability (MAP) estimator, and the set theoretic approach using projection onto convex sets (POCS). This paper utilizes the above known tools to propose a unified methodology toward the more complicated problem of *superresolution restoration*. In the superresolution restoration problem, an improved resolution image is restored from several geometrically warped, blurred, noisy and *downsampled* measured images. The superresolution restoration problem is modeled and analyzed from the ML, the MAP, and POCS points of view, yielding a generalization of the known superresolution restoration methods. The proposed restoration approach is general but assumes explicit knowledge of the linear space- and time-variant blur, the (additive Gaussian) noise, the different measured resolutions, and the (smooth) motion characteristics. A hybrid method combining the simplicity of the ML and the incorporation of nonellipsoid constraints is presented, giving improved restoration performance, compared with the ML and the POCS approaches. The hybrid method is shown to converge to the unique optimal solution of a new definition of the optimization problem. Superresolution restoration from motionless measurements is also discussed. Simulations demonstrate the power of the proposed methodology.

Index Terms—Constrained optimization problems, estimation, image restoration, MAP, ML, POCS, regularization, supperresolution.

I. INTRODUCTION

THE CLASSIC theory of restoration of a single image from linear blur and additive noise has drawn a lot of research attention in the last three decades [1]–[4]. Many algorithms were proposed in the literature for this classic and related problems, contributing to the construction of a unified theory that ties together many of the existing methods [4]. In the single image restoration theory, three major and distinct approaches are extensively used in order to get practical restoration algorithms: 1) maximum likelihood (ML) estimator [1]–[4], 2) maximum *a posteriori* (MAP) probability estimator [1]–[4], and 3) projection onto convex sets (POCS) approach ([5]).

The consistent development of computer technology in recent years has led to a growing interest in image restoration theory. The main directions are nontraditional treatments to the classic problem and looking at new, second-generation restoration problems, allowing for more complicated and more computationally intensive algorithms. Among these new second-generation problems are multiple image restoration [6]–[12] and superresolution image restoration [13]–[27]. This paper focuses on the latter problem of superresolution restoration. Application of such restoration methods arises in the following areas.

1) *Remote sensing:* where several images of the same area are given, and an improved resolution image is sought.
2) *Frame freeze in video:* where typical single frame in video signal is generally of poor quality and is not suitable for hard-copy printout. Enhancement of a freeze image can be done by using several successive images merged together by a superresolution algorithm.
3) *Medical imaging (CT, MRI, ultrasound, etc.):* these enable the acquisition of several images, yet are limited in resolution quality.

The superresolution restoration idea was first presented by Tsay and Huang [13]. They used the frequency domain approach to demonstrate the ability to reconstruct one improved resolution image from several downsampled noise-free versions of it, based on the spatial aliasing effect. Other results suggested a simple generalization of the above idea to noisy and blurred images. A frequency domain recursive algorithm for the restoration of superresolution images from noisy and blurred measurements is suggested in [14]–[16]. A spatial domain alternative, based on Papoulis [17] and Yen [18] generalized sampling theorems is suggested by Ur and Gross [19]. Srinivas and Srinath [20] proposed a superresolution restoration algorithm based on a minimum mean squared error (MMSE) approach for the multiple image restoration problem and interpolation of the restored images into one. All the above superresolution restoration methods [13]–[20] are restricted to global uniform translational displacement between the measured images, linear space-invariant (LSI) blur, and homogeneous additive noise.

A different approach toward the superresolution restoration problem was suggested by Peleg *et al.* [21]–[23], based on the

M. Elad is with Hewlett-Packard Laboratories–Israel, The Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: elad@hp.technion.ac.il).

A. Feuer is with the Electrical Engineering Department, Technion–Israel Institute of Technology, Haifa 32000, Israel.

iterative backprojection (IBP) method adopted from computer aided tomography (CAT). This method starts with an initial guess of the output image, projects the temporary result to the measurements (simulating them), and updates the temporary guess according to this simulation error. This method is not limited as the previous ones to specific motion characteristics and allows arbitrary smooth motion flow, although the convergence of the proposed algorithm is proven only for an affine geometric warp between the measured images [23]. Set theoretic approach to the superresolution restoration problem was also suggested [24]–[26]. The main result there is the ability to define convex sets which represent tight constraints on the required image. Having such constraints, it is straightforward to apply the POCS method. The restoration procedure in [24]–[26] has the same benefits as the IBP method mentioned earlier: arbitrary smooth motion, linear space variant blur, and nonhomogeneous additive noise. Actually, the POCS can even be better than the IBP since nonlinear constraints can be easily combined with the restoration process. However, practical application of the projections might be computationally demanding, thus limiting the POCS applicability.

Another approach toward the super-resolution restoration problem is presented by Schultz and Stevenson [27]. Their approach uses MAP estimator, with the Huber-Markov Random Field (HMRF) prior. The blur of the measured images is assumed to be simple averaging, and the measurements additive noise is assumed to be independent and identically distributed (i.i.d.) Gaussian vector. This choice of prior causes the entire problem to be nonquadratic, thus complicating the resulting minimization problem.

In complete analogy to the single image restoration problem, this paper proposes a unified methodology toward the problem of *superresolution restoration*. In this problem, one improved resolution image is restored from several blurred, noisy, and downsampled measured images. The superresolution restoration problem is modeled by using sparse matrices and analyzed from the ML, the MAP, and the POCS points of view. The result is a direct generalization of the classic problem of single image restoration from one measured image. The three approaches merge into one family of algorithms, which generalizes the single image restoration theory [1]–[4] on one hand, and the existing superresolution algorithms proposed in the literature [13]–[23] on the other hand. The proposed restoration approach is general but assumes explicit knowledge of the linear space- and time-variant blur, the (additive Gaussian) noise, the different measured resolutions, and the (smooth) motion flow. The presented methodology also enables the incorporation of POCS into the ML or MAP restoration algorithms, similar to the way it is done for the iterative single image restoration problem [4], yielding hybrid superresolution restoration algorithm with further improved performance and assured convergence.

The classic superresolution restoration assumes a relative motion between the measured images as part of the model, in order to achieve superresolution restoration capability [13]–[27]. Another question addressed in this paper is whether motion is a necessary condition for a feasible restoration with improved resolution. Theoretical results for the LSI blur

operations can be found in [17] and [32]. In this paper, the general linear space variant case is treated, and it is shown that superresolution can be achieved even without motion.

This paper is organized as follows: Section II presents a new model for the superresolution problem and the application of the ML, the MAP, the POCS, and the hybrid algorithms for the restoration task. Section III presents a short analysis of the motionless superresolution restoration problem. In Section IV we compare the main known superresolution restoration techniques to the approach presented here, showing that the new approach is a generalization of those techniques. Simulations results presented in Section V and Section VI concludes the paper.

II. SUPERRESOLUTION RESTORATION: A NEW APPROACH

In this section, we present a new approach toward the superresolution restoration problem. Simplicity and direct connection to the problem of single image restoration (from one measured image) are the main benefits of this approach. Thus, the various known methods to restore one image from one measured image are easily generalized to the new problem of single image restoration from several measured images. We start our presentation with a new model to the problem and then turn to apply known restoration methods to the suggested model.

A. Modeling the Problem

The key to a comprehensive analysis of the classical superresolution problem is to formulate the problem and to model it as simply and as efficiently as possible. We start by presenting the problem to be solved and then turn to introduce an analytical model describing it. Throughout this paper we represent images columnwise lexicographically ordered for matrix notation convenience.

