Multiframe Resolution-Enhancement Methods for Compressed Video

B. K. Gunturk, Yucel Altunbasak, *Senior Member, IEEE*, and R. M. Mersereau, *Fellow, IEEE*

***Abstract*—Multiframe resolution enhancement ("superresolution") methods are becoming widely studied, but only a few procedures have been developed to work with compressed video, despite the fact that compression is a standard component of most image- and video-processing applications. One of these methods uses quantization-bound information to define convex sets and then employs a technique called "projections onto convex sets" (POCS) to estimate the original image. Another uses a discrete cosine transformation (DCT)-domain Bayesian estimator to enhance resolution in the presence of both quantization and additive noise. The latter approach is also capable of incorporating known source statistics and other reconstruction constraints to impose blocking artifact reduction and edge enhancement as part of the solution. In this article we propose a spatial-domain Bayesian estimator that has advantages over both of these approaches.**

***Index Terms*—High-resolution video, maximum a posteriori probability (MAP), Motion Pictures Experts Group (MPEG), multiframe restoration, projections onto convex sets (POCS), resolution enhancement, superresolution (SR), video quality.**

I. Introduction

In the process of recording an image, there is a natural loss of spatial resolution caused by the nonzero physical dimensions of the individual sensor elements, the nonzero aperture time, optical blurring, noise, and motion. Multiframe resolution enhancement ("superresolution") techniques try to estimate the high-resolution image by combining the nonredundant information that is available into a sequence of low-resolution images.

Superresolution (SR) reconstruction has many application areas including enhancing the imagery for high-definition television (HDTV) sets, extracting still images from a video source for printing, medical imaging, aerial and satellite imaging, remote sensing, surveillance systems, forensic science, and digital cameras. While many methods have been proposed to enhance raw video, only a few have been proposed to operate for the compressed video. Of course, any algorithm that enhances uncompressed-video algorithms can be used with compressed video by first decompressing the material, but this process necessarily discards important information about the quantization of the high-resolution imagery that was introduced by the act of compression itself. Patti and Altunbasak [1], [2] demonstrated the importance of properly handling this quantization information and suggested a solution for enhancing the video signal that explicitly exploits the compression process. This method models the video acquisition and compression processes and uses the quantization information as the basis for a projections onto convex sets (POCS)-based algorithm that operates in the compressed domain. However, with this approach all sources of error except for the quantization error are ignored, which may not be appropriate at medium to high bit rates. The POCS approach is also unable to impose additional constraints on the reconstructed image. In [3], spatial-domain additive noise is modeled and transformed to the compressed domain to establish a stochastic framework that can utilize the quantization information. It is also possible to develop a Bayesian reconstruction algorithm that seeks to minimize the artifacts produced by the compression process [4], [5].

This article briefly summarizes these methods and proposes to transform quantization noise statistics into the spatial domain. An algorithm based on iterated conditional modes is implemented, and the results are compared with the POCS-based algorithm.

II. Superresolution Techniques That Model Compression

All SR techniques model the video acquisition process that relates high-resolution imagery to observations (low-resolution frames or quantized discrete cosine transformation (DCT) coefficients) and attempt to solve this inverse problem by using the model and the observations. The video acquisition process can be formulated as

$$g_d(\mathbf{l},k)=\sum_{\mathbf{n}} f(\mathbf{n},t_r)h(\mathbf{n},t_r;\mathbf{l},k)+v(\mathbf{l},k) \qquad (1)$$

where $h(\mathbf{n},t_r;\mathbf{l},k)$ is the discrete, linear shift-varying (LSV) blur mapping between the discrete high-resolution (HR) image $f(\mathbf{n},t_r)$ at a reference time t_r and the kth discrete low-resolution (LR) image $g_d(\mathbf{l},k)$ [2], [6]. $v(\mathbf{l},k)$ is an additive noise that is usually assumed to be from a white, Gaussian, discrete random process. As depicted in Fig. 1, the compression stages perform motion compensation of the LR frames $g_d(\mathbf{l},k)$ and follow this by a series of 8×8 block-DCTs to produce the DCT

Manuscript received May 10, 2001; revised March 22, 2002. This work was supported in part by the Office of Naval Research (ONR) under Award N000140110619 and the National Science Foundation (NSF) under Award CCR-0113681. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gaetano Scarano.

