Super-Resolution Reconstruction of Compressed Video Using Transform-Domain Statistics

Bahadir K. Gunturk, *Student Member, IEEE*, Yucel Altunbasak, *Senior Member, IEEE*, and Russell M. Mersereau, *Fellow, IEEE*

***Abstract*—Considerable attention has been directed to the problem of producing high-resolution video and still images from multiple low-resolution images. This multiframe reconstruction, also known as super-resolution reconstruction, is beginning to be applied to compressed video. Super-resolution techniques that have been designed for raw (i.e., uncompressed) video may not be effective when applied to compressed video because they do not incorporate the compression process into their models. The compression process introduces quantization error, which is the dominant source of error in some cases. In this paper, we propose a stochastic framework where quantization information as well as other statistical information about additive noise and image prior can be utilized effectively.**

***Index Terms*—DCT-domain reconstruction, MAP, multiframe image reconstruction, POCS, super-resolution.**

I. Introduction

An IMPORTANT problem that arises frequently in visual communications and image processing is the need to enhance the resolution of a still image extracted from a video sequence or of the video sequence itself. This enhanced resolution is possible because the spatial correlations between successive image frames can be exploited. Such a multiframe reconstruction process is usually called super-resolution reconstruction.

Super-resolution reconstruction has many applications. One is in the design of high definition television (HDTV) sets. As the use of HDTV sets becomes widespread, a clear need for systems that enhance standard definition TV signals to match the quality and resolution of HDTV displays will develop. A related super-resolution problem arises when we need to create an enhanced-resolution still image from a video sequence, as when printing stills from video sources. The human visual system requires a higher resolution for a still image than for a sequence of frames, with the same perceptual quality. NTSC video yields at most 480 vertical lines, whereas more than twice as many lines are required to print with reasonable resolution on modern printers. Another application area of super-resolution reconstruction is aerial/satellite imaging. Because of the vast distances involved, some objects may not be properly resolved. When multiple images of the scene are available, super-resolution reconstruction algorithms can be used to resolve details that would be impossible otherwise. Super-resolution reconstruction also finds applications in security/surveillance systems, forensic science, medical imaging, and astronomy.

Most of the work done in the area of super-resolution reconstruction has not considered the compression process [1]–[17]. The input signal (video/image sequence) is assumed to exist in a raw format instead of a compressed format. However, because of the limited resources that are often available (bandwidth, storage space, I/O requirements, etc.), compression has become a standard component of almost every data communication application. Printing from MPEG video sources, by definition, involves compressed video, standardized SDTV signals are MPEG-2 compressed, and digital video cameras typically store images in a compressed format. Unfortunately, super-resolution algorithms designed for uncompressed data do not perform well when directly applied to decompressed image sequences, especially for high compression rates. The reason is that the quantization error introduced during the compression/quantization process is often the dominant source of error when the compression rate is high and this error is not modeled.

In this paper, we propose a Bayesian super-resolution reconstruction technique that models compression and exploits the quantization step size information (available in the data bitstream) in reconstruction. The proposed algorithm allows us to use the statistical information about the quantization noise and the additive noise at the same time. The framework is especially designed for the popular discrete cosine transform (DCT) based video standards such as MPEG, H.261, and DV, although it can easily be generalized to any compression method where the transform involved is linear.

In Section II, we review the state-of-the-art in the area of super-resolution reconstruction; and then we present a general image acquisition and video compression model in Section III. Section IV provides a Bayesian framework for the resolution enhancement of compressed video problem and one possible approach for its solution. Experimental results are presented in Section V. Finally, in Section VI, our conclusions are discussed, and some unsolved problems are given.

II. Previous Work

The super-resolution idea was first addressed by Tsai and Huang [1], who used the aliasing effect to restore a high-resolution image from multiple low-resolution low-resolution

Manuscript received April 10, 2001; revised October 15, 2002. This work was supported in part by the Office of Naval Research (ONR) under Award N00014-01-1-0619 and by the National Science Foundation under Award CCR-0113681. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Timothy J. Schulz.

B. K. Gunturk is with the Louisiana State University, Baton Rouge, LA 70803 USA (e-mail: bahadir@ece.lsu.edu).

Y. Altunbasak and R. M. Mersereau are with the Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: yucel@ece.gatech.edu; rmm@ece.gatech.edu).

Digital Object Identifier 10.1109/TIP.2003.819221

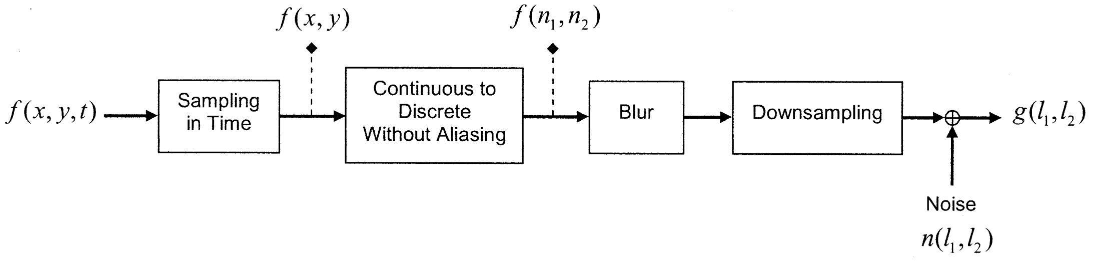


Fig. 1. Image acquisition model used in super-resolution reconstruction. Because the reconstructed signal must be digital, the model includes a discretization block to convert $f(x,y)$ to $f(n_1,n_2)$. $f(n_1,n_2)$ is the signal to be reconstructed.

images. This work was followed by some early approaches [2]–[6], in which the reconstruction was limited to enhancement in the presence of linear shift-invariant blurs and global translational motion. More realistic approaches that allow for linear shift-variant blur, arbitrary motion between the frames, nonwhite noise, etc., can be divided into two distinct groups. One group uses deterministic methods, such as Projections Onto Convex Sets (POCS), to enhance the resolution in the spatial domain without taking any source statistics into account [7]–[13]. Stark and Oskoui [7] used the POCS method to reduce the blur introduced by low-resolution sensors. Tekalp *et al.* extended this POCS formulation to include sensor noise [8]. Later in [9], Patti *et al.* incorporated the time-varying motion blur induced by the motion of the camera, and an arbitrary sampling lattice into Tekalp's model [8]. Other iterative approaches were also proposed. Irani and Peleg [10] used the method of iterated backprojections. Their method assumed translational and rotational motion between the low-resolution frames. Later, [11] and [12] extended this method for more general motion models. Another method proposed by Komatsu *et al.* [13] used a Landweber iteration technique.

The second group of methods is based on a statistical formulation, such as a maximum likelihood or maximum *a posteriori* probability (MAP) estimate [14]–[16]. In [14], Cheeseman *et al.* used a Gaussian model for all distributions, and employed Jacobi's method to solve the problem iteratively. Schultz and Stevenson [15] used a Markov Random Field model for the high-resolution target image, aiming to preserve the edges in the reconstruction by means of a Huber edge penalty function. Elad and Feuer [16] proposed a hybrid method that applies the set theoretic (POCS) and stochastic estimation approaches iteratively. Borman and Stevenson [17] extended the approach in [15] to incorporate temporal smoothness constraints in the prior image model.

