Bayesian Resolution Enhancement of Compressed Video

C. Andrew Segall, Aggelos K. Katsaggelos, *Fellow, IEEE*, Rafael Molina, and Javier Mateos

***Abstract*—Super-resolution algorithms recover high-frequency information from a sequence of low-resolution observations. In this paper, we consider the impact of video compression on the super-resolution task. Hybrid motion-compensation and transform coding schemes are the focus, as these methods provide observations of the underlying displacement values as well as a variable noise process. We utilize the Bayesian framework to incorporate this information and fuse the super-resolution and post-processing problems. A tractable solution is defined, and relationships between algorithm parameters and information in the compressed bitstream are established. The association between resolution recovery and compression ratio is also explored. Simulations illustrate the performance of the procedure with both synthetic and nonsynthetic sequences.**

***Index Terms*—Image coding, image restoration, post-processing, resolution enhancement, super-resolution, video coding.**

I. Introduction

IMPROVING the resolution of an image impacts a wide variety of applications. For example, high-resolution imagery often enhances the precision of scientific, medical and space imaging systems; improves the robustness of image analysis and tracking tasks, and benefits consumer electronic and entertainment applications. In each of these systems, acquiring data with a high-resolution sensor is one method to increase im age resolution. This makes resolution improvement straightforward. Unfortunately though, many applications cannot afford the increased system and transmission complexity required by high-resolution data acquisition. For these tasks, one must turn to algorithmic techniques for increasing resolution.

Super-resolution algorithms increase the resolution of an image without changing the resolution of the image sensor. This is accomplished by exploiting the underlying motion of a video sequence to provide multiple observations for each frame, and it mitigates the requirements for transporting and storing a high-resolution sequence. An accurate system model is the key to the super-resolution approach. Typical models consider the low-resolution sensor as a succession of filtering and sampling operations and assume that the image sequence is corrupted by additive noise during the acquisition process. (The noise also accounts for occlusions within the scene and uncertainties in the motion.) Until recently, this model described a majority of image acquisition tasks. However, with the increased use of video compression prior to digital transmission and storage, such an acquisition model is no longer adequate. Instead, novel algorithms must be developed that exploit the information available in the compressed bitstream.

The rest of this paper develops a super-resolution algorithm for compressed video. Estimating high-resolution video from a sequence of low-resolution and compressed observations is the focus, and we are interested in hybrid motion-compensation and transform coding methods, such as the MPEG and ITU family of standards [1]–[6]. These compression systems introduce several disparities to the super-resolution approach. As a first deviation, the low-resolution observations are no longer a sequence of intensity images. Instead, the compressed bitstream serves as the input to the super-resolution algorithm. This bitstream describes the original image sequence as a combination of quantized transform coefficients and motion vectors, and it introduces other departures into the super-resolution problem. For example, noise introduced by the quantization operator degrades the low-resolution images with a frequency and spatially varying noise process. Furthermore, the structure of the encoder introduces a variety of coding errors such as blocking, ringing, and temporal flicker. As a final difference, motion vectors are present in the bitstream. These vectors provide a noisy observation of the subpixel displacement within the image sequence.

The paper is organized as follows. In the next section, we present background for the super-resolution problem. This includes a discussion of super-resolution as well as post-processing methods. In Section III, we define a system model that contains a compression process. In Section IV, we propose a problem formulation for the super-resolution of compressed video. The formulation relies on the Bayesian framework, and it incorporates both the transform coefficients and motion vectors from the compressed bitstream. In Section V, we describe a realization of the super-resolution algorithm. In Section VI, we illustrate the efficacy of the proposed approach through simulation. Finally, we discuss conclusions and future work in Section VII.

II. Background

To be successful, a resolution-enhancement algorithm requires that a low-resolution image contain additional information about the original high-resolution scene. This is first

Manuscript received February 4, 2002; revised November 19, 2003. This work was supported in part by the "Comisión Nacional de Ciencia y Tecnología" under contracts TIC2000-1275 and TIC2003-00880. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Tamas Sziranyi.