The authors are with the Center for Signal and Image Processing, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: bahadir@ece.gatech.edu; rmm@ece.gatech.edu; yucel@ece.gatech.edu).

Publisher Item Identifier 10.1109/LSP.2002.800503.

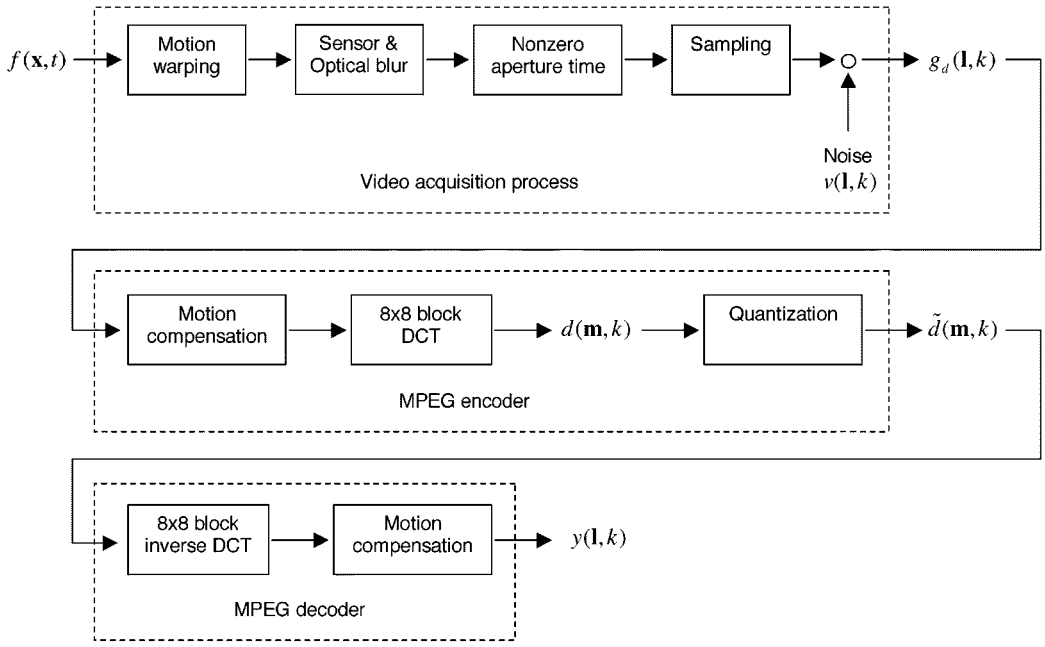

Fig. 1. Video acquisition model and encoder/decoder stages.

coefficients $d(\mathbf{m},k)$. Defining $\hat{g}(\mathbf{l},k)$ as the prediction frame and $\mathcal{DCT}\{\cdot\}$ as the DCT operator, we can write

$$\begin{aligned} d(\mathbf{m},k) &= \mathcal{DCT}\left\{g_d(\mathbf{l},k) - \hat{g}(\mathbf{l},k)\right\} \\ &= \sum_{\mathbf{n}} f(\mathbf{n},t_r) h_{\mathrm{DCT}}(\mathbf{n},t_r;\mathbf{m},k) \\ &\quad - \hat{G}(\mathbf{m},k) + V(\mathbf{m},k) \end{aligned} \tag{2}$$

where $\hat{G}$ and V are the block-DCTs of $\hat{g}$ and v, respectively. $h_{DCT}(\mathbf{n},t_r;\mathbf{m},k)$ is the 8×8 block-DCT of $h(\mathbf{n},t_r;\mathbf{l},k)$ [2]. The DCT coefficients $d(\mathbf{m},k)$ are then quantized to produce the quantized DCT coefficients $\tilde{d}(\mathbf{m},k)$. The quantization operation is a nonlinear process that will be denoted by the operator $\mathcal{Q}\{\cdot\}$. Defining $G(\mathbf{m},k) \triangleq \sum_{\mathbf{n}} f(\mathbf{n},t_r) h_{DCT}(\mathbf{n},t_r;\mathbf{m},k)$, we write