All of these methods are based on the assumption that the low-resolution images are available in the spatial domain, i.e., it is assumed that there is no compression stage. In this paper, we focus on super-resolution from a video source that is available only in a compressed format such as MPEG, H.263 or DV. In contrast to the abundance of methods proposed to enhance raw video, there are only a few methods that have been proposed for MPEG-compressed video. Chen and Schultz [18] propose to decompress the MPEG video and then use the uncompressed-video algorithm given in [15]. The drawback is that decompression discards important information about the quantization error that was introduced when the video was compressed. [19] demonstrated the importance of properly handling the quantization information, and suggested a solution that explicitly incorporates the compression process. This method extends the model given in [9] by adding the MPEG stages, and uses the quantization information as the basis for a POCS-based algorithm that operates in the compressed domain. However, in this approach, all sources of error except for the quantization error are ignored, which may not be a good assumption at medium-to-high bit rates. It is also difficult with the POCS approach to impose additional constraints on the reconstructed frame. There are also several Bayesian algorithms that are designed for compressed video. In [20] and [21], the quantization noise is modeled in the DCT domain and transformed back to the spatial domain. In [22], the algorithm is designed to penalize any artifacts formed during the quantization process. [23] proposes to compute the joint statistics of the spatial quantization and additive noises.

In this paper, we also propose a Bayesian method; however, it is different from the previous approaches in the sense that the quantization information is utilized directly in the DCT domain. It is possible to treat the DCT coefficients separately in a way that depends on their statistical distribution or reliability. The method also allows the use of source statistics and additional reconstruction constraints, such as those that might aid in blocking artifact reduction and edge enhancement. We will assume Gaussian models and derive the equations necessary for a maximum *a posteriori* probability estimator. The proposed method is compared with spatial-domain MAP and POCS approaches that do not consider the compression process.

III. Imaging Model

This section extends a general video acquisition model to accommodate block-DCT based compression. The result is a linear set of equations that relates the (unobserved) high-resolution source images to the observed data: the quantized DCT coefficients of low-resolution frames. We use this set of equations to establish the Bayesian framework in the next section.

We begin by reviewing a typical image acquisition model depicted in Fig. 1. According to this model, a spatially and temporally continuous input signal $f(x,y,t)$ is sampled in time to form a spatially continuous image $f(x,y)$. Here, (x,y) represents the continuous spatial coordinates, and t represents time. Because we are dealing with digital images, this continuous image is converted to a discrete image $f(n_1,n_2)$. (n_1,n_2) are the discrete spatial coordinates, and $f(n_1,n_2)$ is the high-reso-

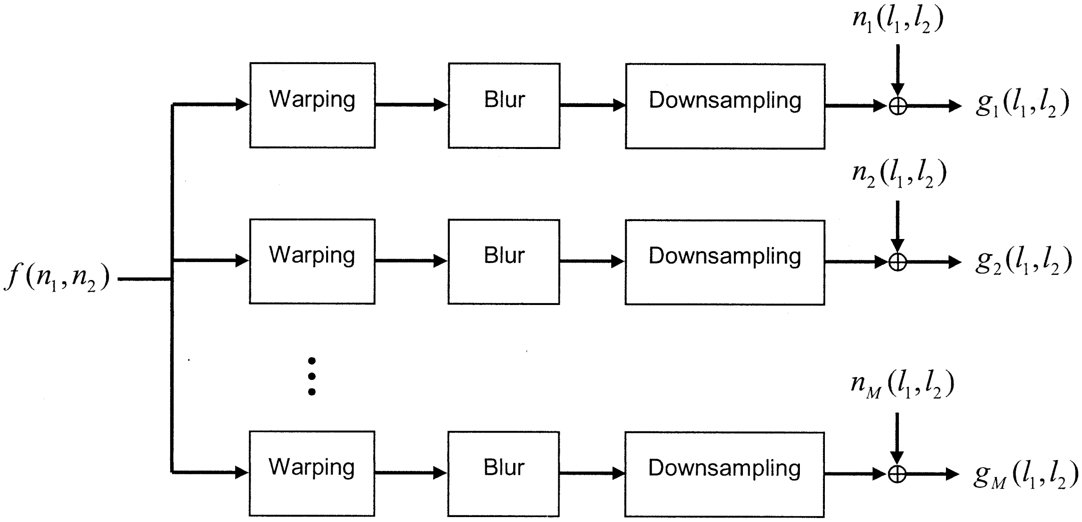


Fig. 2. Video acquisition model.

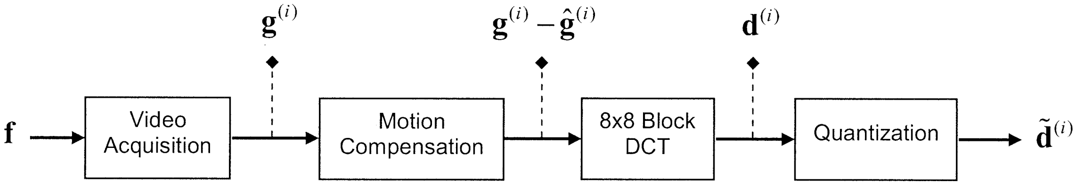


Fig. 3. MPEG compression is appended to the video acquisition.

lution image that we are trying to reconstruct through super-resolution reconstruction. $f(n_1, n_2)$ is then distorted by sensor and optical blurs. Sensor blur is caused by integrating the received light over the finite nonzero sensor cell area, while optical blur includes optical distortions such as out of focus. It is also possible to include motion blur to the blurring process, which is due to nonzero shutter time. The blurred image is then downsampled to account for the insufficient sensor density. There may also be additive noise causing further degradation. The final is image is represented by $g(l_1, l_2)$, where (l_1, l_2) are the discrete spatial coordinates.

We now extend this image acquisition model to a video acquisition model by incorporating the relative motion among different frames of a video. The model is depicted in Fig. 2. $f(n_1, n_2)$ is the high-resolution discrete image we want to reconstruct and $g_i(l_1, l_2)$ are the low-resolution frames that are observed. i is the frame number, and M is the total number of frames. The warping block accounts for the relative motion between the observations, which can be global as well as dense. The rest of the model (blurring, downsampling, and additive noise) is same as in the image acquisition model. All the processes in this video acquisition model are linear, and the relationship between the high-resolution image $f(n_1, n_2)$ and the recorded low-resolution images $g_i(l_1, l_2)$ can be formulated as follows [24]

$$g_i(l_1, l_2) = \sum_{n_1, n_2} h_i(l_1, l_2; n_1, n_2) f(n_1, n_2) + n_i(l_1, l_2) \quad (1)$$

where $h_i(l_1, l_2; n_1, n_2)$ is the linear mapping that includes warping, blurring, and downsampling operations, and $n_i(l_1, l_2)$ is the additive noise. The high-resolution and low-resolution sampling lattice indices (i.e., pixel coordinates) are (n_1, n_2) and (l_1, l_2), respectively. Note that (1) provides a linear set of equations that relates the high-resolution image to the low-resolution frames $g_i(l_1, l_2)$ for different values of i. The relation in (1) can also be expressed in a simpler matrix-vector notation, which we will use in the remainder of this paper. Letting $\mathbf{f}$, $\mathbf{g}^{(i)}$, and $\mathbf{n}^{(i)}$ denote the lexicographically ordered versions of $f(n_1, n_2)$, $g_i(l_1, l_2)$, and $n_i(l_1, l_2)$, respectively, we write

$$\mathbf{g}^{(i)} = \mathbf{H}^{(i)}\mathbf{f} + \mathbf{n}^{(i)}, \quad i = 1 \cdots M \quad (2)$$

where M is the total number of observations and $\mathbf{H}^{(i)}$ is a matrix constructed from the blur mapping $h_i(l_1, l_2; n_1, n_2)$.