C. A. Segall and A. K. Katsaggelos are with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 USA (e-mail: asegall@ece.nwu.edu; andrew@pixonics.com; aggk@ece.nwu.edu; aggk@ece.northwestern.edu).

R. Molina and J. Mateos are with the Departamento de Ciencias de la Computación e I. A., Universidad de Granada, 18071 Granada, Spain (e-mail: rms@decsai.ugr.es; jmd@decsai.ugr.es)

Digital Object Identifier 10.1109/TIP.2004.827230

considered in [7], where a set of still images is addressed. These images differ from one another by a single displacement vector, and the super-resolution algorithm solves the necessary problems of registration and interpolation. The approach is subsequently extended in [8], where the high-resolution images are filtered before sampling and corrupted by noise. This incorporates a restoration component into the algorithm, as the blur operator must be considered. Several other methods for the super-resolution of a set of still images have also been suggested. In [9]–[11], the tasks of interpolation and restoration are considered, while the registration problem is treated separately. In [12]–[14], methods are proposed that consider the degradations concurrently. For a complete review of resolution enhancement methods, see [15] and the references therein.

Extending the super-resolution approach to video sequences places greater emphasis on the registration process. This is necessary since objects move during acquisition, which makes the displacement between two images in the sequence spatially varying. One method for addressing the problem is to estimate the displacements and high-resolution data independently. For example, an estimate for the motion field is developed and utilized for resolution enhancement in [16]; motion is estimated with a hierarchical block-matching algorithm and followed by a maximum *a posteriori* estimate for the super-resolved image in [17], and a projection onto convex sets methodology that assumes a pre-computed motion estimate is developed in [18]. Recognizing that the displacement estimates often limit the super-resolution algorithm, information about the accuracy of the estimates is incorporated into the super-resolution problem in [19]–[21].

In this paper, we integrate yet another task into the super-resolution procedure. Low-resolution image sequences are compressed prior to resolution enhancement, which introduces additional artifacts into the observed images. Algorithms that attenuate these coding errors belong to the field of post-processing. As an example, filtering a decoded image with a spatially invariant kernel is proposed in [22]. This enhancement technique removes blocking errors. Unfortunately though, it also attenuates semantically meaningful edge features. Addressing this flaw is the goal of many subsequent enhancement approaches, such as [23]–[25], that reduce the amount of smoothing in the vicinity of edges.

Recovery methods can also be utilized for post-processing. They lead to a more rigorous approach to the problem, as prior knowledge of both the original image and compression system are considered. In [26], this information is derived numerically, and the recovery procedure takes the form of a table lookup operation. In the majority of approaches though, explicit models are utilized. For example, distributions for the transform coefficients and original intensities are defined in [27]–[29]. These distributions lead to a maximum *a posteriori* estimate for the post-processed image. In [30] and [31], information is expressed by sets of valid solutions. As the sets are convex, the theory of projection onto convex sets provides the necessary solution. As a third approach, constrained least-squares solutions incorporate deterministic models and are considered in [32]–[34].

With a large body of work devoted to both post-processing and super-resolution methods, it seems wise to combine the two techniques and solve the super-resolution of compressed video problem. This has been pursued in various forms. For example, in [35] and [36], the quantization operator is incorporated into a super-resolution procedure. The resulting algorithms consider the spatially varying noise process and treat the interpolation, restoration and post-processing problems. All necessary displacement values are assumed known. As a second example, motion vectors within the bitstream influence the registration problem in [17] and [37]. The post-processing problem is ignored though, and estimates for the motion and high-resolution data are computed sequentially.

In this paper, we build on previous post-processing and super-resolution work [38]–[43] and develop a novel algorithm that solves the registration, interpolation, restoration, and post-processing problem simultaneously. We utilize the Bayesian framework to incorporate information from the compressed bitstream, knowledge of the encoder structure and models for the high-resolution intensities and displacement values. This is especially appealing with hybrid motion-compensation and transform coding methods, as the compressed bitstream contains information about both the down-sampled intensities and the inter-frame displacements.