$$\begin{aligned} \tilde{d}(\mathbf{m},k) &= \mathcal{Q}\left\{d(\mathbf{m},k)\right\} \\ &= \mathcal{Q}\left\{G(\mathbf{m},k) - \hat{G}(\mathbf{m},k) + V(\mathbf{m},k)\right\}. \end{aligned} \tag{3}$$

In popular video compression schemes, quantization is realized by dividing each DCT coefficient by a quantization step size followed by rounding to the nearest integer. The quantized DCT coefficients $\tilde{d}(\mathbf{m},k)$ and the corresponding step sizes are available at the decoder, i.e., are either embedded in the compressed bit-stream or specified as part of the coding standard.

A. DCT-Domain Projections Onto Convex Sets (POCS) Method

Since the quantization step size is known for each DCT coefficient, upper and lower bounds within which the DCT coefficients $d(\mathbf{m},k)$ of the reconstructed HR image lie can be determined from the observations $\tilde{d}(\mathbf{m},k)$. These bounds form a constraint set for each DCT coefficient. The video acquisition and compression stages [i.e., the model given by (2)] are applied to an initial HR image estimate to compute its DCT coefficients. If the computed DCT coefficients lie outside the bounds that are derived from the MPEG bitstream, the error is back projected onto the initial HR estimate so as to enforce consistency with these bounds. SR reconstruction is achieved by repeating this procedure iteratively for all observation sets. The details of this method can be found in [1], [2].

B. DCT-Domain Maximum A Posteriori Probability (MAP) Estimator

The DCT-domain POCS method performs impressively, especially when the compression ratio is high and quantization is the major source of error. However, it ignores all other sources of error, which can become important at lower compression ratios (higher bit rates). Bayesian estimators are capable of partially compensating for these error sources, which are generally modeled as additive noise (see Fig. 1). In addition, with the use of Bayesian estimators, the source statistics as well as additional constraints that affect blocking artifact reduction and edge enhancement can also be included in the reconstruction. Addressing these issues, it is possible to formulate a MAP estimator that uses statistical models in the DCT domain [3]:

$$\begin{aligned} \hat{f}(\mathbf{n},t_r) = \arg\max_{f(\mathbf{n},t_r)} \Big\{ & p_{\tilde{d}(\mathbf{m},k_1),\ldots,\tilde{d}(\mathbf{m},k_p)|f(\mathbf{n},t_r)}(\cdot) \\ & \cdot p_{f(\mathbf{n},t_r)}(\cdot)\Big\}. \end{aligned} \tag{4}$$

In [3], an analytical formula for the conditional probablitiy density function (PDF) $p_{\tilde{d}(\mathbf{m},k_1),\ldots,\tilde{d}(\mathbf{m},k_p)|f(\mathbf{n},t_r)}(\cdot)$ is given. It is also possible to model the conditional PDF differently under some other assumptions [7]–[9].

C. Spatial-Domain Maximum A Posteriori Probability (MAP) Estimator

Spatial-domain SR reconstruction techniques are generally less computationally complex than DCT-domain techniques. This is because of the fact that each pixel in the spatial domain maps to 64 DCT coefficients when an 8×8 block-DCT is taken. This increases the blur mapping size significantly, which has a direct effect on the computational cost.