We now add the MPEG compression stages to this model. As shown in Fig. 3, the low-resolution frame $\mathbf{g}^{(i)}$ is motion compensated (i.e., the prediction frame is computed and subtracted from the original to get a residual image), and the residual is transformed using a series of 8 × 8 block-DCTs to produce the DCT coefficients $\mathbf{d}^{(i)}$. Defining $\hat{\mathbf{g}}^{(i)}$ as the prediction frame and $\mathbf{T}$ as the DCT matrix for the lexicographically ordered images, we write

$$\mathbf{d}^{(i)} = \mathbf{T}\mathbf{H}^{(i)}\mathbf{f} - \mathbf{T}\hat{\mathbf{g}}^{(i)} + \mathbf{T}\mathbf{n}^{(i)}. \quad (3)$$

The prediction frame $\hat{\mathbf{g}}^{(i)}$ is obtained using neighboring frames except for the case of intra-coded frames, where the prediction frame is zero. The DCT coefficients $\mathbf{d}^{(i)}$ are then quantized to produce the quantized DCT coefficients $\tilde{\mathbf{d}}^{(i)}$. The quantization operation is a nonlinear process that will be denoted by the operator $\mathcal{Q}\{\cdot\}$

$$\tilde{\mathbf{d}}^{(i)} = \mathcal{Q}\left\{\mathbf{T}\mathbf{H}^{(i)}\mathbf{f} - \mathbf{T}\hat{\mathbf{g}}^{(i)} + \mathbf{T}\mathbf{n}^{(i)}\right\}. \tag{4}$$

In MPEG compression, quantization is realized by dividing each DCT coefficient by a quantization step size followed by rounding to the nearest integer. The quantization step size is determined by the location of the DCT coefficient, the bit rate, and the macroblock mode [25]. The quantized DCT coefficients $\tilde{\mathbf{d}}^{(i)}$ and the corresponding step sizes are available at the decoder, i.e., they are either embedded in the compressed bit-stream or specified as part of the coding standard. Since the quantization takes place in the transform domain, the natural way to exploit this information is to use it in the DCT domain without reverting back to the spatial domain.

Equation (4) is the fundamental equation that represents the relation between the high-resolution image $\mathbf{f}$ and the quantized DCT coefficients $\tilde{\mathbf{d}}^{(i)}$. In the next section, we formulate a Bayesian super-resolution reconstruction framework based on this equation.

IV. Bayesian Super-Resolution Reconstruction

With a Bayesian estimator, not only the source statistics but also various regularizing constraints can be incorporated into the solution. Bayesian estimators have been frequently used for super-resolution reconstruction. However, in these approaches either the video source is assumed to be available in uncompressed form, or it is simply decompressed prior to enhancement without considering the quantization process. Additive noise is considered as the only source of error. On the other hand, the POCS-based approaches treat the quantization error as the only source of error without considering the additive noise [24]. Clearly, neither of these approaches provides a complete framework for super-resolution. As will be shown, a Bayesian estimator that considers the quantization process can be applied successfully.

In the maximum *a posteriori* probability (MAP) formulation, the quantized DCT coefficients $\tilde{\mathbf{d}}^{(i)}$, the original high-resolution frame $\mathbf{f}$, and the additive noise $\mathbf{n}^{(i)}$ are all assumed to be random processes. Denoting $p\left(\mathbf{f}|\tilde{\mathbf{d}}^{(1)},\ldots,\tilde{\mathbf{d}}^{(M)}\right)$ as the conditional probability density function (PDF), the MAP estimate $\hat{\mathbf{f}}$ is given by

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}}\left\{p\left(\mathbf{f}|\tilde{\mathbf{d}}^{(1)}\cdots\tilde{\mathbf{d}}^{(M)}\right)\right\}. \tag{5}$$

Using the Bayes rule, (5) can be rewritten as

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}}\left\{p\left(\tilde{\mathbf{d}}^{(1)}\cdots\tilde{\mathbf{d}}^{(M)}|\mathbf{f}\right)p\left(\mathbf{f}\right)\right\} \tag{6}$$

where we used the fact that $p\left(\tilde{\mathbf{d}}^{(1)},\ldots,\tilde{\mathbf{d}}^{(M)}\right)$ is independent of $\mathbf{f}$. In order to find the MAP estimate $\hat{\mathbf{f}}$, we need to model the conditional PDF $p\left(\tilde{\mathbf{d}}^{(1)},\ldots,\tilde{\mathbf{d}}^{(M)}|\mathbf{f}\right)$ and the prior PDF $p\left(\mathbf{f}\right)$. Before proceeding, we rewrite (4) by letting $\mathbf{e}^{(i)}$ denote the error introduced by quantization

$$\tilde{\mathbf{d}}^{(i)} = \mathbf{T}\mathbf{H}^{(i)}\mathbf{f} - \mathbf{T}\hat{\mathbf{g}}^{(i)} + \mathbf{T}\mathbf{n}^{(i)} + \mathbf{e}^{(i)}. \tag{7}$$

The quantization error $\mathbf{e}^{(i)}$ is a deterministic quantity that is defined as the difference between $\tilde{\mathbf{d}}^{(i)}$ and $\mathbf{d}^{(i)}$, but it can also be treated as a stochastic vector for reconstruction. There have been a number of studies directed toward modeling the statistical distribution of the quantization error $\mathbf{e}^{(i)}$. In this chapter, we model it as a zero-mean independent identically distributed (IID) Gaussian random process, which leads to a mathematically tractable solution. Using the notation $N\left(\boldsymbol{\mu},\mathbf{C}\right)$ for a normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$, we write

$$\mathbf{e}^{(i)} \sim N\left(\mathbf{0},\mathbf{C}_{\mathbf{e}}\right) \tag{8}$$

where $\mathbf{C}_{\mathbf{e}}$ is the covariance matrix of the quantization error $\mathbf{e}^{(i)}$. The additive noise $\mathbf{n}^{(i)}$ is also modeled as a zero-mean IID Gaussian process

$$\mathbf{n}^{(i)} \sim N\left(\mathbf{0},\mathbf{C}_{\mathbf{n}}\right) \tag{9}$$

where $\mathbf{C}_{\mathbf{n}}$ is the covariance matrix of the additive noise. Since the discrete cosine transform is unitary, the DCT of the noise $\mathbf{n}^{(i)}$ is also an IID Gaussian random process with covariance matrix $\mathbf{T}\mathbf{C}_{\mathbf{n}}\mathbf{T}^{T}$

$$\mathbf{T}\mathbf{n}^{(i)} \sim N\left(\mathbf{0},\mathbf{T}\mathbf{C}_{\mathbf{n}}\mathbf{T}^{T}\right). \tag{10}$$

Because the additive noise and the quantization error have independent Gaussian distributions, the overall noise $\mathbf{T}\mathbf{n}^{(i)} + \mathbf{e}^{(i)}$ is also a Gaussian distribution with a mean equal to the sum of the means, and a covariance matrix equal to the sum of the covariance matrices of $\mathbf{T}\mathbf{n}^{(i)}$ and $\mathbf{e}^{(i)}$

$$\mathbf{T}\mathbf{n}^{(i)} + \mathbf{e}^{(i)} \sim N\left(\mathbf{0},\mathbf{T}\mathbf{C}_{\mathbf{n}}\mathbf{T}^{T} + \mathbf{C}_{\mathbf{e}}\right). \tag{11}$$