Given are N measured images $\{\underline{Y}_k\}_{k=1}^N$, where each image is (in the general case) of different size $[M_k \times M_k]$. We assume that these images are different representations of a single high-resolution image $\underline{X}$ of size $[L \times L]$, where typically $L > M_k$ for $1 \leq k \leq N$. More specifically, each measured image is the result of an arbitrary geometric warping, linear space-variant blurring, and uniform rational decimating performed on the ideal high-resolution image $\underline{X}$. We further assume that each of the measured images is contaminated by nonhomogeneous additive Gaussian noise, uncorrelated between different measurements. In order to treat the most general case, it is assumed that each measurement is the result of different blur, noise, motion, and decimation parameters. Translating the above description to an analytical model, we get

$$\underline{Y}_k = D_k C_k F_k \underline{X} + \underline{E}_k \quad \text{for} \quad 1 \leq k \leq N \quad (2.1)$$

where F_k is a $[L^2 \times L^2]$ matrix representing the geometric warp performed on the image $\underline{X}$, C_k is the linear space-variant blur matrix of size $[L^2 \times L^2]$, D_k is a $[M_k^2 \times L^2]$ matrix representing the decimation operator resulting in $\underline{Y}_k$. $\underline{E}_k$ stands for the additive zero mean Gaussian noise in the kth measurement with positive definite autocorrelation matrix
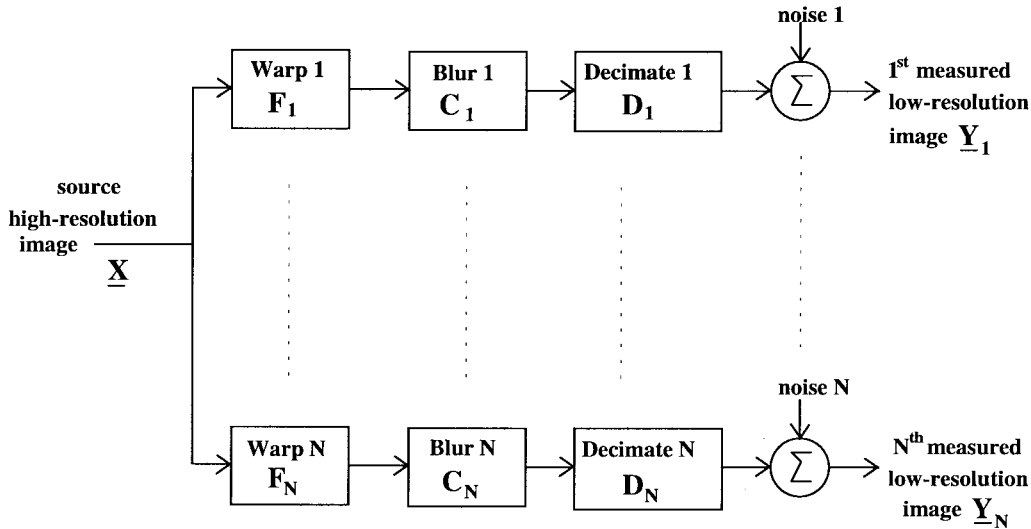
Fig. 1. Degradation model for the superresolution restoration problem

W_k^{-1} of size $[M_k^2 \times M_k^2]$. All these matrices (F_k, C_k, D_k, W_k) are assumed to be known in advance. Fig. 1 illustrates the model schematically.

Before we turn to use the above model, we justify the above assumption regarding the availability of the matrices involved. The geometric warp matrix F_k is a one-to-one representation of the optic flow between the *nondecimated noiseless* version of the kth measured image and the ideal image $\underline{X}$. This optic flow can be reliably estimated from the noisy and downsampled measurement if it is smooth enough. Such a case corresponds to global motion characteristics. Note that the orientation of the ideal image should be chosen arbitrarily as equal to the orientation of one of the measurements.

The assumption on the *a priori* knowledge of the blurring matrix can be explained in some applications by referring the blur to measurable phenomena, such as optics and sensor blur. In other cases, we may assume that the superresolution restoration process is robust to errors in the blurring function, and thus use rough guess for the blur function and allow such errors to exist. Of course, such robustness should be proven, a point that will not be treated in this paper. A different approach which will not be discussed in this framework is estimating the blur function in parallel to the restoration process, as suggested for single image restoration algorithms [4].

The decimation ratio between the ideal image and the kth measurement image is the only parameter determining the matrix D_k. This ratio is directly drawn from the ratio between the number of pixels in the measured image M_k^2 and the ideal image L^2. The determination of the ideal image pixels number is arbitrary. Choosing a very high value will cause the problem to be ill-posed, with the ability to relax it by regularization. Choosing a low value can result in underutilization of information in the measurements, but will improve the noise suppression results. One intuitive rule that can help us determine L is the requirement $L^2 < M_1^2 + \cdots + M_k^2 + \cdots + M_N^2$, which can be explained as a requirement that the amount of given data (in the measurements) should be larger than the amount of information required in the restored image.

The autocorrelation matrix W_k^{-1} can be chosen as the identity matrix if no *a priori* knowledge on the additive noise is given. Such a choice corresponds to the assumption that the noise is white, which is typically the case for many restoration problems, including superresolution applications. As mentioned before, a colored noise can be assumed with its parameters estimated as part of the restoration process [4], though this approach will not be further discussed here.

Having the above model, grouping of the N equations into one can be done for notational convenience. This way we get

$$\begin{bmatrix} \underline{Y}_1 \\ \vdots \\ \underline{Y}_N \end{bmatrix} = \begin{bmatrix} D_1 C_1 F_1 \\ \vdots \\ D_N C_N F_N \end{bmatrix} \underline{X} + \begin{bmatrix} \underline{E}_1 \\ \vdots \\ \underline{E}_N \end{bmatrix} = \begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix} \underline{X} + \underline{\mathbf{E}}$$
$$\Leftrightarrow \underline{\mathbf{Y}} = \mathbf{H}\underline{X} + \underline{\mathbf{E}} \tag{2.2}$$

where we have defined $H_k = D_k C_k F_k$, and the autocorrelation of the Gaussian random vector $\underline{\mathbf{E}}$ is

$$E\left\{\underline{\mathbf{E}}\,\underline{\mathbf{E}}^T\right\} = \begin{bmatrix} W_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & W_N \end{bmatrix}^{-1} = \mathbf{W}^{-1}. \tag{2.3}$$

The obtained model equation $\underline{\mathbf{Y}} = \mathbf{H}\underline{X} + \underline{\mathbf{E}}$ is a classic restoration problem model [1]–[4]. Thus, we can easily apply the ML estimator, the MAP or the POCS methods in order to restore the image $\underline{X}$, which is exactly our purpose here. In the following sections we shall briefly present the way to apply each of those tools.

B. ML Restoration

According to the ML estimator [1]–[4], the estimation of the unknown image $\underline{X}$ is done by maximizing the conditional probability density function of the measurements, given the ideal image $P\{\underline{\mathbf{Y}}/\underline{X}\}$. Assuming that the measurements

additive noise is zero mean Gaussian random process with autocorrelation matrix $\mathbf{W}^{-1}$, performing several algebraic steps we get that the ML is reduced to the weighted least squares (WLS) estimation of the form

$$\hat{\underline{X}}_{ML} = \underset{\underline{X}}{\text{argmax}} \, P\{\underline{\mathbf{Y}}/\underline{X}\}$$

$$= \underset{\underline{X}}{\text{argmax}} \left\{ [\underline{\mathbf{Y}} - \mathbf{H}\underline{X}]^T \mathbf{W} [\underline{\mathbf{Y}} - \mathbf{H}\underline{X}] \right\} \quad (2.4)$$

Differentiating with respect to $\underline{X}$ and equating to zero gives the well-known classic pseudoinverse result

$$\mathbf{R}\hat{\underline{X}}_{ML} = \underline{\mathbf{P}} \quad (2.5)$$

where:

$$\mathbf{R} = \mathbf{H}^T \mathbf{W} \mathbf{H} = \sum_{k=1}^{N} H_k^T W_k H_k$$

$$\underline{\mathbf{P}} = \mathbf{H}^T \mathbf{W} \underline{\mathbf{Y}} = \sum_{k=1}^{N} H_k^T W_k \underline{Y}_k. \quad (2.6)$$

Locally adaptive regularization can be included in the above analysis with both algebraic and physical interpretations [1]–[4]. Using the Laplacian operator S and a weighting matrix V (penalizing nonsmoothness according to the *a priori* knowledge on the smoothness required at each pixel), we get

$$\hat{\underline{X}}_{ML} = \underset{\underline{X}}{\text{argmax}} \left\{ [\underline{\mathbf{Y}} - \mathbf{H}\underline{X}]^T \mathbf{W} [\underline{\mathbf{Y}} - \mathbf{H}\underline{X}] \right.$$
$$\left. + \beta [S\underline{X}]^T V [S\underline{X}] \right\} \quad (2.7)$$

Differentiating again with respect to $\underline{X}$ and equating to zero yields the equation $\mathbf{R}\hat{\underline{X}}_{ML} = \underline{\mathbf{P}}$, which is the same as in (2.5), but a new term, $\beta S^T V S$, is added to the matrix $\mathbf{R}$. Various iterative methods for practical ways to solve this large set of sparse linear equations have been suggested in the literature (see, e.g., [29]).