III. System Model

Images that are captured in rapid succession typically contain similar image content. That is, we can model the image sequence as

$$f_l(x,y) = f_k\left(x + d^x_{l,k}(x,y), y + d^y_{l,k}(x,y)\right) + n^r_{l,k}(x,y) \tag{1}$$

where $f_l\,(x,y)$ and $f_k\,(x,y)$ are spatial locations in the high-resolution images at times l and k, respectively, $d^x_{l,k}(x,y)$ and $d^y_{l,k}(x,y)$ denote the x and y components of the displacement that relates the pixel at time k to the pixel at time l, and $n^r_{l,k}(x,y)$ is an additive noise process that accounts for image locations that are poorly described by the displacement model.

The relationship in (1) is also expressed in matrix-vector form as

$$\mathbf{f}_l = \mathrm{C}(\mathbf{d}_{l,k})\mathbf{f}_k + \mathbf{n}^r_{l,k} \tag{2}$$

where vectors $\mathbf{f}_l$ and $\mathbf{f}_k$ are formed by lexicographically ordering each image frame, $\mathrm{C}(\mathbf{d}_{l,k})$ is the two-dimensional matrix that describes the displacement across the entire frame, $\mathbf{d}_{l,k}$ is the column vector defined by lexicographically ordering the values $[d^x_{l,k}(x,y)\; d^y_{l,k}(x,y)]^{\mathrm{T}}$, and $\mathbf{n}^r_{l,k}$ is the registration noise process. When each image frame is of dimension $PM \times PN$, then $\mathbf{f}_l, \mathbf{f}_k$, and $\mathbf{n}^r_{l,k}$ are column vectors with length $PMPN$ and $\mathrm{C}(\mathbf{d}_{l,k})$ has dimension $PMPN \times PMPN$.

Many applications do not allow for the direct observation of the high-resolution frames in (1) and (2). Instead, only a low-resolution frame is available, which is related to the original high-resolution frame by

$$\mathbf{g}_k = \mathbf{A}\mathbf{H}\mathbf{f}_k + \mathbf{n}_k \tag{3}$$

where $\mathbf{g}_k$ is a vector that contains the low-resolution image with dimension $MN \times 1$, $\mathbf{f}_k$ is the high-resolution image, $\mathbf{n}_k$ is a vector that describes the acquisition noise, $\mathbf{A}$ is an $MN \times PMPN$ matrix that subsamples the high-resolution image, and $\mathbf{H}$ is an $PMPN \times PMPN$ matrix that filters the high-resolution image [44]. Combining (2) and (3), the relationship between any low-resolution observation and a frame in the high-resolution image sequence is defined as

$$\mathbf{g}_l = \mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k + \mathbf{n}_{l,k} \tag{4}$$

where $\mathbf{n}_{l,k}$ contains both the registration and acquisition errors.

In classical imaging scenarios, the noise appearing in (4) is the dominant noise in the low-resolution observation. However, this may not be the case when the sequence of low-resolution images is compressed before transmission. Compression reduces the bandwidth required for transmission, and it further degrades the observations. In this paper, we consider a system where compression is the dominant source of noise. This allows us to focus on integrating the compressed bitstream into a super-resolution algorithm. (Please note that the resulting procedure can be easily extended to systems with additional noise components, which we address in later sections.)

A. Video Compression

Hybrid motion compensation and transform based coding algorithms are common methods for compressing a sequence of images. In these techniques, which include all of the MPEG and ITU video coding standards, each image is first divided into equally sized blocks. Then, the blocks are encoded with one of two available methods. As a first approach, a linear transform such as the discrete cosine transform (DCT) is applied to the block. The transform coefficients are then quantized and efficiently transmitted to the decoder, where an estimate of the uncompressed image is generated. As a second coding approach, a prediction for the block is found in previously encoded frames. The location of the prediction is represented with a motion vector, which defines the spatial offset between the current block and its prediction, and is transmitted to the decoder. Computing the prediction error, transforming it with a linear transform, quantizing the transform coefficients, and transmitting the quantized information refine the prediction.