Equation (11) gives us an elegant way of combining the statistical information of two different noise processes. This is in contrast to the previous approaches, where only one noise source is considered. We now pursue with the derivation by writing the explicit forms of the probability density functions. Denoting $\mathbf{u}^{(i)} \equiv \mathbf{T}\mathbf{n}^{(i)} + \mathbf{e}^{(i)}$ as the total noise term, and $\mathbf{K} \equiv \mathbf{T}\mathbf{C}_{\mathbf{n}}\mathbf{T}^{T} + \mathbf{C}_{\mathbf{e}}$ as the overall covariance matrix, the probability distribution function of $\mathbf{u}^{(i)}$ is

$$p(\mathbf{u}^{(i)}) = \frac{1}{Z}\exp\left(-\frac{1}{2}\left(\mathbf{u}^{(i)}\right)^{T}\mathbf{K}^{-1}\left(\mathbf{u}^{(i)}\right)\right) \tag{12}$$

where Z is a normalization constant. Using (7) and the PDF of the noise $\mathbf{u}^{(i)}$, the conditional PDF $p\left(\tilde{\mathbf{d}}^{(i)}|\mathbf{f}\right)$ is found to be (see (13) at the bottom of the page). Since the noise is assumed to be

$$p\left(\tilde{\mathbf{d}}^{(i)}|\mathbf{f}\right) = \frac{1}{Z}\exp\left(-\frac{1}{2}\left(\tilde{\mathbf{d}}^{(i)} - \mathbf{T}\mathbf{H}^{(i)}\mathbf{f} + \mathbf{T}\hat{\mathbf{g}}^{(i)}\right)^{T}\mathbf{K}^{-1}\left(\tilde{\mathbf{d}}^{(i)} - \mathbf{T}\mathbf{H}^{(i)}\mathbf{f} + \mathbf{T}\hat{\mathbf{g}}^{(i)}\right)\right) \tag{13}$$

an IID process, the joint PDF $p\left(\tilde{\mathbf{d}}^{(1)}\cdots\tilde{\mathbf{d}}^{(M)}|\mathbf{f}\right)$ is the product of the individual PDFs. As a result, we obtain (see (14) at the bottom of the page) where Z is again a normalization constant. We now need to model the prior distribution $p(\mathbf{f})$ to complete the MAP formulation. Again, we will assume a joint Gaussian model

$$p(\mathbf{f}) = \frac{1}{Z}\exp\left(-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^T\boldsymbol{\Lambda}^{-1}(\mathbf{f}-\boldsymbol{\mu})\right) \quad (15)$$

with $\boldsymbol{\Lambda}$ being the covariance matrix, $\boldsymbol{\mu}$ being the mean of $\mathbf{f}$, and Z being a normalization constant. Substituting (14) and (15) into (6), we end up with the following MAP estimate: (see (16) at the bottom of the page). We finish this section by presenting an approach to solve (16). In the next section we detail the implementation and selection of the parameters and covariance matrices. We will explain how to incorporate the quantization step size information in reconstruction through selection of covariance matrices.

One approach to obtain the MAP estimate in (16) is to use an iterative steepest descent technique. Let $E(\mathbf{f})$ be the cost function to be minimized, then the high-resolution image $\mathbf{f}$ can be updated in the direction of the negative gradient of $E(\mathbf{f})$. At the nth iteration, the high-resolution image estimate is

$$\mathbf{f}_n = \mathbf{f}_{n-1} - \alpha\nabla E(\mathbf{f}_{n-1}) \quad (17)$$

where α is the step size.

From (16), we can choose a slightly generalized cost function as follows:

$$\begin{aligned} &E(\mathbf{f}) \\ &= \frac{1-\lambda}{2} \\ &\quad\times\sum_{i=1}^{M}\left[\left(\mathbf{d}^{(i)}-\mathbf{T}\mathbf{H}^{(i)}\mathbf{f}+\mathbf{T}\hat{\mathbf{g}}^{(i)}\right)^T\mathbf{K}^{-1}\left(\mathbf{d}^{(i)}-\mathbf{T}\mathbf{H}^{(i)}\mathbf{f}+\mathbf{T}\hat{\mathbf{g}}^{(i)}\right)\right] \\ &\quad+\frac{\lambda}{2}(\mathbf{f}-\boldsymbol{\mu})^T\boldsymbol{\Lambda}^{-1}(\mathbf{f}-\boldsymbol{\mu}) \quad (18)\end{aligned}$$

where λ is a number, ($0 \le \lambda \le 1$), that controls the relative contributions of the conditional and prior information in the reconstruction. When λ is set to zero, the estimator behaves like a maximum likelihood (ML) estimator. When λ is made larger, the prior information is more and more important to reconstruction.

Taking the derivative of $E(\mathbf{f})$ with respect to $\mathbf{f}$, the gradient of $E(\mathbf{f})$ can be calculated as

$$\begin{aligned}\nabla E(\mathbf{f}) = -(1-\lambda)\sum_{i=1}^{M}\mathbf{H}^{(i)^T}\mathbf{T}^T\mathbf{K}^{-1}\left(\tilde{\mathbf{d}}^{(i)}-\mathbf{T}\mathbf{H}^{(i)}\mathbf{f}+\mathbf{T}\hat{\mathbf{g}}^{(i)}\right) \\ +\lambda\boldsymbol{\Lambda}^{-1}(\mathbf{f}-\boldsymbol{\mu}). \quad (19)\end{aligned}$$

The step size α in (17) can be fixed or updated adaptively during the iterations. One way is to update it using the Hessian of $E(\mathbf{f})$. In that case, α is updated at each iteration using the formula

$$\alpha = \frac{(\nabla E(\mathbf{f}_{n-1}))^T(\nabla E(\mathbf{f}_{n-1}))}{(\nabla E(\mathbf{f}_{n-1}))^T H(\nabla E(\mathbf{f}_{n-1}))} \quad (20)$$

where H is the Hessian matrix found by

$$H = (1-\lambda)\sum_{i=1}^{M}\mathbf{H}^{(i)^T}\mathbf{T}^T\mathbf{K}^{-1}\mathbf{T}\mathbf{H}^{(i)} + \lambda\boldsymbol{\Lambda}^{-1}. \quad (21)$$

In the reconstruction, everything but $\mathbf{f}$ is known or can be computed in advance. For a specific observation sequence, the quantized DCT coefficients $\tilde{\mathbf{d}}^{(i)}$, the prediction frames $\hat{\mathbf{g}}^{(i)}$, and the quantization step sizes are known; the blur mappings $\mathbf{H}^{(i)}$ and the other statistical/reconstruction parameters are computed or determined beforehand.

V. Experimental Results

We have designed a set of experiments to examine the performance of the proposed algorithm for different quantization levels. We also tested the spatial-domain POCS [9] and spatial-domain MAP [15], [16] algorithms, and compared their results with the results of our DCT-domain MAP algorithm. Before getting into details of the experiments, we want to address some of the implementation issues.