C. MAP Restoration

According to the MAP estimator, the additive noise, the measurements, and the ideal image are all assumed stochastic signals. The MAP estimation of the unknown image $\underline{X}$ is done by maximizing the conditional probability density function of the ideal image given the measurements $P\{\underline{X}/\underline{\mathbf{Y}}\}$. Based on Bayes rule, maximizing $P\{\underline{X}/\underline{\mathbf{Y}}\}$ is equivalent to maximizing the function $P\{\underline{\mathbf{Y}}/\underline{X}\}P\{\underline{X}\}$ [1]–[4]. Therefore, the MAP estimator is equivalent to the ML estimator, with the uniform probability distribution assumption of $\underline{X}$.

If we assume that the measurements additive noise is zero mean Gaussian random process with auto-correlation matrix $\mathbf{W}^{-1}$, and $\underline{X}$ is a zero mean Gaussian random process also, with autocorrelation matrix Q, the MAP estimator becomes the MMSE estimator. Performing several algebraic steps as

was done before, the MAP estimation gives

$$\hat{\underline{X}}_{MAP} = \underset{\underline{X}}{\text{argmax}} \, \{P\{\underline{\mathbf{Y}}/\underline{X}\}P\{\underline{X}\}\}$$

$$= \underset{\underline{X}}{\text{argmax}} \left\{ [\underline{\mathbf{Y}} - \mathbf{H}\underline{X}]^T \mathbf{W} [\underline{\mathbf{Y}} - \mathbf{H}\underline{X}] \right.$$
$$\left. + \underline{X}^T Q^{-1} \underline{X} \right\}. \quad (2.8)$$

Minimizing the above function with respect to $\underline{X}$ yields the following result:

$$\mathbf{R}\hat{\underline{X}}_{MAP} = \underline{\mathbf{P}} \quad (2.9)$$

where

$$\mathbf{R} = Q^{-1} + \mathbf{H}^T \mathbf{W} \mathbf{H} = Q^{-1} + \sum_{k=1}^{N} H_k^T W_k H_k$$

$$\underline{\mathbf{P}} = \mathbf{H}^T \mathbf{W} \underline{\mathbf{Y}} = \sum_{k=1}^{N} H_k^T W_k \underline{Y}_k \quad (2.10)$$

and the resemblance to the ML result is evident. It can be shown [4] that if an autoregressive (AR) model is assumed on the image $\underline{X}$, a simple and direct connection between the Laplacian regularization matrix and the AR coefficients can be established. As before for the ML, the MAP estimator reduces to a large set of sparse equations that can be solved iteratively [29].

D. Set Theoretic Restoration

According to the set theoretic approach [5], each *a priori* knowledge on the required restored image should be formulated as a constraining convex set containing the restored image as a point within this set. Using the model presented earlier, we can suggest a group of such convex sets based on the ℓ_2 distance measure

$$G_k = \left\{ \underline{X} \middle| \|D_k C_k F_k \underline{X} - \underline{Y}_k\|_{W_k}^2 \leq 1 \right\} \quad 1 \leq k \leq N. \quad (2.11)$$

This defines a group of N convex sets—ellipsoids, in this case. If the measurements additive noise is white, then $W_k = \sigma_k^{-2} \cdot I$, other forms of constraints can be proposed, based on ℓ_∞ distance measure [5], [24], [25]

$$G_k(m,n) = \left\{ \underline{X} \middle| \left| [D_k C_k F_k \underline{X}]_{(m,n)} - y_k(m,n) \right| \leq \delta_k(m,n) \right\}$$
$$1 \leq k \leq N; \forall (m,n) \in \theta_k \quad (2.12)$$

where θ_k is the support region of the kth measured image, and δ_k stands for the uncertainty of the model [24], [25]. This convex set is actually a polytop constructed of M_k^2 scalar constraints—each corresponds to one pixel of the measured image $\underline{Y}_k$, requiring that the absolute value of the model error at this point be bounded.

Another set that can be used is the one constraining smoothness. According to the proposed smoothness constraint [4], we can suggest convex set versions as before.

For ℓ_2, we have

$$G_S = \left\{ \underline{X} \middle| \|S\underline{X}\|_V^2 \leq 1 \right\} \quad (2.13)$$

or for ℓ_∞:

$$G_S(m,n) = \left\{ \underline{X} \middle| \left| [S\underline{X}]_{(m,n)} \right| \le \delta_0 \right\} \quad \forall(m,n) \in \theta_0 \quad (2.14)$$

where θ_0 is the support region of the ideal image. We can incorporate additional nonlinear constraints such as constraints on the output energy, phase, support, and others. An often used constraint is the one posed on the amplitude of the result

$$G_A = \{x(m,n) | A_1 \le x(m,n) \le A_2\} \quad \forall(m,n) \in \theta_0 \tag{2.15}$$

Having a group of $\mathcal{M}$ convex sets, each containing the required image, the POCS method suggests the following iterative algorithm for the recovery of a point within the intersection of these sets [5]:

$$\underline{X}_{k+1} = P_{\mathcal{M}} P_{\mathcal{M}-1} \cdots P_2 P_1 \{\underline{X}_k\} \tag{2.16}$$

where P_j is the projection of a given point onto the jth convex set. Relaxed projections can be used instead of direct ones in order to improve convergence rate [5]. Projecting onto an ellipsoid, as required for the sets given in (2.11) and (2.13), is a computationally complicated task. Projection onto a multidimensional cube, as is given in the convex sets in (2.12) and (2.14), does not require any parameter setup, or large matrix inversion, and thus is much simpler to apply, compared to the projection onto an ellipsoid.

A different approach toward the POCS idea is the bounding ellipsoid method [6], [30], [33]. This method is valid for the case where all the constraints are ellipsoids. The basic idea here is to find the ellipsoid bounding the intersection of all the participating constraints, and to choose its center as the output result. In [6], [30], and [33], the bounding ellipsoid method and its properties are discussed. Using the convex sets given in (2.11) and (2.13), we get that the bounding ellipsoid center is [6], [30], [33]

$$\mathbf{R}\underline{\hat{X}}_{BE} = \underline{\mathbf{P}} \tag{2.17}$$

where

$$\mathbf{R} = \rho_0 S^T V S + \sum_{k=1}^{N} \rho_k H_k^T W_k H_k$$

$$\underline{\mathbf{P}} = \sum_{k=1}^{N} \rho_k H_k^T W_k \underline{Y}_k$$

and $\{\rho_k\}_{k=0}^{N}$ must satisfy

$$\forall 0 \le k \le N \quad \rho_k \ge 0 \quad \text{and} \quad \sum_{k=1}^{N} \rho_k \le 1. \tag{2.18}$$

For each different choice of $\{\rho_k\}_{k=0}^{N}$ a different bounding ellipsoid is obtained, with a different center. For the special case where all ρ_k are identical, we get the results obtained by the ML and the MAP methods.

E. Hybrid Restoration

We have seen that the ML, the MAP, and the uniform-weights bounding ellipsoid estimators give similar linear set of equations, to be solved using iterative algorithms [29]. The benefit in using these approaches is the relative simplicity of the restoration process. Their main drawback is the fact that additional nonquadratic constraints representing additional *a priori* knowledge of the ideal signal are not incorporated into the restoration process. This section presents the way to combine the simplicity of the above iterative algorithms with the application of the nonquadratic constraints. We start by defining a new convex optimization problem, which combines a quadratic scalar error with M convex constraints as follows:

$$\text{minimize } e^2 = \left\{ [\underline{Y} - \mathbf{H}\underline{X}]^T \mathbf{W} [\underline{Y} - \mathbf{H}\underline{X}] + \beta [S\underline{X}]^* V [S\underline{X}] \right\}$$

$$\text{subject to } \{\underline{X} \in \Im_k \quad 1 \le k \le M\} \tag{2.19}$$

where the quadratic error takes care of the model and the smoothness errors, and the M additional constraints refer to the nonellipsoids *a priori* knowledge. The quadratic error term is the same as was defined in the ML, the MAP, and the bounding ellipsoid methods. Our aim is to construct an efficient iterative algorithm to solve this constrained convex optimization problem. The benefit of the proposed new formulation is in the fact that it combines both set theoretic and stochastic estimation approaches. This way, all the *a priori* knowledge is utilized effectively and, in contrast to the POCS method, there is a single optimal solution (see appendix A) and [4], [33].