The compressed observation of the low-resolution frame is therefore expressed as

$$\mathbf{y}_k = \mathbf{T}^{-1}Q\left[\mathbf{T}\left(\mathbf{g}_k - \sum_{\forall i} C(\mathbf{v}_{k,i})\mathbf{y}_i\right)\right] + \sum_{\forall i} C(\mathbf{v}_{k,i})\mathbf{y}_i \tag{5}$$

where $\mathbf{y}_k$ is the vector that contains the decoded image for frame k, $\mathbf{v}_{k,i}$ is the vector that contains the transmitted motion vectors that predict frame k from previously compressed frame i, matrix $C(\mathbf{v}_{k,i})$ represents the prediction process with a matrix, $Q[\cdot]$ represents the quantization procedure, and $\mathbf{T}$ and $\mathbf{T}^{-1}$ are the forward and inverse-transform operations, respectively. Combining (4) and (5), we state the relationship between any low and high-resolution image as

$$\mathbf{y}_l = \mathbf{T}^{-1}Q\left[\mathbf{T}(\mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k - \mathbf{y}_l^{\mathrm{MV}} + \mathbf{n}_{l,k})\right] + \mathbf{y}_l^{\mathrm{MV}} \tag{6}$$

where the motion-compensated estimate defined by $\sum_{\forall i} C(\mathbf{v}_{k,i})\mathbf{y}_i$ is denoted as y_l^{MV} for notional convenience.

B. Noise Models

The quantization operator in (6) introduces compression noise into the decoded frames. These errors correspond to information discarded during quantization, and they describe a deterministic process. The resulting compression errors are stochastic though, since we are dealing with random quantities. Here, we express the compression noise in the spatial domain. Since the noise in the transform and spatial domains is related as

$$\mathbf{n}^{\mathrm{Spatial}} = \mathbf{T}^{-1}\mathbf{n}^{\mathrm{Transform}} \tag{7}$$

noise in the spatial domain is a linear sum of independent noise components. The resulting noise process then approaches the Gaussian distribution, as the Central Limit Theorem is satisfied.

Since we are assuming that the quantization noise is dominant, we approach the quantity

$$\mathbf{T}^{-1}Q\left[\mathbf{T}\left(\mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k - \mathbf{y}_l^{\mathrm{MV}} + \mathbf{n}_{l,k}\right)\right] \tag{8}$$

by

$$\mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k - \mathbf{y}_l^{\mathrm{MV}} + \boldsymbol{\varepsilon}_{Q,l} \tag{9}$$

where

$$\boldsymbol{\varepsilon}_{Q,l} \sim \mathrm{N}(0, \mathbf{K}_{Q,l}) \tag{10}$$

and $\mathbf{K}_{Q,l}$ is the covariance matrix of the quantization noise in the spatial domain at frame l. Combining (6) and (9), we then have

$$\mathbf{y}_l = \mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k + \boldsymbol{\varepsilon}_{Q,l}. \tag{11}$$

Defining the covariance matrix $\mathbf{K}_{Q,l}$ thus becomes the critical step in modeling the compression system. Since errors in the spatial domain are related to errors in the transform domain by the inverse-transform operation, we can express the needed covariance matrix as

$$\mathbf{K}_{Q,l} = \mathbf{T}^{-1}\mathbf{K}_{\mathrm{Transform},l}\mathbf{T}^{-1^T} \tag{12}$$

where $\mathbf{K}_{\mathrm{Transform},l}$ is the diagonal matrix containing the expected noise power for each transform coefficient. In order to estimate the variances, it is reasonable to assume an independent uniform distribution within each quantization level when the quantization step size is small. Furthermore, the uniform assumption also holds when the magnitude of the quantized transform coefficient is large. This is true since the distribution of the transform coefficients typically contain significant tails, which leads to a uniform distribution within the quantization intervals

distant from the mean. Whenever the assumption of a uniform distribution is viable, the noise variance σ_l^2 for transform index l is defined completely by the bitstream and expressed as

$$\sigma_l^2 = \frac{q_l^2}{12} \tag{13}$$

where q_l is the quantizer step-size.