A. Implementation

Although the matrix-vector notation provides a neat formulation, implementing the algorithm with images converted into vectors is problematic. When dealing with large images, the matrices become large enough to cause memory problems and slow reconstruction. Instead, the algorithm was implemented using simple image operations, such as warping, convolution, sampling, scaling, and block-DCT transformation.

$$p\left(\tilde{\mathbf{d}}^{(1)}\cdots\tilde{\mathbf{d}}^{(M)}|\mathbf{f}\right) = \frac{1}{Z}\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\left(\tilde{\mathbf{d}}^{(i)}-\mathbf{T}\mathbf{H}^{(i)}\mathbf{f}+\mathbf{T}\hat{\mathbf{g}}^{(i)}\right)^T\mathbf{K}^{-1}\left(\tilde{\mathbf{d}}^{(i)}-\mathbf{T}\mathbf{H}^{(i)}\mathbf{f}+\mathbf{T}\hat{\mathbf{g}}^{(i)}\right)\right) \quad (14)$$

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}}\left\{\sum_{i=1}^{M}\left[\left(\mathbf{d}^{(i)}-\mathbf{T}\mathbf{H}^{(i)}\mathbf{f}+\mathbf{T}\hat{\mathbf{g}}^{(i)}\right)^T\mathbf{K}^{-1}\left(\mathbf{d}^{(i)}-\mathbf{T}\mathbf{H}^{(i)}\mathbf{f}+\mathbf{T}\hat{\mathbf{g}}^{(i)}\right)\right]+(\mathbf{f}-\boldsymbol{\mu})^T\boldsymbol{\Lambda}^{-1}(\mathbf{f}-\boldsymbol{\mu})\right\} \quad (16)$$

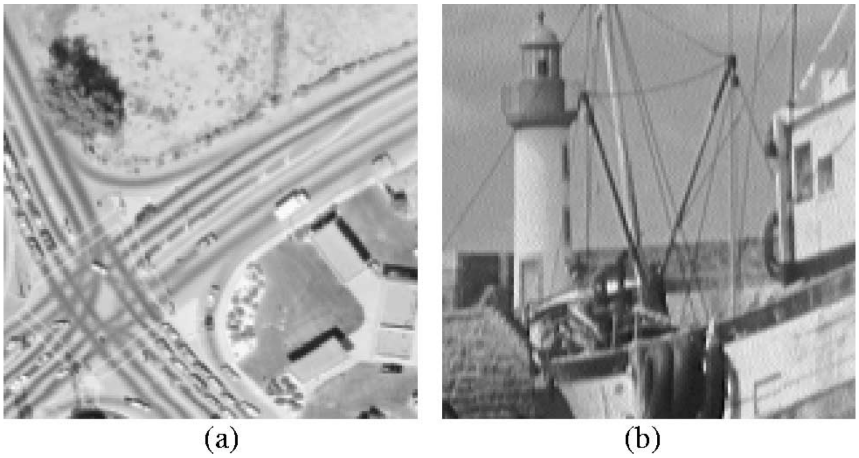


Fig. 4. (a) Original AERIAL image. (b) Original BOAT image.

TABLE I
MEAN SQUARE ERROR (MSE) COMPARISON OF DIFFERENT METHODS FOR AERIAL IMAGE

	MSE of the reconstructed image for different quantization factors					
Method	0.25	0.50	0.75	1.00	1.25	1.50
Bilinear intepolation	109.1	130.4	153.5	176.6	197.9	215.5
Spatial-domain POCS	21.1	63.4	105.0	139.4	168.1	193.2
Spatial-domain MAP	21.2	60.2	102.2	141.9	172.6	196.3
DCT-domain MAP	19.8	55.3	90.3	119.4	147.5	169.6

After determining the blur point spread function (PSF) and the motion vectors between the observations, the high-resolution image is reconstructed as follows. We start by interpolating one of the observed images to obtain an initial estimate $\mathbf{f}_0$. According to (17), we need to calculate the gradient of the cost function and the step size α. Referring to (19), we first need to calculate $\mathbf{H}^{(i)}\mathbf{f}_0$. This is done by motion warping $\mathbf{f}_0$ for the ith frame, convolving with the PSF, and then downsampling. The resulting image and the prediction image $\hat{\mathbf{g}}^{(i)}$ are then transformed to the DCT domain by 8 × 8 block-DCTs. After finding the residual, we need to apply the operations $\mathbf{K}^{-1}$, $\mathbf{T}^T$, and ${\mathbf{H}^{(i)}}^T$. As we mentioned earlier, we assumed statistical independence between the DCT quantization errors, and this results in the covariance matrix $\mathbf{K}$ being diagonal. Therefore, in our implementation, the $\mathbf{K}^{-1}$ is simply computed by dividing each DCT coefficient of the residual with the corresponding variance in $\mathbf{K}$. This is followed by the $\mathbf{T}^T$ operation, which is done by taking the inverse block-DCT. Finally, ${\mathbf{H}^{(i)}}^T$ is implemented by upsampling the image (with zero padding), convolving with the flipped PSF, and motion warping back to the reference frame. (If we let $h(n_1, n_2)$ denote the PSF, the flipped PSF is then $h(-n_1, -n_2)$.) Similar to $\mathbf{K}^{-1}$, $\mathbf{\Lambda}^{-1}$ is also implemented by scaling each pixel by a number because the pixels are also modeled as being statistically independent.

In the computation of α, we need to calculate $(\nabla E(\mathbf{f}_{n-1}))^T (\nabla E(\mathbf{f}_{n-1}))$, which is done by taking the square of each element of $\nabla E(\mathbf{f}_{n-1})$, and then summing them up. The denominator of (20) is obtained similarly. (Apply the operations in (21) on $\nabla E(\mathbf{f}_{n-1})$, multiply the result element by element with $\nabla E(\mathbf{f}_{n-1})$, and then sum them up.)

With this procedure, the reconstruction is achieved faster than working with lexicographically ordered images. We now turn to the experiments.