In this section, we propose to use quantization error statistics in a spatial-domain MAP estimation framework. The spatial-domain LR frames, $y(\mathbf{l}, k)$, are obtained by taking the inverse DCT of the quantized DCT coefficients and then motion-compensating the residual frames (see Fig. 1). Using (3), this can be formulated as

$$\begin{aligned} y(\mathbf{l},k) &= \mathcal{IDCT}\left\{\tilde{d}(\mathbf{m},k)\right\} + \hat{g}(\mathbf{l},k) \\ &= \mathcal{IDCT}\left\{\mathcal{Q}\left\{G(\mathbf{m},k) - \hat{G}(\mathbf{m},k) + V(\mathbf{m},k)\right\}\right\} \\ &\quad + \hat{g}(\mathbf{l},k) \end{aligned} \tag{5}$$

where $\mathcal{IDCT}(\cdot)$ is the 8×8 block inverse DCT operator. The MAP estimator can be written as

$$\begin{aligned} &\hat{f}(\mathbf{n},t_r) \\ &\quad = \arg\max_{f(\mathbf{n},t_r)} \left\{p_{y(\mathbf{l},k_1),\ldots,y(\mathbf{l},k_p)|f(\mathbf{n},t_r)}(\cdot)\, p_{f(\mathbf{n},t_r)}(\cdot)\right\}. \end{aligned} \tag{6}$$

Equation (6) shows that the conditional PDF $p_{y(\mathbf{l},k_1),\ldots,y(\mathbf{l},k_p)|f(\mathbf{n},t_r)}(\cdot)$ and the prior image PDF $p_{f(\mathbf{n},t_r)}(\cdot)$ have to be known for MAP estimation. Simple Gaussian models or more complicated ones can be assumed for the prior image PDF $p_{f(\mathbf{n},t_r)}(\cdot)$ (see [11]). Unfortunately, there is no closed-form solution for the conditional PDF $p_{y(\mathbf{l},k_1),\ldots,y(\mathbf{l},k_p)|f(\mathbf{n},t_r)}(\cdot)$, since the $\mathcal{IDCT}(\cdot)$ and $\mathcal{Q}(\cdot)$ operators [in (5)] do not commute. However, this distribution can be computed experimentally and saved in a lookup table. It is also possible to fit a model to the distribution. This can be done once and used for all reconstructions afterward.

One way of computing the conditional PDF $p_{y(\mathbf{l},k_1),\ldots,y(\mathbf{l},k_p)|f(\mathbf{n},t_r)}(\cdot)$ is to assume that the noise is an independent identically distributed additive white Gaussian process, which results in an analytical formula for the error PDF at the quantizer output [3]. This error PDF is the convolution of a Gaussian PDF (coming from the additive noise) with a uniform PDF (coming from the quantization process). Since the error PDFs at different DCT coefficients are independent of each other, the distribution of the error PDFs after the IDCT can be found without difficulty. Another analysis given in [7] suggests that for DCT coefficients that are observed to be zero, the Laplacian model fits well for the quantization error; for the nonzero-quantized DCT coefficients, the uniform model is better.

To demonstrate the effectiveness of using statistical models, the following experiment was conducted. The aerial image shown in Fig. 2(a) is jittered to have multiple frames that are slightly different than each other. The motion vectors are saved for use in reconstruction. Each frame is then blurred with a Gaussian low-pass filter having a support of five pixels and a variance of 1. The filtered frames are than downsampled by four and quantized using the ISO MPEG-2 intraframe quantization matrix with the quantizer scale parameter set to 0.5. One of the quantized frames that is bilinearly interpolated is given in Fig. 2(b). We then applied the spatial-domain POCS method proposed in [10]. Sixteen images are used in reconstruction, and after two iterations the image in Fig. 2(c) is obtained. We also applied the spatial-domain MAP estimator. For the conditional PDF, we assumed a Gaussian distribution. Letting $\mathcal{H}(\mathbf{l}, k; \mathbf{n}, t_r)$ be the operator applying the blocks given in Fig. 1 on $f(\mathbf{n}, t_r)$, the conditional PDF is proportional to $\exp\{-\|y(\mathbf{l},k) - \mathcal{H}(\mathbf{l},k;\mathbf{n},t_r) f(\mathbf{n},t_r)\|^2\}$. For the prior PDF, we assumed a local conditional PDF based on Markov Random Fields. It penalizes the difference between a pixel intensity and an estimate value that is obtained by averaging the intensities of its four neighbor pixels. That is, the local conditional PDF is proportional to $\exp\{-\|f(\mathbf{n},t_r) - (1/4)\sum_{\mathbf{i}\in\mathcal{N}} f(\mathbf{i},t_r)\|^2\}$, where $\mathcal{N}$ represents the four neighbors of the pixel at $(\mathbf{n}, t_r)$.