Of course, other definitions for $\mathbf{K}_{\mathrm{Transform},l}$ could be constructed to address the inaccuracies of assuming a uniform distribution within the quantization interval. This requires additional information about the distribution of the original transform coefficients, and it may necessitate an additional estimation procedure. Besides improving the accuracy of the covariance matrix though, other reasons for modifying $\mathbf{K}_{\mathrm{Transform},l}$ are also envisioned. For example, if the noise introduced during compression is not the dominant noise process, then the compressed bitstream does not completely specify the covariance matrix. Instead, information about these other corrupting processes must also be included. This identifies how the noise model in (4) is incorporated.

Finally, it is interesting to note that when the diagonal entries of $\mathbf{K}_{\mathrm{Transform},l}$ are equal, the resulting covariance matrix describes an independent and identically distributed (IID) noise process in the transform domain. For the large class of linear transforms where $\mathbf{T}^{-1}$ is proportional to $\mathbf{T}^{\mathrm{T}}$ (such as the DCT), the noise in the spatial domain is also IID under these conditions. However, when the noise is not identically distributed in the transform domain (but still independent), it then becomes correlated in the spatial domain. In the case of standards based coding, both of these situations occur in practice. Intra-coded frames employ perceptually motivated quantization strategies. This leads to coarser quantization of the high-frequency transform coefficients and a $\mathbf{K}_{\mathrm{Transform},l}$ that is not IID. When transmitting the residual between the original frame and a motion compensated prediction though, the quantization strategy often utilizes the same quantization step size for each coefficient. The compression noise in this case is IID in both the transform and spatial domains.

In addition to the quantization intervals, information about the subpixel displacements also appears in the compressed bitstream. This data is encapsulated in the motion vectors that provide a noisy observation of the original displacements. The significance of these vectors is determined by several variables. For example, the intensity values of the low-resolution sequence have an impact. When these values describe significant image features, such as large-scale edges or corners, then the motion vector and actual subpixel displacement are often similar. When smooth image regions are considered though, the motion vector and actual displacement may vary significantly. As a second difference, encoders utilize motion vectors to increase the compression ratio. Thus, it may be advantageous to select a motion vector that poorly describes the current image if it reduces the bit-rate. The quality of the motion vectors is then apparent in the error residual transmitted to the decoder.

To model the accuracy of the motion vectors, we define the relationship between the motion compensated prediction and the original image frame as

$$\mathbf{y}_l^{\mathrm{MV}} = \mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k + \boldsymbol{\eta}_{\mathrm{MV},l} \tag{14}$$

where

$$\boldsymbol{\eta}_{Q,l} \sim \mathrm{N}(0, \mathbf{K}_{\mathrm{MV},l}). \tag{15}$$

The motivation for this model is that it encapsulates the influence of the original and compressed image frames on the significance of the motion vectors. Additionally, it makes the meaning of the covariance matrix $\mathbf{K}_{\mathrm{MV},l}$ intuitive. It is the covariance matrix for the displaced frame difference that is internal to the decoder.

Like the previous covariance matrix, an estimate for $\mathbf{K}_{\mathrm{MV},l}$ can be extracted from the compressed bitstream. This covariance matrix is found from the transmitted displaced frame difference, and the variance at each transform index l could be defined as

$$\sigma_l^2 = c_l^2 + \frac{q_l^2}{12} \tag{16}$$

where c_l is the transform coefficient decoded from the bitstream and q_l is the width of the quantization interval. The relationship in (12) then maps the variance information to the spatial domain. Other definitions could also be constructed. As with the covariance matrix $\mathbf{K}_{Q,l}$ though, when information about additional noise processes is available, it should also be incorporated into $\mathbf{K}_{\mathrm{MV},l}$. This includes the noise model in (4).

IV. Problem Formulation

The system model in (6) is used for the formulation of an algorithm that recovers high-resolution frames from a sequence of low-resolution and compressed frames. The approach is based on the fact that information about a single high-resolution frame appears in multiple low-resolution observations. When this information is not redundant, as provided with subpixel displacements in the mapping of frame $\mathbf{f}_k$ to $\mathbf{f}_l$, the introduction of aliasing by the sampling procedure $\mathbf{AH}$, and the preservation of aliasing during compression, then each observation provides additional information about the high-resolution image frame.