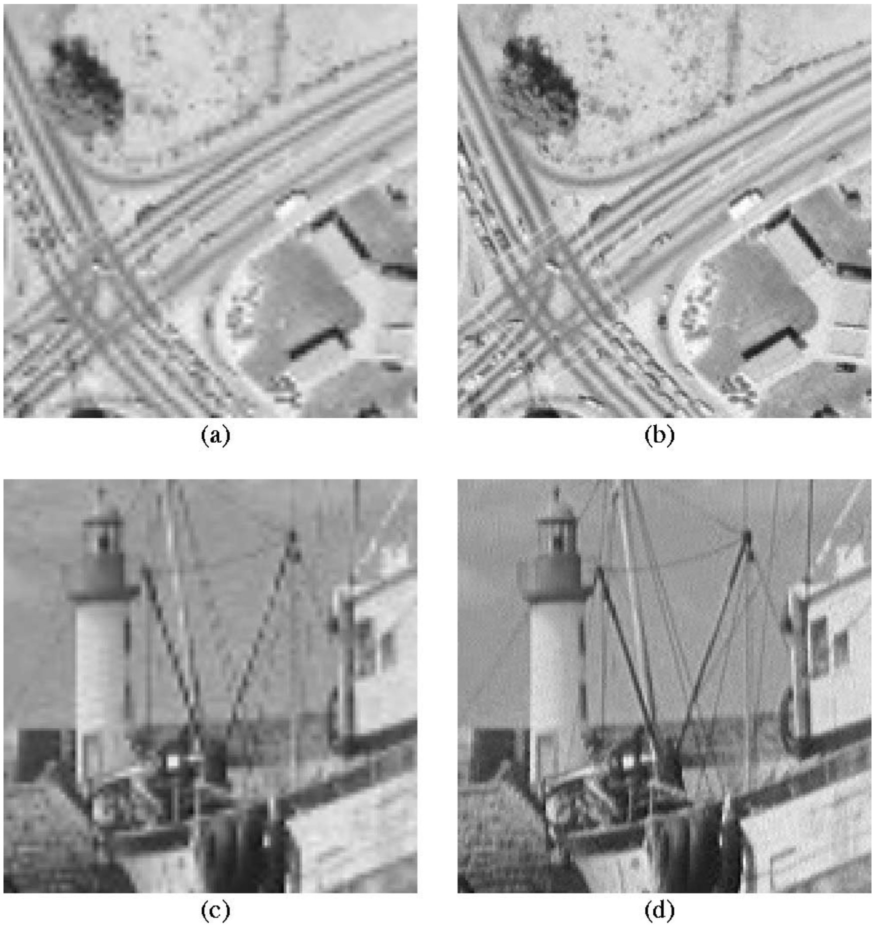
B. Experimental Setup

In order to test the proposed algorithm, we designed a controlled experiment. The AERIAL and BOAT images shown in Fig. 4 are downsampled by two horizontally and vertically to create four low-resolution observations. These observations are then block transformed using 8×8 DCTs. The DCT coefficients are then quantized using the MPEG-2 quantization table for the luminance channel. Spatial-domain POCS, spatial-domain MAP, and the proposed DCT-domain MAP algorithms are tested. In the reconstructions, all four observations are used. The experiments are repeated for different quantization scales, which is done by multiplying the quantization table by a positive real number. The scaling factors used in the experiments are 0.25, 0.5, 0.75, 1.0, 1.25, and 1.5.

In addition to the simulated data, we also tested the algorithm with observations captured with a digital camera. We captured six images from a text document with slightly different viewing positions and six images (zooming license plate) from a moving car. The observations are then quantized with the quantization scaling factor set to 0.25. For the TEXT sequence, the dense motion fields between the observations are calculated using a two-level hierarchical block-based motion estimation algorithm. Block sizes of 12 pixels are used with mean absolute difference as the matching criterion. In the final level of search, quarter pixel motion vectors are sought. For the LICENSE PLATE sequence, we used the Harris corner

TABLE II
MEAN SQUARE ERROR (MSE) COMPARISON OF DIFFERENT METHODS FOR BOAT IMAGE

	MSE of the reconstructed image for different quantization factors					
Method	0.25	0.50	0.75	1.00	1.25	1.50
Bilinear intepolation	143.5	154.9	167.8	179.9	192.0	206.2
Spatial-domain POCS	17.1	42.5	67.1	89.1	109.6	129.6
Spatial-domain MAP	17.1	43.4	68.3	89.7	108.4	126.3
DCT-domain MAP	16.4	38.6	60.8	81.4	101.7	119.1



(a)

(b)

(c)

(d)

Fig. 5. (a) Bilinearly interpolated AERIAL image. (b) Reconstructed AERIAL image. (c) Bilinearly interpolated BOAT image. (d) Reconstructed BOAT image.

detector [26] to select a set of points in the reference image. We then find the correspondence points in the other images using normalized cross correlation. From these correspondences, a least-mean-square estimate of the affine motion parameters is found.

C. Parameter Selection

The spatial-domain POCS algorithm [9] starts with an initial estimate of the high-resolution image, which is obtained by bilinearly interpolating one of the observations. The mapping $\mathbf{H}^{(i)}$ is applied to this initial estimate to compute a prediction of one of the observed images. The difference between the predicted image and the real observed image is backprojected to update the initial estimate. This is repeated for a predetermined number of iterations (typically 15) or until the change in the mean-square-error (MSE) is less than 0.1.

The formulation of the spatial-domain MAP [15], [16] algorithm is similar to the DCT-domain MAP algorithm that we derived in this chapter. In the spatial-domain MAP algorithm, the observations are the low-resolution images, not the quantized DCT coefficients. The spatial-domain MAP algorithm requires the covariance matrices for the additive noise and the prior image. Our DCT-domain MAP algorithm requires the covariance matrix for the quantization error in addition to the additive noise and prior image covariance matrices.

The parameters in the experiments are chosen first intuitively and then finalized by trial-and-error. (Obviously, this is not an

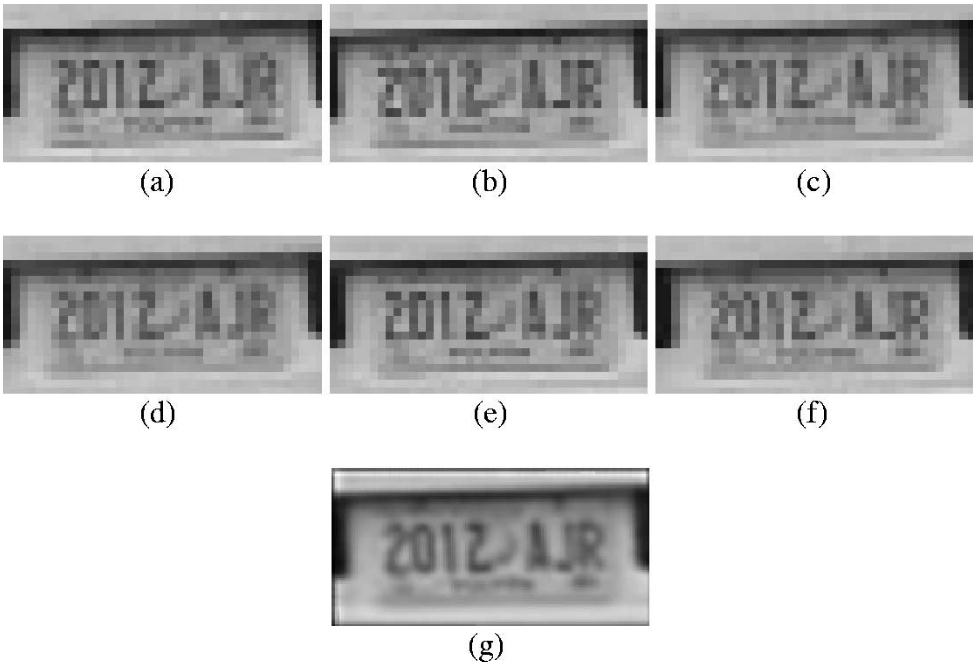


Fig. 6. Results for License Plate sequence; quantization factor is 0.25; enhancement factor is four. (a) to (f) are the images used in reconstruction. (g) The reconstructed image using the DCT-domain MAP algorithm.

optimal approach; we will comment on this issue later in this chapter.) In our formulation, all of the covariance matrices are diagonal as a result of the IID assumption. The variance of the additive noise is set to two, i.e., $\mathbf{C_n}$ is chosen to be a diagonal matrix with twos along the diagonal. After some trial-and-error, λ is set to 0.15, and the diagonal entries of $\mathbf{\Lambda}$ are set to 25. The mean vector $\boldsymbol{\mu}$ is set to the bilinearly interpolated reference frame. Again, the iterations are repeated for a predetermined number of times or until the change in the MSE is less than 0.1.