Following the iterative methods presented in [4], we propose a simple yet effective two-phase iterative algorithm to solve the above optimization problem. Analysis of this method can be found in Appendix A. Suppose that an efficient iterative algorithm that is known to converge to the minimum of the scalar squared error e^2 is given, denoted by I_t [29]. Algorithms such as the conjugate gradient (CG) or the Gauss–Siedel can be considered as excellent candidates for I_t. Beyond this first iterative algorithm I_t, $\mathcal{M}$ projection operators denoted by J_t^k $k = 1, 2, \cdots, \mathcal{M}$ can be constructed, each projects onto a convex set and represents a single given constraint. Assuming that the $\mathcal{M}$ projections are all given using the Euclidiean metric, we suggest the following global iterative step:

$$\underline{X}_{k+1} = J_t^{\mathcal{M}} \{ J_t^{\mathcal{M}-1} \{ \cdots \cdots J_t^1 \{ I_t \{\underline{X}_k\} \} \} \} \tag{2.20}$$

This interlaced approach is generally converging to a *suboptimal* point of the problem given in (2.19) (see Appendix A). Adding several new iterations, where now I_t is replaced by the (notoriously slow [29]) *steepest descent*, updates the previous result and assures that the final convergence is to the optimal point, as is proved in Appendix A. Off course, such convergence is assured only if the iterative algorithms I_t in both phases (conjugate-gradient/Gauss–Siedel, and steepest descent) are converging to the minimum of e^2 [29].

Appendix A also shows that the Gauss–Siedel or the CG algorithms [29] can serve for the first phase, because of their relatively fast convergence, but indeed, might converge to a suboptimal result, whereas the steepest descent assures convergence to the optimal result. It should be noted that

applying the CG algorithm combined with projection operations might cause instability behavior, because the projections might destroy the $\mathbf{R}$-orthogonality of the search directions, generated by the CG algorithm. However, our simulations gave no evidence for such problem.

III. MOTION-FREE SUPERRESOLUTION

The classic superresolution restoration assumes a relative motion between the measured images as part of the model, in order to achieve superresolution restoration capability [13]–[27]. This section refers to the question of whether the motion is a necessary condition for a feasible restoration with improved resolution. An example for such application is a fixed camera filming fixed objects. The question is whether superresolution image can be obtained from several such images with different defocusing. This section contains a short analysis of the above question. More details can be found in [33].

A. Theoretic Analysis

As was shown in the previous section, the ML, MAP, and POCS (only for ellipsoids and ℓ_2 norm) methods give similar restoration procedures, namely the solution of a large set of sparse linear equations $\mathbf{R}\hat{X} = \mathbf{P}$. We say that superresolution restoration is possible if the linear set of equations is well posed, which means that the matrix $\mathbf{R}$ is nonsingular. But, when regularization is combined as an interpolator, the matrix $\mathbf{R}$ is always nonsingular, even if the other terms in $\mathbf{R}$ are singular. Therefore, we omit the regularization term for the following analysis to concentrate on the restoration part without the need for interpolation.

Another way of looking at the equations set is the following: The spatial aliasing effect in the frequency domain is intimately connected to the superresolution restoration idea. This connection can be directly seen in frequency domain methods such as [13]–[16]. Spatial aliasing means that high-frequency components are folded and added to lower frequency ones. By getting several measurements with different spatial aliasing effect, these high frequencies can be identified and restored; this is the main idea behind the superresolution restoration methods. Intuitively, we can understand that having a measured image with no aliasing effect means that higher frequencies components do not exist in the original image, thus removing need for restoring them, and in such cases, the obtained restoration equation becomes ill posed. Taking this reasoning further, we can say that when the defined superresolution restoration problem reduces to an ill-posed equations set, this corresponds directly to an overresolution demand from the restoration process. Regularizing the ill-posed problem is equivalent to introducing interpolation since the data available is insufficient.

We are interested in the case where $F_k = I, \forall k$, which means that there is no relative geometric motion between the measured images. The question to be solved is: Can the matrix $\mathbf{R}$ under the above assumption on F_k be nonsingular? We further simplify the analysis by assuming that

1) all the measured images are of the same size M^2, i.e., $M_1^2 = M_2^2 = \cdots = M_N^2 = M^2$;

2) the decimation is the same for all the measured images: $D_k = D, \forall k$;
3) the weight matrices are the same for all measurements and equals I: $W_k = W = I, \forall k$;
4) the blur matrices are block-Toeplitz, which corresponds to LSI blur;
5) the blur matrices' kernel has zero center of mass in the two axes. This assumption is crucial since nonzero center of mass implies a global translational motion of the corresponding measured image. We could replace this requirement by a tighter requirement and restrict the analysis to the case of symmetric blurring kernel (in the two axes), but this of course only limits the results of the analysis.

Having all these assumptions, we have the following matrix to check for singularity:

$$\mathbf{R} = \sum_{k=1}^{N} C_k^T D^T D C_k = \begin{bmatrix} DC_1 \\ \vdots \\ DC_N \end{bmatrix}^T \begin{bmatrix} DC_1 \\ \vdots \\ DC_N \end{bmatrix} = \mathbf{H}^T\mathbf{H}. \tag{3.1}$$

It is easy to see that if the matrix $\mathbf{H}$ is of full rank, then $\mathbf{R}$ is nonsingular [31]. One immediate necessary condition for this requirement comes from the dimensions on the matrix $\mathbf{H}$: the number of its rows, NM^2, must be at least as high as the number of its columns, L^2. This necessary condition is therefore

$$L^2 \leq NM^2 \Rightarrow \left[L/M \right]^2 \leq N \tag{3.2}$$

The rank of the term DC_k is M^2 at the most (since $L > M$). Taking the first row of each such term DC_k, we get N vectors. These vectors must span at least an $(L/M)^2$ dimensional subspace in order to enable $\mathbf{H}$ to be full rank. Since these rows consist of only $[2p+1]^2$ nonzero entries at fixed locations, and since the center of mass for each kernel must be zero (see the assumption 5 made above) we get a second necessary condition for $\mathbf{H}$ to be full rank, as follows:

$$\left[L/M \right]^2 \leq [2p+1]^2 - 2 \tag{3.3}$$

where $(2p+1) \times (2p+1)$ is the kernel size of the blurring operators C_k [1]–[3]. From the above two necessary (but not sufficient) conditions, we have that superresolution restoration is impossible if the following condition is not met:

$$\left[L/M \right]^2 \leq \min \left\{ [2p+1]^2 - 2, N \right\} \tag{3.4}$$

which poses a restriction over the resolution improvement, governed by the blur-kernel size and the number of measurements. Note, however, that the above inequality only presents necessary conditions for the superresolution restoration possibility. One interesting point with regard to the above result is that increasing the number of measurements cannot increase the restored resolution beyond the upper bound posed by the kernel size. However, increased number of measurements with the same output resolution means that better reduction of noise can be achieved.

$$c_1 = \frac{1}{19}\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 3 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \; ; \; c_2 = \frac{1}{16}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 2 & 4 & 2 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \; ; \; c_3 = \frac{1}{14}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \; ; \; c_4 = \frac{1}{18}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$c_5 = \frac{1}{18}\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 2 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \; ; \; c_6 = \frac{1}{25}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \; ; \; c_7 = \frac{1}{28}\begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 4 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} \; ; \; c_8 = \frac{1}{26}\begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Fig. 2. The eight blurring kernels for the test.

TABLE I
RESULTS FOR THE TWO SUPERRESOLUTION RESTORATION TESTS

N	Cond. # of $\mathbf{R}$ kernels $[5 \times 5]$	Rank of $\mathbf{R}$ kernels $[5 \times 5]$	Cond. # of $\mathbf{R}$ kernels $[3 \times 3]$	Rank of $\mathbf{R}$ kernels $[3 \times 3]$
1	(∞)	64	(∞)	64
2	(∞)	128	(∞)	128
3	(∞)	192	(∞)	192
4	$2.64 \cdot 10^8$	256 (full)	(∞)	192
5	$2.61 \cdot 10^6$	256 (full)	(∞)	192
6	$7.62 \cdot 10^5$	256 (full)	(∞)	192
7	$7.85 \cdot 10^5$	256 (full)	(∞)	192
8	$7.83 \cdot 10^5$	256 (full)	(∞)	192

B. Motionless Superresolution: Simulation Results

To demonstrate that superresolution is possible without motion, we have chosen the following case. Let $L = 16$, $M = 8$, and $p = 2$. We then need to have $N \geq 4$ in order to satisfy the necessary condition (3.4). By assuming a symmetric blur kernel, we have only $p^2 + p + 1 = 7$ free parameters (see [33] for more details). Eight different experiments have been conducted varying in the number of measurement images used, from one to eight. Each of the available measurements resulted from using different blur kernel; these are given in Fig. 2. The C_k, $k = 1, 2, \cdots, 8$ are constructed each from a corresponding kernel in a block Toeplitz form. The matrix $\mathbf{R}$ was constructed for each test and its rank calculated. The results are summarized in Table I. We note that for our choice of blur kernels $\mathbf{R}$ is nonsingular when $N \geq 4$. Namely, in this case, motionless superresolution restoration is possible.