A suboptimal iterated conditional modes (ICM) implementation of the algorithms updates each pixel intensity iteratively by minimizing the weighted sum of the exponents

$$\begin{aligned} \hat{f}(\mathbf{n},t_r) = \arg\min_{f(\mathbf{n},t_r)} \Bigg\{ &\alpha \left\| y(\mathbf{l},k) - \mathcal{H}(\mathbf{l},k;\mathbf{n},t_r)\, f(\mathbf{n},t_r)\right\|^2 \\ &+ (1-\alpha)\left\| f(\mathbf{n},t_r) - \frac{1}{4}\sum_{\mathbf{i}\in\mathcal{N}} f(\mathbf{i},t_r)\right\|^2 \Bigg\} \end{aligned} \tag{7}$$

where α determines the relative contribution of the conditional and prior PDFs. One way of implementation is as follows:

1. Choose a reference frame from the video sequence; bilinearly interpolate it to obtain an initial estimate.
2. Estimate the motion between the reference and other frames and compute $\mathcal{H}(\mathbf{l}, k; \mathbf{n}, t_r)$
3. Repeat the following until a stopping criterion is reached:
   (a) Choose a low-resolution observation.
   (b) For each pixel in the low-resolution observations $y(\mathbf{l}, k)$,
      i. Determine the set of pixels in $f(\mathbf{n}, t_r)$ that affects the particular pixel in the low-resolution image through the mapping $\mathcal{H}(\mathbf{l}, k; \mathbf{n}, t_r)$.
      ii. Keep all the pixels intensities except for one pixel in the given set fixed and find the intensity that minimizes the sum given in **(7)**.
      iii. Update the pixel intensity and repeat for the other pixels in the set.
   (c) Choose another low-resolution observation and go to Step (b).