For the DCT-domain MAP algorithm, *the standard deviation of the quantization noise is proportional to the quantization step size.* Although there is no exact analytical relationship, this is a valid assumption and directly affects the reconstruction. Heuristically, we set the standard deviation to one-fifth of the quantization step size in our experiments. The diagonal entries of $\mathbf{C_e}$ are computed according to the corresponding quantization step sizes, which are available in the data bitstream. The covariance matrix $\mathbf{K}$ is then calculated using the formula $\mathbf{K} = \mathbf{T}\mathbf{C_n}\mathbf{T}^T + \mathbf{C_e}$. (Since $\mathbf{C_n}$ and $\mathbf{C_e}$ are diagonal, and $\mathbf{T}$ is a unitary transform, $\mathbf{K}$ is also diagonal.)

For the real video experiments, the same reconstruction parameters except for the λ are used. For the Text sequence, λ is set to 0.1; for the License Plate sequence, λ is set to 0.7. An additional problem with the real video experiment is that the PSF of the camera is unknown. Therefore, we chose a typical PSF, and tested the reconstruction using that PSF. In the experiments, the PSF was set to a 7 × 7 Gaussian blur with standard deviation of two.

D. Results

The spatial-domain MAP, spatial-domain POCS, and DCT-domain MAP algorithms are tested for quantization scaling factors of 0.25, 0.5, 0.75, 1.0, 1.25, and 1.5. The same observations are used for all algorithms. The MSE comparison for Aerial and Boat images are given in Tables I and II, respectively. As seen in those tables, in all the experiments, the DCT-domain MAP algorithm performed better than the spatial-domain MAP and POCS algorithms, which do not utilize the quantization information. As the quantization step size increases, the relative performance of the DCT-domain MAP algorithm improves. Although the quantitative comparison shows that the DCT-domain MAP algorithm performs better than the other algorithms, the difference is not visually obvious.

In Fig. 5, we provide visual results of the DCT-domain MAP algorithm for Aerial and Boat images. The downsampling factor is two, and four observations are used in both experiments. (The quantization scaling factor in these experiments is 0.25.)

The real video experiments were performed to assess the robustness of the algorithm when the true motion vectors and the PSF are not known. The parameters used in the experiments are given in the previous two subsections. The observations from the License Plate sequence and the reconstructed image are given in Fig. 6. The resolution enhancement factor is four in horizontal and vertical directions. For the Text sequence, the resolution enhancement factor is two. The observations and the reconstructed image for the Text is given in Fig. 7. Although we do not have the true motion vectors and the exact PSF, we still observe improvement in the readability in these experiments.

VI. Conclusion

In this paper, we introduced a super-resolution algorithm that incorporates both the quantization operation and additive sensor noise in a stochastic framework. Since the resolution enhancement problem is cast in a Bayesian framework, additional constraints can easily be incorporated in the form of prior image models. Although a block-based hybrid transform coder was

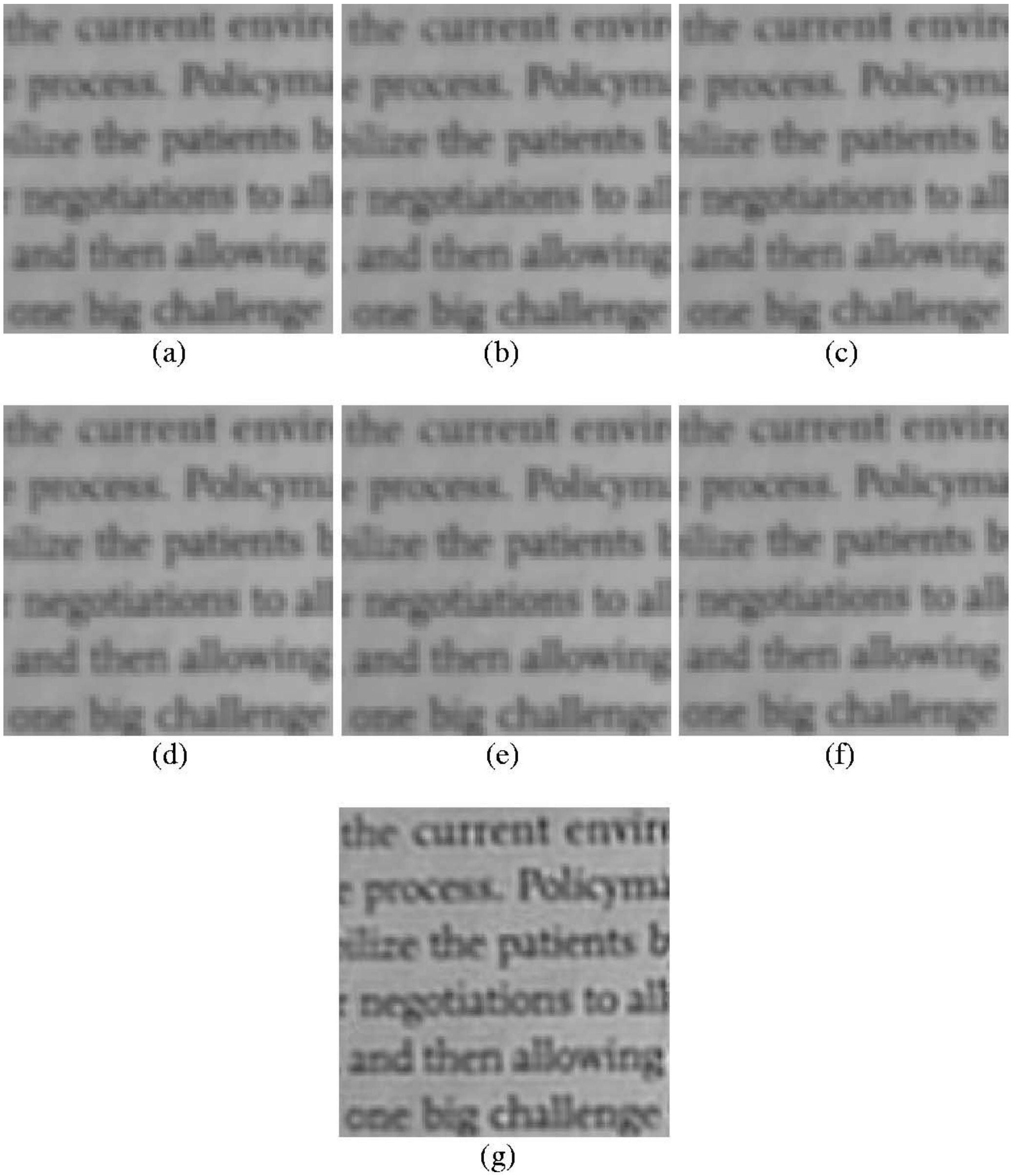


Fig. 7. Results for TEXT sequence; quantization factor is 0.25; enhancement factor is two. (a) to (f) are the images used in reconstruction. (g) The reconstructed image using the DCT-domain MAP algorithm.

emphasized throughout our derivations, the framework is still valid for all video coding standards where the transform utilized is linear.

The proposed algorithm enables distinct treatment of each DCT coefficient. This can be exploited for better performance. For instance, the information coming from high-frequency DCT coefficients can be discarded altogether since they are quantized severely, and the information obtained from those coefficients are likely to be noise. One step beyond this idea is to realize the whole reconstruction in transform domain. That is, the high-resolution image is reconstructed in a transform domain, and then converted back to spatial domain at the end. This way, we can suppress noise and achieve reconstruction at a lower computational complexity.