As part of this same experiment, eight measurements images from an ideal image $\underline{X}$, blurred by C_k, decimated by D and contaminated by white Gaussian noise with $\sigma = 1$ were created. The ideal image was then restored using four, five, six, seven, and eight measurements, with and without regularization. The regularization constant was chosen to be $\beta = 0.3$. The weight matrix for the regularization term was diagonal matrix, with "1" for all the image plane, except for a rectangle covering the letter A, which got the weight 0.1.

The restored images without regularization were computed using the CG algorithm using only five iterations. It was found

that using more iterations could cause unstable results. It is known that limiting the iteration number stands for a variation of regularization [4]. Thus, the nonregularized restoration results are in a way also regularized. The restoration results with smoothness regularization were computed using 20 iterations. The restored images are presented in Fig. 3. As can be seen from the images, raising the number of measurements indeed improves the restoration result as expected. Regularizing using the Laplacian operator also improves the results significantly, especially for a low number of measurements.

This experiment was repeated using different blur kernels, all $[3 \times 3]$, created from those in Fig. 3 by eliminating the outer elements. In this case, since the number of free parameters is $p^2 + p + 1 = 3$, requiring $L/M = 2$ is theoretically impossible according to (3.4). Indeed, Table I shows that in this case the rank of the matrices $\mathbf{R}$ is bounded by 192, meaning that $\mathbf{R}$ is singular for all N, as expected.

Summarizing this section, we have shown that superresolution restoration is possible even without motion between the measurements. We have established two necessary conditions that apply to the case of LSI blur and uniform size measurements.

IV. RELATION TO OTHER METHODS

Since this paper proposes a new approach to the superresolution restoration problem, it is appropriate to relate this new approach to the methods already known in the literature. In the sequel, we will present a brief description of each of the existing methods in light of the new results. The four main known methods for superresolution restoration are the IBP method [21]–[23], the frequency domain approach [14]–[16], the POCS approach [24]–[25], and the MAP approach [27]. This section will concentrate on these four methods.

A. The IBP Method

The IBP method [21]–[23] is an iterative algorithm that projects the temporary result onto the measurements, simulating them this way. The above simulation error is used to update the temporary result. If we take this exact reasoning and apply it on our proposed model in (2.1), denoting the
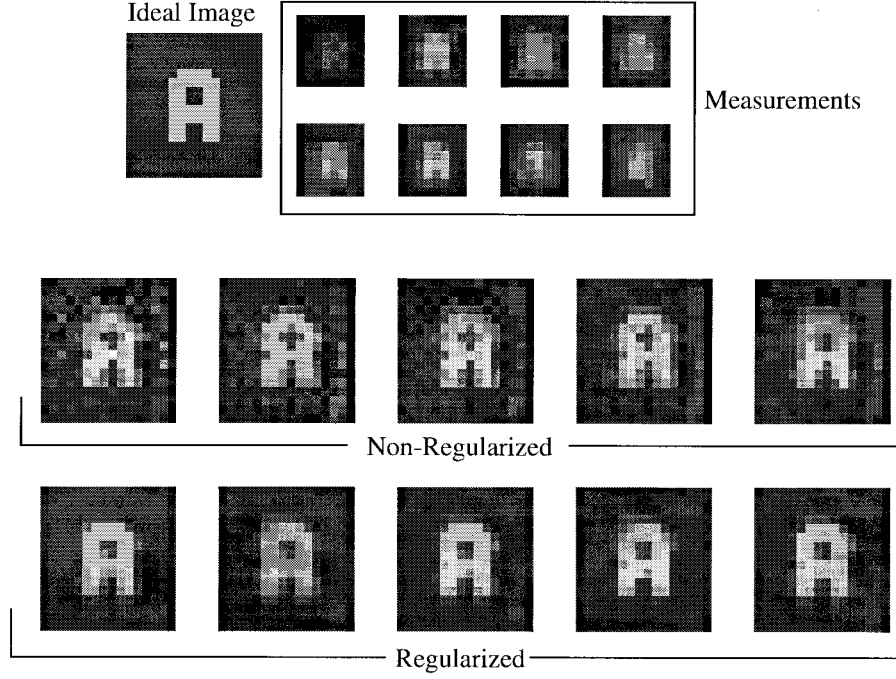
Fig. 3. Motionless superresolution restoration demonstration. $N = 4$ (left) to $N = 8$ (right).

temporary result at the mth step by $\hat{\underline{X}}_m$, we get for the simulated measurements $\hat{\underline{Y}}_{k,m} = D_k C_k F_k \hat{\underline{X}}_m$. The proposed update equation in the IBP method [21]–[23] is given in scalar form, but when put in matrix notations, we get

$$\hat{\underline{X}}_{m+1} = \hat{\underline{X}}_m + \sum_{k=1}^{N} Q_k \left[\underline{Y}_k - \hat{\underline{Y}}_{k,m} \right]$$

$$= \hat{\underline{X}}_m + \sum_{k=1}^{N} Q_k \left[\underline{Y}_k - D_k C_k F_k \hat{\underline{X}}_m \right] \quad (4.1)$$

where Q_k are some error relaxation matrices to be chosen. The configuration obtained in (4.1) is a simple error relaxation algorithm (such as the steepest descent, the Gauss–Siedel algorithms, or other algorithms), which minimizes a quadratic error as defined in (2.4). This analogy means that the IBP method is none other than the ML (or least squares) method proposed here without regularization. In the IBP method presented in [21]–[23], the matrices Q_k were chosen to be $Q_k = F_k^{-1} \tilde{C}_k \tilde{D}_k$ where $\tilde{C}_k$ is a reblurring operator, and $\tilde{D}_k$ is an interpolation to be determined [21]–[23]. If we choose the simple SD algorithm for the solution of (2.5), we get that $Q_k = F_k^T C_k^T D_k^T$. This result implies that choosing the transpose of the blur matrix as the reblurring operator, and zero padding as the interpolation operator gives almost the same result as the IBP method. The only difference is the choice of the warp matrix F_k in the above two configurations. Since $F_k^{-1} = \left(F_k^T F_k \right)^{-1} F_k^T$, the IBP method uses the additional positive-definite inverse of the matrices $\left(F_k^T F_k \right)$ to the error relaxation matrices proposed by the SD algorithm. These additional terms may compromise the convergence properties of the IBP algorithm, whereas the SD (and others) approach performed directly on the ML optimization problem assures convergence.

According to the above discussion, therefore, the new approach has thus several benefits when compared to the IBP method, as follows.

1) There is a freedom to choose faster iterative algorithms (such as the CG) to the quadratic optimization problem.
2) Convergence is assured for arbitrary motion characteristic, linear space variant blur, different decimation factors for the measurements, and different additive noise statistics.
3) Locally adaptive regularization can be added in a simple fashion, with improved overall performance.

B. The Frequency Domain Method

A frequency domain analysis is possible only for an LSI case, where the blur, the motion, and the decimation are all space invariant. In order for the motion to be LSI, only global translational motion is allowed. It can be shown [31]–[33] that in such cases, the matrix $\mathbf{R}$ is unitarily similar to a block diagonal matrix using the 2-D discrete Fourier transform (DFT) matrix. Actually, if we take the previous section discussion and assume that the blur kernels are with nonzero center of mass, we get a treatment for the general LSI superresolution case. As is shown in [33], since the matrix to be inverted as part of the restoration procedure is block diagonal matrix, small groups of pixels in the output image (in the frequency domain) can be calculated independently. Thus, the overall restoration algorithm is separable and can be implemented most efficiently in a parallel scheme.

The frequency domain algorithm proposed in [14]–[16] gives the same result as the one discussed here. In addition, the recursive least squares (RLS) algorithm is proposed there in order to add new measurements to the process, using the previous restoration result. The frequency domain method as

9

proposed in [14]–[16] is again the ML approach presented here for the LSI case. Thus, its result is the same as the one that is obtained by the IBP method. As mentioned before, regularization can be added in order to improve the restoration results, a tool that is not proposed in [14]–[16]. However, since a regularization weight matrix that assigns locally adaptive smoothness weight to each pixel is linear space *variant*, locally adaptive regularization is not possible in the frequency domain approach. Instead, we can suggest a different scheme where the superresolution restoration is performed twice using LSI formulation—once without regularization, yielding $\underline{X}_{out}^1$, and once with LSI regularization yielding $\underline{X}_{out}^2$. The locally adaptive restoration result can then be constructed by the equation

$$\underline{X}_{out} = [I - V] \cdot \underline{X}_{out}^1 + V \cdot \underline{X}_{out}^2 \qquad (4.2)$$

where V is the regularization weight matrix. This way, non-smooth regions will be chosen from the nonregularized restoration process, and smooth regions will be chosen from the regularized image.