The Bayesian maximum *a posteriori* (MAP) estimate provides the appropriate framework for recovering high-resolution information from a sequence of compressed observations [43]. Since information about both the intensities and displacements in the original image sequence are present in the compressed bitstream, we advocate the joint estimate in (17), shown at the bottom of the page, where $\hat{\mathbf{f}}_k$ and

$$\begin{aligned}\hat{\mathbf{f}}_k, \hat{\mathbf{D}}_{\mathrm{TB,TF}} &= \arg\max_{\mathbf{f}_k, \mathbf{D}} \{p(\mathbf{f}_k, \mathbf{D}_{\mathrm{TB,TF}} \,|\, \mathbf{Y}_{\mathrm{TB,TF}}, \mathbf{V}_{\mathrm{TB,TF}})\} \\ &= \arg\max_{\mathbf{f}_k, \mathbf{D}} \left\{ \frac{p(\mathbf{Y}_{\mathrm{TB,TF}}, \mathbf{V}_{\mathrm{TB,TF}} \,|\, \mathbf{f}_k, \mathbf{D}_{\mathrm{TB,TF}}) p(\mathbf{f}_k, \mathbf{D}_{\mathrm{TB,TF}})}{p(\mathbf{Y}_{\mathrm{TB,TF}}, \mathbf{V}_{\mathrm{TB,TF}})} \right\}\end{aligned} \tag{17}$$

$\hat{\mathbf{D}}_{\mathrm{TB,TF}}$ are estimates of the high-resolution image and displacement field, respectively, $\mathbf{D}_{\mathrm{TB,TF}}$ is the matrix defined as $(\mathbf{d}_{k,k-\mathrm{TB}},\ldots,\mathbf{d}_{k,k+\mathrm{TF}})$, $\mathbf{Y}_{\mathrm{TB,TF}}$ is the matrix defined as $(\mathbf{y}_{k-\mathrm{TB}},\ldots,\mathbf{y}_{k+\mathrm{TF}})$, $\mathbf{V}_{\mathrm{TB,TF}}$ is the matrix defined as $(\mathbf{v}_{k-\mathrm{TB}},\ldots,\mathbf{v}_{k+\mathrm{TF}})$, $\mathbf{v}_k$ is the column vector $(\mathbf{v}_{k,0,}^T\ldots\mathbf{v}_{k,.}^T)^T$ that allows an encoding method to transmit multiple motion vectors for each block, TF and TB indicate the number of frames utilized by the recovery algorithm along the forward and backward directions of the temporal axis, and $\mathrm{p}(\cdot,\cdot)$ and $p(\cdot,\cdot\,|\,\cdot,\cdot)$ denote the joint and conditional probability density functions, respectively.

By observing that $p(\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{V}_{\mathrm{TB,TF}})$ does not depend on the optimization variables, and with the use of the monotonic log function, the MAP estimate becomes

$$\hat{\mathbf{f}}_k,\hat{\mathbf{D}}_{\mathrm{TB,TF}}=\arg\max_{\mathbf{f}_k,\mathbf{D}}\{\log p(\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{V}_{\mathrm{TB,TF}}\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}})+\log p(\mathbf{f}_k)+\log p(\mathbf{D}_{\mathrm{TB,TF}})\}\quad(18)$$

where $\mathbf{f}_k$ and $\mathbf{D}_{\mathrm{TB,TF}}$ are assumed to be independent.

V. Proposed Algorithm

Obtaining a frame with enhanced resolution according to (17) requires definitions of the probability density functions $p(\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{V}_{\mathrm{TB,TF}}\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}}),p(\mathbf{f}_k)$ and $p(\mathbf{D}_{\mathrm{TB,TF}})$. These functions incorporate information about the compression system, as well as *a priori* knowledge of the original high-resolution images sequence into the recovery framework. In this section, we propose models for the density functions that are applicable to a wide variety of coding scenarios. Moreover, we show that these models lead to a tractable method for solving (18).