The problem of high-quality video reconstruction from compressed data is also expected to have broader implications in information fusion and information theory. The proposed research activities deal with fusing information spread over multiple frames in an optimal manner. In addition, they are closely related to the information content and the compressibility of video data.

The experimental results were encouraging. However, there are still several open issues, as follows.

- Both the POCS-based and Bayesian-based solutions are computationally expensive. For software implementations, fast, but perhaps suboptimal, solutions need to be investigated. For hardware implementations, algorithm parallelization issues also need to be examined.
- All super-resolution methods, including ours, require accurate motion estimates. However, for a typical video sequence, we will almost surely have inaccurate motion estimates for some frames or regions because of the ill-posed nature of motion estimation. We need to deal with these model failure regions for successful application of

the proposed super-resolution algorithms to general video sequences.

- Here, we demonstrated the algorithm for a limited set of parameters, which are probably not optimal. The algorithm enables distinct treatment of each DCT coefficient. This can be exploited for better performance. For instance, the information coming from high-frequency DCT coefficients can be discarded since they are quantized severely, and the information obtained from those coefficients are likely to be noise. Again, more thorough analysis needs to be done.
- The stochastic entities in the problem were assumed to be Gaussian random processes. Different statistical models need to be investigated. Even if they do not have analytical solutions, improved reconstruction may be achievable.

REFERENCES

[1] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*. Greenwich, CT: JAI, 1984.

[2] 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.

[3] S. P. Kim and W. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Processing*, vol. 2, pp. 534–539, Oct. 1993.

[4] B. C. Tom, A. K. Katsaggelos, and N. P. Galatsanos, "Reconstruction of a high resolution image from registration and restoration of low resolution images," in *Proc. IEEE Int. Conf. Image Processing*, Nov. 1994, pp. 13–16.

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

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

[7] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. Amer. A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[8] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Mar. 1992, pp. 169–172.

[9] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, pp. 1064–1076, Aug. 1997.

[10] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 231–239, May 1991.

[11] S. Mann and R. W. Picard, "Virtual bellows: Constructing high quality stills from video," in *IEEE Int. Conf. Image Processing*, Nov. 1994, pp. 13–16.

[12] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, pp. 324–335, Dec. 1993.

[13] T. Komatsu, K. Aizawa, and T. Saito, "Very high resolution imaging scheme with multiple different-aperature cameras," *Signal Process.: Image Commun.*, vol. 5, pp. 511–526, Dec. 1993.

[14] P. Cheeseman, B. Kanefsky, and R. Hanson, "Super-Resolved Surface Reconstruction from Multiple Images," NASA Rep., 1993.

[15] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, pp. 996–1011, June 1996.

[16] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec. 1997.

[17] S. Borman and R. L. Stevenson, "Simultaneous multi-frame map super-resolution video enhancement using spatio-temporal priors," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 1999, pp. 469–473.

[18] D. Chen and R. R. Schultz, "Extraction of high-resolution video stills from MPEG image sequences," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, 1998, pp. 465–469.

[19] Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from MPEG coded video," *IEEE Trans. Circuits, Syst., Video Technol.*, vol. 12, no. 4, pp. 217–226, 2002.

[20] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2001, pp. 185–188.

[21] ——, "Restoration of compressed video using temporal information," *Proc. SPIE*, vol. 4310, pp. 21–29, 2001.

[22] C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Reconstruction of high-resolution image frames from a sequence of low-resolution and compressed observations," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, 2002, pp. 1701–1704.

[23] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Multiframe resolution-enhancement methods for compressed video," *IEEE Signal Processing Lett.*, vol. 9, pp. 170–174, June 2002.

[24] A. J. Patti and Y. Altunbasak, "Artifact reduction for set theoretic superresolution image reconstruction with edge adaptive constraints and higher-order interpolants," *IEEE Trans. Image Processing*, vol. 10, pp. 179–186, Jan. 2001.

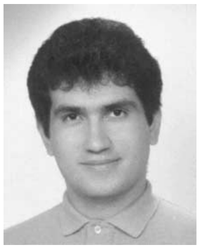
[25] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[26] C. J. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, 1988, pp. 147–151.

Bahadir K. Gunturk (S'01) received the B.S. degree from Bilkent University, Ankara, Turkey, in 1999, and the M.S. and Ph.D. degrees from Georgia Institute of Technology, Atlanta, in 2001 and 2003.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Louisiana State University. His research interests include image processing, multimedia communications, and computer vision. He spent the summer of 2002 as an Intern in the Imaging Science Division of the Eastman Kodak Research Laboratories, Rochester, NY.

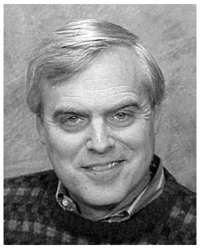
Dr. Gunturk received the Outstanding Research Award from the Center for Signal and Image Processing, Georgia Institute of Technology, in 2001.



Yucel Altunbasak (S'94–M'97–SM'01) received the B.S. degree from Bilkent University, Ankara, Turkey, in 1992 with highest honors. He received the M.S. and Ph.D. degrees from the University of Rochester, Rochester, NY, in 1993 and 1996, respectively.

He joined Hewlett-Packard Research Laboratories (HPL), Palo Alto, CA, in July 1996. His position at HPL provided him with the opportunity to work on a diverse set of research topics, such as video processing, coding and communications, multimedia streaming, and networking. He also taught digital video and signal processing courses at Stanford University, Stanford, CA, and San Jose State University, San Jose, CA, as a Consulting Assistant Professor. He joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, in 1999 as an Assistant Professor. He is currently working on industrial- and government-sponsored projects related to video and multimedia signal processing, inverse problems in imaging, and network distribution of compressed multimedia content. His research efforts resulted in over 75 publications and 12 patents/patent applications.

Dr. Altunbasak is an area/associate editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, *Signal Processing: Image Communications*, and the *Journal of Circuits, Systems and Signal Processing*. He is a member of the IEEE Signal Processing Society's IMDSP Technical Committee. He serves as a co-chair for Advanced Signal Processing for Communications Symposia at ICC'03. He also serves as a session chair in technical conferences, as a panel reviewer for government funding agencies, and as a technical reviewer for various journals and conferences in the field of signal processing and communications. He received the National Science Foundation (NSF) CAREER Award in 2002.



Russell M. Mersereau (F'83) received the S.B. and S.M. degrees in 1969 and the Sc.D. degree in 1973 from the Massachusetts Institute of Technology, Cambridge.

He joined the School of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, in 1975. His current research interests are in the development of algorithms for the enhancement, modeling, and coding of computerized images, synthesis aperture radar, and computer vision. In the past, this research has been directed to problems of distorted signals from partial information of those signals, computer image processing and coding, the effect of image coders on human perception of images, and applications of digital signal processing methods in speech processing, digital communications, and pattern recognition. He is the coauthor of the text *Multidimensional Digital Signal Processing*.

Dr. Mersereau has served on the editorial board of the PROCEEDINGS OF THE IEEE and as Associate Editor for Signal Processing of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, and the IEEE SIGNAL PROCESSING LETTERS. He is the corecipient of the 1976 Bowder J. Thompson Memorial Prize of the IEEE for the best technical paper by an author under the age of 30, a recipient of the 1977 Research Unit Award of the Southeastern Section of the ASEE, and three teaching awards. He was awarded the 1990 Society Award of the Signal Processing Society. He is currently the Vice President for Awards and Membership of the Signal Processing Society.