C. The POCS Method

The approach taken in [24]–[26] is the direct application of the POCS method for the restoration of superresolution image. The suggested approach did not use the smoothness constraint as proposed here, and chose to use the ℓ_∞ distance measure in order to get simpler projection operators. In the sequel, we have presented the bounding ellipsoid method as a tool to relate the POCS results to the stochastic estimation methods. We have seen that applying only ellipsoids as constraints gives a very similar result to the ML and the MAP methods [33]. In [24]–[26], it is suggested to add only the amplitude constraint given in (2.15) to the trivial ellipsoid constraints. We have shown that instead, we can suggest a hybrid method that has a unique solution, and yet is very simple to implement.

D. The Map Method with Huber–Markov Prior

The MAP approach with the Huber–Markov prior was suggested by Schultz and Stevenson [27]. Their approach starts with a linear model describing the relationship between the measurements and the required higher resolution image. This model is very similar to ours, given in (2.2). However, they restrict their treatment to simple uniform blur, and the measurements noise is assumed to be i.i.d. Gaussian vector, with variance which linearly decays as a function of the image index, related to the center index. This property gives higher influence to near images, and low influence to distant ones.

As we have seen above, the MAP estimator suggests some sort of regularization, originated from stochastic modeling. In the Huber–Markov prior, this regularization is a Gibbs prior that penalizes high activity regions. No attempt is made to adopt this penalty to be locally varying, according to the image content. The Huber–Markov prior is simply a quadratic function for low activity values, and linear for higher values. As such, the overall resulting minimization problem becomes nonquadratic, and is typically more complicated to solve.
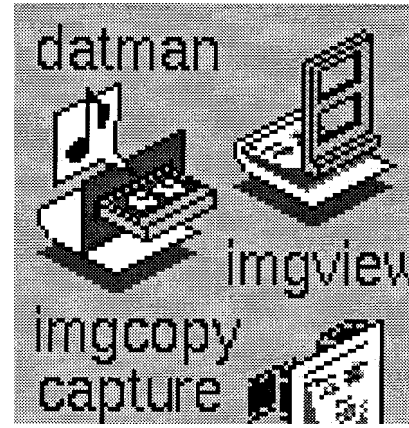

Fig. 4.   Ideal image.

According to the above discussion, our new approach thus has several advantages when compared to the MAP–Huber method given in [27], listed below.

1) In our approach, relatively simple and efficient iterative algorithms can be applied, with assured convergence, whereas the MAP–Huber method results in a more complicated optimization problem.
2) The MAP–Huber method as given in [27] treats only simple blur and white noise.
3) Locally adaptive regularization can be added in a simple fashion, with improved overall performance.

E. Conclusion: Relation to Other Methods

In this section we have shown that the four main known methods for superresolution restoration are highly connected to the new approach presented here. Moreover, since our approach is strongly connected to the classic restoration theory, various tools and ideas to enhance the superresolution restoration result are revealed. These tools enable the improvement of the restoration procedure both from the computational and the output quality points of view. In a sense, we can say that the methodology presented here gives a unified approach toward the superresolution restoration problem and solution, and generalizes (to some extent) the already known methods.

V. SIMULATIONS AND ANALYSIS

In this section, we present simple examples that demonstrate the effectiveness of the proposed method for the superresolution restoration problem. All the simulations correspond to synthetic data, in order to bypass problems which are beyond the scope of this paper such as motion estimation, and the blurring function estimation. We start with a single high-quality image of size $[100 \times 100]$ shown in Fig. 4, from which we generated 16 blurred, downsampled, and noisy images of size $[50 \times 50]$.

The degradation includes affine motion (with zoom ratio in the range [0.9,1.1], rotation in the range [0,50°], and translation in the range [−5,5] pixels), blur with the 1-D separable kernel $[0.7 \quad 1.0 \quad 0.7]/2.4$, a $2:1$ decimation ratio, and additive white Gaussian noise with $\sigma = 3$. All the degraded images are shown in Fig. 5.

10

Fig. 5.   The 16 measured images.

Figs. 6–9 show the restored images using the hybrid restoration algorithm. In the first ten iterations ($k = 1 \cdots 10$) the step I_t [see (2.20)] consists of using the CG algorithm. In the next 30 iterations ($k = 11 \cdots 40$) the CG is replaced by the SD algorithm. In both phases of the algorithm there is one projection to the set presented in (2.15) limiting the output image gray level to the range [0,63]. The initialization image was chosen to be the first measured image after additional blur, and interpolation. We made an attempt to use the initialization as was recommended in [22], but found out that the benefit was minor in terms of convergence rate. Exact known motion flow, blur function, and decimation ratio were used in the restoration process. The various restored images correspond to different regularization terms as is listed in the figures. The first image (Fig. 6) contains no regularization [$\beta = 0$ in (2.19)]. The second image (Fig. 7) corresponds to LSI regularization ($V = I$) Laplacian as the smoothness operator, and $\beta = 2$. Fig. 8 presents the restored image with local adaptive regularization, using the β as before, and diagonal matrix V with diagonal values in the range [0.01,1], defined by

$$V(i + N(j - 1)) = [\alpha \cdot \log \{1 + g(i,j)\} + 1]^{-1} \qquad (5.1)$$

where $g(i,j)$ is the smoothness measured by the gradient on the initialization image and $\alpha = 99$. The final image, Fig. 9, presents the result of an alternative approach to the locally adaptive regularization procedure as was presented in
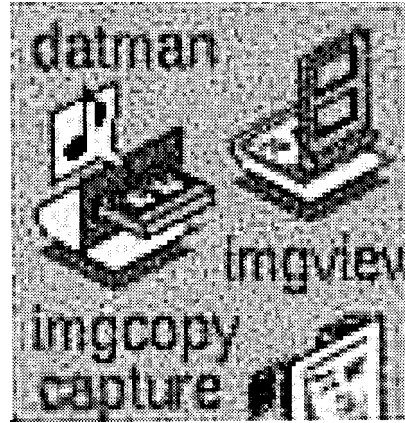
Fig. 6.   Superresolution restoration result without regularization (MSE = 96.17).

the previous section for the frequency domain superresolution restoration. Instead of solving the optimization problem with arbitrary diagonal matrix V, we use two restored images, one without regularization $\underline{X}^1_{out}$ and one with LSI regularization, yielding $\underline{X}^2_{out}$, as is presented in Figs. 6 and 7. The locally adaptive restoration result is constructed by (4.2).

Summarizing the obtained results, we first see that a superresolution image can be generated; the text is readable and the icons are recognizable, in all cases, after the restoration. As expected, the restoration without regularization gives ringing effects [4]. The ringing effects are reduced effectively by the
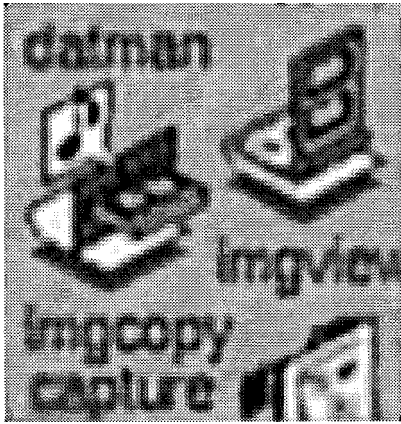
Fig. 7.   Superresolution restoration result with LSI regularization (MSE = 154.82).

Fig. 8.   Superresolution restoration result with locally adaptive regularization: the direct approach (MSE = 112.22).

LSI regularization, causing degradation in the sharpness of the output edges. The two LSV regularization approaches give relatively sharp images, as can be seen in Figs. 8 and 9, with reduced ringing in the smooth regions. It is difficult to compare the two LSV regularization methods for the general case, because different definitions of the matrix V might change significantly the restoration results.

VI. CONCLUSION

This paper addresses the superresolution restoration problem: Namely, given a number of moved, blurred, _decimated_, and noisy versions of a single ideal image, one wants to restore the original image. To solve this problem, a new general model was introduced here. This model enabled the direct generalization of classic tools from restoration theory to the new problem. In this context, the ML, the MAP, and the POCS methods are all shown to be directly and simply applicable to superresolution restoration with equivalencies between these methods. The restoration problem at hand in each of these approaches reduces to the problem of solving a very large set of sparse linear equations.