A. Fidelity Constraints

The first density function in (17) defines the relationship between information in the compressed bitstream and the original high-resolution image sequence. We now take into account that

$$p(\mathbf{Y}_{\mathrm{TB,TF}}\mathbf{V}_{\mathrm{TB,TF}}\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}})=p(\mathbf{Y}_{\mathrm{TB,TF}}\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}})\cdot p(\mathbf{V}_{\mathrm{TB,TF}}\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}},\mathbf{Y}_{\mathrm{TB,TF}}).\quad(19)$$

The first conditional distribution on the right hand side of (19) can be approximated as follows. First, note that when $\mathbf{f}_k$ and $\mathbf{D}_{\mathrm{TB,TF}}$ are given, then the variables $\mathbf{y}_l$ can be assumed to be independent. Thus, we have

$$p(\mathbf{Y}_{\mathrm{TB,TF}}\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}})=\prod_l p(\mathbf{y}_l\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}})\quad(20)$$

and from (11), we have

$$p(\mathbf{y}_l\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}})\propto\exp\left\{-\frac{1}{2}(\mathbf{y}_l-\mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k)^T\times\mathbf{K}_{Q,l}^{-1}(\mathbf{y}_l-\mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k)\right\}\quad(21)$$

where $\mathbf{K}_{Q,l}$ is the covariance of the quantization noise in the spatial domain at frame l.

In order to now obtain $p(\mathbf{V}_{\mathrm{TB,TF}}\,|\,\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}},\mathbf{Y}_{\mathrm{TB,TF}})$ in (19), we note that when $\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}}$ and $\mathbf{Y}_{\mathrm{TB,TF}}$ are given, then (14) can be used to obtain $p(\mathbf{v}_l\,|\,\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{f}_k,\mathbf{d}_{l,k})$. To do so, we have to change variables from $C(\mathbf{v}_{l,k})\forall i$ to $\mathbf{v}_l$ and also remove $\mathbf{y}_i\forall i$ from the left hand side of (14). However, the Hessian of these changes does not depend on the variables $\mathbf{f}_k$ and $\mathbf{d}_{l,k}$ that we are trying to estimate. We then have

$$p(\mathbf{v}_l\,|\,\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{f}_k,\mathbf{d}_{l,k})\propto\exp\left\{-\frac{1}{2}(\mathbf{y}_l^{\mathrm{MV}}-\mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k)^T\times\mathbf{K}_{\mathrm{MV},l}^{-1}(\mathbf{y}_l^{\mathrm{MV}}-\mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k)\right\}.\quad(22)$$

We can assume that given $\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}}$, and $\mathbf{Y}_{\mathrm{TB,TF}}$, the variables $\mathbf{v}_i\forall l$ are independent, and so we write

$$p(\mathbf{V}_{\mathrm{TB,TF}}\,|\,\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{f}_k,\mathbf{D}_{\mathrm{TB,TF}})=\prod_l p(\mathbf{v}_l\,|\,\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{f}_k,\mathbf{d}_{l,k})\quad(23)$$

where $p(\mathbf{v}_l\,|\,\mathbf{Y}_{\mathrm{TB,TF}},\mathbf{f}_k,\mathbf{d}_{l,k})$ is given by (22).

B. Prior Models

The structure of the compression system motivates the selection of the other density functions in (17). The purpose of the first density function $p(\mathbf{f}_k)$ is to incorporate *a priori* information about the original high-resolution images into the recovery method. Most critical here is that the original images rarely contain any of the artifacts introduced during coding. Common errors include ringing artifacts that are caused by the coarse quantization of high-frequency information and blocking errors that arise from the independent processing of blocks. Both compression errors are penalized with the density

$$p(\mathbf{f}_k)\propto\exp\{-(\lambda_1\|\mathbf{Q}_1\mathbf{f}_k\|^2+\lambda_2\|\mathbf{Q}_2\mathbf{AHf}_k\|^2)\}\quad(24)$$

where $\mathbf{Q}_1$ represents a linear high-pass operation, $\mathbf{Q}_2$ represents a linear high-pass operation across block boundaries, and λ_1 and λ_2 control the influence of the two norms. By increasing the value for λ_1, the density describes a smoother image frame. Increasing the value for λ_2 denotes a frame with smooth block boundaries.