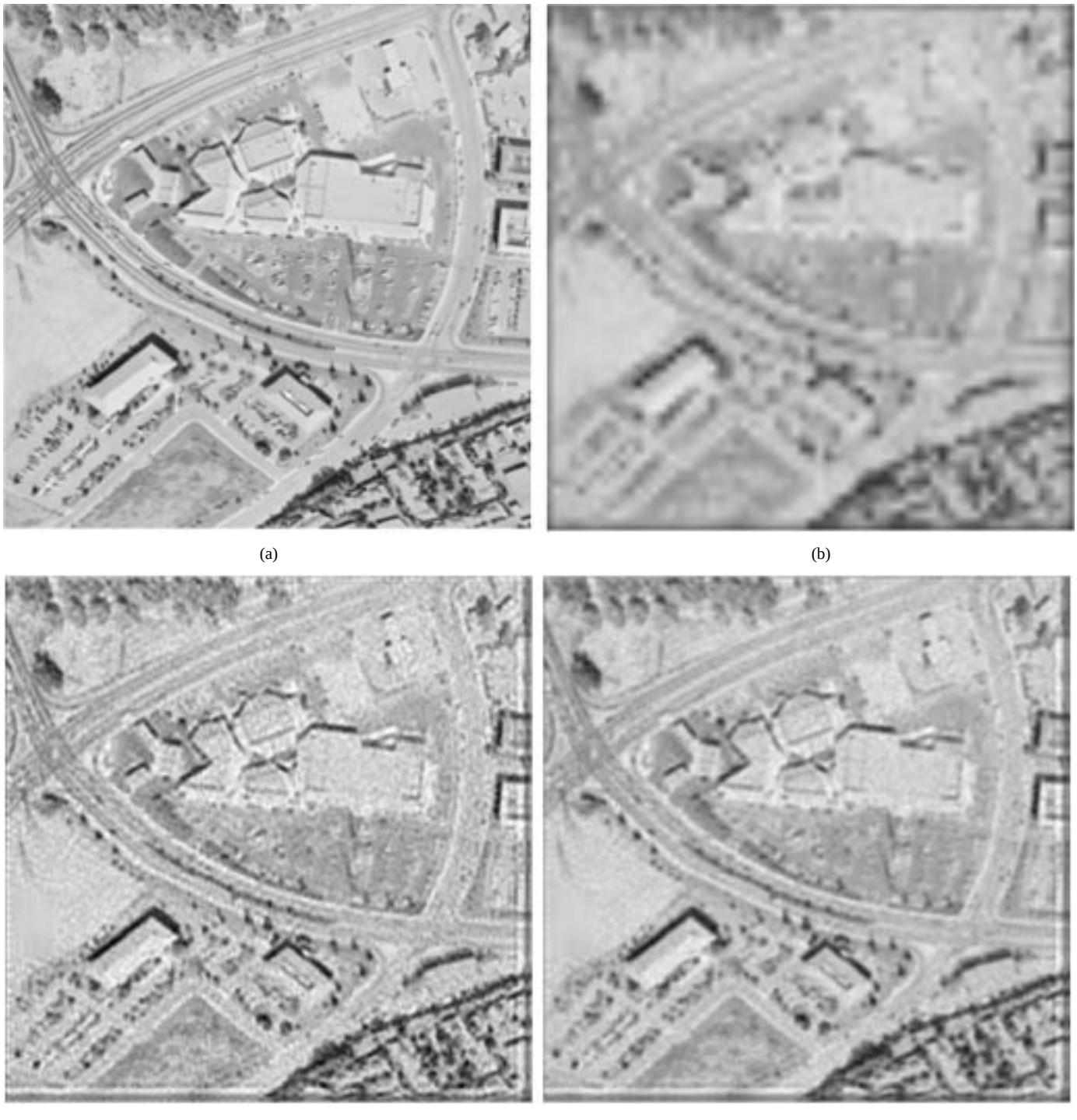

Fig. 2. A comparison of POCS and Bayesian reconstructions for compressed video. (a) Original image. (b) Bilinearly interpolated observation. (c) POCS reconstruction. (d) MAP reconstruction.

In the implementation, we set α to 0.65 and the number of iterations to 2. For this particular choice of PDFs, there is a closed-form solution for Step 3(b)ii in the above algorithm. The same set of low-resolution observations is used as in the POCS-based reconstruction. The reconstructed image is given in Fig. 2(d). When compared to the result of the POCS-based implementation, the reconstructed image of the MAP implementation is smoother and much more artifact-free.

III. Conclusion

As mentioned earlier, there are two main SR reconstruction approaches that were designed for compressed video. One is based on the POCS technique, and it enforces the consistency of the solution with the quantization-bound information. The other approach is based on the MAP estimation. In addition to the prior information about the solution, MAP-based methods can utilize the quantization statistics in either the transform or spa-

tial domain. In the experiments, we assumed a Gaussian model for the conditional PDF, and a local Markov Random Fields model for the prior PDF. Although these are not necessarily the best models, the MAP-based reconstruction produced a more artifact-free and smoother solution than the POCS-based reconstruction did.

REFERENCES

[1] Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from MPEG-coded video," *IEEE Trans. Circuits Syst. Vid. Technol.*, vol. 12, pp. 217–226, Apr. 2002.

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

[3] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Bayesian resolution-enhancement framework for transform-coded video," in *Proc. IEEE ICIP*, vol. 2, 2001, p. 41.

[4] J. Mateos, A. K. Katsaggelos, and R. Molina, "Resolution enhancement of compressed low resolution video," in *Proc. IEEE ICASSP*, vol. 4, 2000, pp. 1919–1922.

[5] C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian high-resolution reconstruction of low-resolution compressed video," in *Proc. IEEE ICIP*, vol. 2, 2001, pp. 25–28.

[6] 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.

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

[8] G. Lakhani, "Distribution-based restoration of DCT coefficients," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 819–823, Aug. 2000.

[9] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Processing*, vol. 9, pp. 1661–1666, Aug. 2000.

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

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