A hybrid algorithm is proposed that combines the benefits of the simple ML estimator, and the ability of the POCS to incorporate nonellipsoids constraints. This hybrid algorithm solves a constrained convex minimization problem, combining all the _a priori_ knowledge on the required result into the restoration process. An efficient iterative two-phase algorithm is presented for solving the defined problem, and convergence is assured to the optimal point. Simulations are performed to demonstrate superresolution restoration using the hybrid algorithm.

An interesting question with regard to superresolution restoration is raised and treated in this paper. Typically, superresolution restoration methods assume that motion exists between the measured images [13], [27]. The question whether the motion is necessary for superresolution restoration ability is not treated in the literature. We demonstrate that, indeed, there is an ability to restore an image with improved resolution, based on several motionless blurred, decimated, and noisy images.

The proposed methodology is compared to known superresolution methods [13]-[25]. It is shown that the presented methodology gives a unified approach toward the superreso-
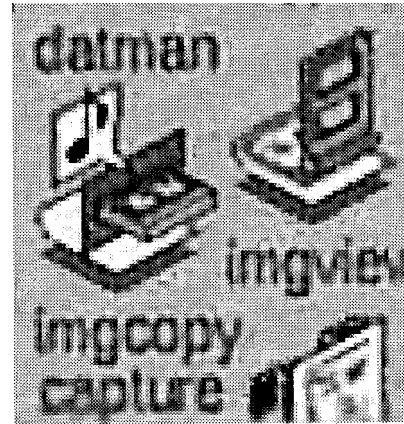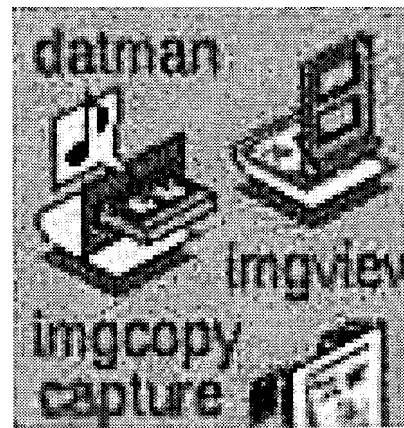

Fig. 9.   Superresolution restoration result with locally adaptive regularization: by linear combination (MSE = 94.57)

lution restoration problem and solution, and generalizes the already known methods. However, this generalization is not ideal, since MAP estimator with nonquadratic priors, POCS with L_∞ norm, and other possible methods are not special cases of the proposed approach. Rather, we have shown in this paper a generalization of well-known methods from restoration theory to the superresolution problem. Establishing such connection enables to import various other known tools from restoration theory, to be implemented in the more complicated problem of superresolution restoration.

APPENDIX A
HYBRID RESTORATION METHOD

Definition A-1: The following constrained quadratic optimization problem is defined as **P**:

$$\text{Minimize} \quad \varepsilon^2(\underline{X}) = \underline{X}^T R \underline{X} - 2\underline{X}^T \underline{P} + q$$
$$\text{subject to} \quad \{\underline{X} \in \Im_k \quad 1 \le k \le M\}$$

where R is a positive definite matrix, q is a constant which ensures that $\varepsilon^2(\underline{X}) \ge 0$, and $\Im_k$ are all closed convex sets. ∎

Let $\underline{X}_{opt}$ denote the optimal solution of the problem **P** when no constraints are added to the optimization criteria. Clearly, this vector is given by the equation $\underline{X}_{opt} = R^{-1}\underline{P}$ and is unique. Let $\Im_0$ denote the (assumed) nonempty intersection of the M convex sets $\Im_k$, namely, $\Im_0 = \bigcap_{k=1}^M \Im_k$. This intersection set is closed and convex too. The problem **P**

is equivalent to a problem where the M convex constraints are replaced by a single constraint $\underline{X} \in \Im_0$.

Theorem A1: If $\underline{X}_{opt} \notin \Im_0$ then the solution to the problem **P** is a point on the boundary of $\Im_0$.

Proof: The proof is by contradiction. Let $\underline{X}_0 \in int\{\Im_0\}$ be the solution of the problem **P**. Then, because of the convexity of the quadratic error $\varepsilon^2(\underline{X})$, it is clear that

$$\varepsilon^2(\underline{X}) < \varepsilon^2(\underline{X}_0) \tag{A.1}$$

for $\underline{X} = \alpha\underline{X}_{opt} + (1-\alpha)\underline{X}_0$ for all $0 < \alpha \leq 1$. Since $\underline{X}_0 \in int\{\Im_0\}$, there exists $0 < \alpha_1 \leq 1$ such that

$$\underline{X} = \alpha_1\underline{X}_{opt} + (1-\alpha_1)\underline{X}_0 \in \Im_0$$

which, together with (A.1) leads to a contradiction. ∎

Theorem A2: The problem **P** has a unique solution.

Proof: We refer to the problem **P** with the constraint $\underline{X} \in \Im_0$. If $\underline{X}_{opt} \in \Im_0$ then $\underline{X}_{opt}$ is the solution of the problem **P**, and since this point is unique, we have a unique solution. Now, assuming that $\underline{X}_{opt} \notin \Im_0$, we have from the previous theorem that the solution (or solutions if there are several) is on the boundary of $\Im_0$. Assume that $\underline{X}_1$ and $\underline{X}_2$ are two distinct solutions to the problem **P**. Then

1. $\underline{X}_1; \underline{X}_2 \in$ Boundary of $\Im_0$
2. $\varepsilon^2(\underline{X}_1) = \varepsilon^2(\underline{X}_2)$
3. $\forall \underline{X} \neq \underline{X}_1; \underline{X}_2 \varepsilon^2(\underline{X}) > \varepsilon^2(\underline{X}_1) = \varepsilon^2(\underline{X}_2)$

$$\tag{A.2}$$

Since $\Im_0$ is convex $\underline{X}_0 = \alpha\underline{X}_1 + (1-\alpha)\underline{X}_2 \in \Im_0 \; \forall 0 \leq \alpha \leq 1$. Furthermore, because of the nonsingularity of **R**, it can readily be shown that

$$\varepsilon^2(\underline{X}_0) = \varepsilon^2(\alpha\underline{X}_1 + (1-\alpha)\underline{X}_2) < \alpha\varepsilon^2(\underline{X}_1)$$
$$+ (1-\alpha)\varepsilon^2(\underline{X}_2) = \varepsilon^2(\underline{X}_1)$$
$$\text{for all} \quad 0 \leq \alpha \leq 1 \tag{A.3}$$

and since $\underline{X}_0 \in \Im_0$, this leads to contradiction and we must have a unique solution. ∎

In Section II, we have presented an iterative algorithm to solve the problem **P**. We show that the proposed algorithm converges to the solution of **P**.

Theorem A3: Let J be the projection onto the closed convex set $\Im_0$ (using the Euclidian norm), and I_t defined as the steepest descent mapping, namely

$$I_t\{\underline{X}\} = \underline{X} - \mu[R\underline{X} - \underline{P}] \tag{A.4}$$

where R and $\underline{P}$ are the matrix and vector defining **P**, and $\mu > 0$ is the stepsize chosen such that $(I - \mu R)$ has all its eigenvalues in (-1,1). Define the algorithm

$$\underline{X}_{k+1} = J\{I_t\{\underline{X}_k\}\} \tag{A.5}$$

Then, (A.5) converges globally to the solution of **P**.

Proof: From the properties of POCS [see, e.g., [5] and (A.4)] we have

$$\|J\{I_t\{\underline{X}\}\} - J\{I_t\{\underline{Y}\}\}\| \leq \|I_t\{\underline{X}\} - I_t\{\underline{Y}\}\|$$
$$\leq \|I - \mu R\| \cdot \|\underline{X} - \underline{Y}\| \leq \|\underline{X} - \underline{Y}\| \quad \forall \underline{X}, \underline{Y} \in \mathbf{R}^n. \tag{A.6}$$

Hence, $J\{I_t\{\bullet\}\}$ is a contraction mapping and as such has a unique fixed point. Then, clearly the algorithm in (A.5) converges to $\underline{X}_\infty$ regardless of the initialization. $\underline{X}_\infty$ satisfies the following equation:

$$J\{I_t\{\underline{X}_\infty\}\} = \underline{X}_\infty. \tag{A.7}$$

It can be readily shown (see [5]) that for any $\underline{X} \notin \Im_0$ we have

$$\langle \underline{X} - J\{\underline{X}\}, \underline{Y} - J\{\underline{X}\} \rangle$$
$$= [\underline{X} - J\{\underline{X}\}]^T[\underline{Y} - J\{\underline{X}\}] \leq 0 \quad \forall \underline{Y} \in \Im_0. \tag{A.8}$$

In case $\underline{X}_{opt} \in \Im_0$, $I_t\{\underline{X}_{opt}\} = \underline{X}_{opt}$ so $J\{\underline{X}_{opt}\} = \underline{X}_{opt}$ and the theorem holds. In case $\underline{X}_{opt} \notin \Im_0$, $I_t\{\underline{X}_\infty\} = \underline{X}_\infty - \mu[R\underline{X}_\infty - \underline{P}]$. Since $\underline{X}_{opt} \notin \Im_0$ we have that $\underline{X}_\infty \neq \underline{X}_{opt}$ and thus, the local gradient of $\varepsilon^2(\underline{X})$ at the point $\underline{X}_\infty$ is nonzero and thus $I_t\{\underline{X}_\infty\} = \underline{X}_\infty - \mu[R\underline{X}_\infty - \underline{P}] \notin \Im_0$. Otherwise we would have

$$J\{I_t\{\underline{X}_\infty\}\} = I_t\{\underline{X}_\infty\} \neq \underline{X}_\infty$$

which contradicts (A.7). Then, by (A.7) and (A.8) we have

$$[I_t\{\underline{X}_\infty\} - J\{I_t\{\underline{X}_\infty\}\}]^T[\underline{Y} - J\{I_t\{\underline{X}_\infty\}\}]$$
$$= \mu[\underline{P} - R\underline{X}_\infty]^T[\underline{Y} - \underline{X}_\infty] \leq 0 \forall \underline{Y} \in \Im_0 \tag{A.9}$$

Hence

$$\varepsilon^2(\underline{Y}) - \varepsilon^2(\underline{X}_\infty) = \underline{Y}^T R\underline{Y} - \underline{X}_\infty^T R\underline{X}_\infty$$
$$- 2\underline{P}^T(\underline{Y} - \underline{X}_\infty) \forall \underline{Y} \in \Im_0 \tag{A.10}$$

and by using the inequality given in (A.9) in the above equation, we get

$$\varepsilon^2(\underline{Y}) - \varepsilon^2(\underline{X}_\infty) \geq \underline{Y}^T R\underline{Y} - 2\underline{Y}^T R\underline{X}_\infty + \underline{X}_\infty^T R\underline{X}_\infty$$
$$= [\underline{X}_\infty - \underline{Y}]^T R[\underline{X}_\infty - \underline{Y}] \geq 0 \forall \underline{Y} \in \Im_0. \tag{A.11}$$

and therefore $\underline{X}_\infty$ is the solution of the problem **P**. ∎

Note that, from the above, we have that if the iterative algorithm is not the steepest descent, then the limiting point is not *necessarily* the optimal solution of **P**.

References

[1] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. New York: Addison-Wesley, 1987.

[2] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1991.

[3] A. K. Jain, *Fundamentals in Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[4] R. L. Lagendijk and J. Biemond *Iterative Identification and Restoration of Images*. Boston, MA: Kluwer, 1991.

[5] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 694–702, 1978.

[6] A. K. Katsaggelos, "A multiple input image restoration approach," *J. Vis. Commun. Image Representat.*, vol. 1, pp. 93–103, Sept. 1990.

[7] D. C. Ghiglia, "Space-invariant deblurring given N independently blurred images of a common object," *J. Opt. Soc. Amer.*, vol. 1, pp. 398–402, Apr. 1984.

[8] B. R. Hunt and O. Kubler, "Karhunen–Loeve multispectral image restoration, Part I: Theory," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 592-599, June 1984.

[9] S. J. Ko and Y. H. Lee, "Nonlinear spatio-temporal noise suppression techniques with applications in image sequence processing," *IEEE Int. Symp. CIS*, 1991, vol. 5, pp. 662–665.

[10] T. S. Huang, *Image Sequence Analysis*. New York: Springer-Verlag, 1981.

[11] ———, *Image Sequence Processing and Dynamic Scene Analysis*. New York: Springer-Verlag, 1983.

[12] A. J. Patti, A. M. Tekalp, and M. I Sezan, *Image Sequence Restoration and De-Interlacing by Motion-Compensated Kalman Filtering*, SPIE, vol. 1903, 1993.

[13] T. S. Huang and R. Y. Tsay, "Multiple frame image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1, T. S. Huang, Ed. Greenwich, CT: JAI, pp. 317–339, 1984.

[14] S. P. Kim, N. K. Bose, and H. M Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1013–1027, June 1990.

[15] N. K. Bose, S. P. Kim, and H. M Valenzuela, "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Minneapolis MN, vol. V, pp. 269–272, 1993.

[16] H. C. Kim, "High resolution image reconstruction from undersampled multiframes," Ph.D. dissertation, Pennsylvania State Univ., Univ. Park, PA, 1994.

[17] A. Papoulis, "Generalized sampling theorem," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 652–654, Nov. 1977.

[18] L. J. Yen, "On nonuniform sampling of bandwidth limited signals," *IRE Trans. Circuits Theory*, vol. 3, pp. 251–257, Apr. 1956.

[19] H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP: Graph. Models Image Process.*, vol. 54, pp. 181–186, Mar. 1992.

[20] C. Srinivas and M. D. Srinath, "A stochastic model-based approach for simultaneous restoration of multiple miss-registered images," *SPIE*, vol 1360, pp. 1416-1427, 1990.

[21] S. Peleg, D. Keren, and L. Schweitzer, "Improving image resolution using subpixel motion," *Pattern Recognit. Lett.*, vol. 5, pp. 223–226, Mar. 1987.

[22] M. Irani and S. Peleg, "Improving resolution by Image Registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 231–239, Mar. 1991.

[23] ———, "Motion analysis for image enhancement: resolution, occlusion, and transparency," *J. VCIR*, vol. 4, pp. 324–335, Dec. 1993.

[24] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration," *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, San Francisco, CA., Mar. 1992, vol. III, pp. 169–172.

[25] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur," in *Proc. ICIP*, Austin, TX, Nov. 1994, pp. 343-347.

[26] ———, "High-resolution standards conversion of low-resolution video," in Proc. *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Detroit, MI., May 1995, vol. II, pp. 2197-2200.

[27] R. R. Schultz and R. L. Stevenson, "Improved definition video frame enhancement," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Detroit, MI., May 1995, vol. IV, pp. 2169-2172.

[28] M. Elad and A. Feuer, "On restoration and super-resolution for continuous image sequence-adaptive filtering approach," Int. Rep. 942, The Technion–Israel Inst. Technol., Haifa, Oct. 1994.

[29] D. M. Young, *Iterative Solution of Large Linear Systems*. New York: Academic, 1971.

[30] F. C. Schweppe, *Uncertain Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[31] R. A. Horn and C. J. Johnson, *Matrix Analysis*. Cambridge, MA: Cambridge Univ. Press, 1985.

[32] J. L. Brown, "Multi-channel sampling low-pass signals," *IEEE Trans. Circuits Syst.*, vol. CAS-28, Feb. 1981.

[33] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred noisy and under-sampled measured images," Int. Rep. 973, The Technion, Israel Inst. Technol., Haifa, June 1995.

Michael Elad was born in Haifa, Israel, on December 10, 1963. He received the B.Sc., M.Sc., and D.Sc. degrees from the Electrical Engineering Department, Technion–Israel Institute of Technology, Haifa, in 1986, 1988, and 1997, respectively.

He is currently employed by Hewlett-Packard Laboratories–Israel, Technion, Haifa. His current research interests include reconstruction and estimation problems in image processing and low-level vision, adaptive filtering algorithms, and pattern recognition.

Arie Feuer (S'74–M'86–SM'93) received the B.Sc and M.Sc. degrees from the Technion–Israel Institute of Technology, Haifa, in 1967 and 1973 respectively, and the Ph.D. degree from Yale University, New Haven, CT, in 1978.

From 1967 to 1970, he was with Technomatic–Israel, working in factory automation. From 1978 through 1983, he was with Bell Labs, Holmdel, NJ, studying telephone network performance. Since 1983, he has been with the Faculty of Electrical Engineering, Technion. His research interests are in adaptive systems and sampled data systems, both in control and in signal and image processing.