The last distribution appearing in (17) provides an *a priori* model for the displacement between image frames. Like the distribution in (24), we define the displacements to be smooth and absent of coding artifacts. Coding errors are largely attributable to the quantization and decimation of the motion field. To penalize these errors, the prior displacement is given by

$$p(\mathbf{D}_{\mathrm{TB,TF}})\propto\exp\left\{\sum_{l=k-\mathrm{TB}}^{k+\mathrm{TF}}-(\lambda_3\|\mathbf{Q}_3\mathbf{d}_{l,k}\|^2)\right\}\quad(25)$$

where $\mathbf{Q}_3$ is a linear high-pass operator, and λ_3 a control parameter. We note that dependencies between the displacements of different frames could also be incorporated into (25).

C. Optimization Procedure

Substituting (19)–(25) into (18), we get

$$\begin{aligned}\hat{\mathbf{f}}_k, \hat{\mathbf{D}} = \arg\min_{\mathbf{f}_k, \mathbf{D}} \Bigg\{ & \sum_{l=k-\mathrm{TB}}^{k+\mathrm{TF}} (\mathbf{y}_l - \mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k)^T \\ & \times \mathbf{K}_{Q,l}^{-1}(\mathbf{y}_l - \mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k) \\ & + \sum_{l=k-\mathrm{TB}}^{k+\mathrm{TF}} \left(\mathbf{y}_l^{\mathrm{MV}} - \mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k\right)^T \\ & \times \mathbf{K}_{\mathrm{MV},l}^{-1} \left(\mathbf{y}_l^{\mathrm{MV}} - \mathbf{AH}C(\mathbf{d}_{l,k})\mathbf{f}_k\right) \\ & + \lambda_1 \|\mathbf{Q}_1\mathbf{f}_k\|^2 + \lambda_2 \|\mathbf{Q}_2\mathbf{AHf}_k\|^2 + \lambda_3 \|\mathbf{Q}_3\mathbf{d}_{l,k}\|^2 \Bigg\}. \end{aligned} \tag{26}$$

The minimization of (26) can be accomplished with the cyclic coordinate-descent optimization method [45]. In this approach, an estimate for the motion field is first found by assuming that the high-resolution image is known. Then, estimates for the displacements are found while the high-resolution image is assumed known. The high-resolution image is then re-estimated with the recently found displacements, which is then re-estimated with the most recent high-resolution estimate. The process iterates until the estimates converge.

Treating the high-resolution image as a known parameter, the estimate for the motion field in (26) can be found by the method of successive approximations as follows:

$$\begin{aligned}\hat{\mathbf{d}}_{l,k}^{m+1} = \hat{\mathbf{d}}_{l,k}^{m} - \alpha_d^{l,k} \Bigg\{ & \frac{\partial C\left(\hat{\mathbf{d}}_{l,k}^{m}\right)\mathbf{f}_k}{\partial \hat{\mathbf{d}}_{l,k}^{m}} \mathbf{H}^T\mathbf{A}^T \\ & \times \left[\mathbf{K}_{Q,l}^{-1}\left(\mathbf{y}_l - \mathbf{AH}C\left(\hat{\mathbf{d}}_{l,k}^{m}\right)\mathbf{f}_k\right)\right. \\ & \left. + \mathbf{K}_{\mathrm{MV},l}^{-1}\left(\mathbf{y}_l^{\mathrm{MV}} - \mathbf{AH}C\left(\hat{\mathbf{d}}_{l,k}^{m}\right)\mathbf{f}_k\right)\right] + \lambda_3 \mathbf{Q}_3^T\mathbf{Q}_3\hat{\mathbf{d}}_{l,k}^{m} \Bigg\} \end{aligned} \tag{27}$$

where $\hat{\mathbf{d}}_{l,k}^{m+1}$ and $\hat{\mathbf{d}}_{l,k}^{m}$ are respectively the $(m+1)$th and mth estimates of the displacement between frame k and l, $\mathbf{A}^T$ defines the up-sampling operation, and $\alpha_d^{l,k}$ controls the convergence and rate of convergence of the algorithm.

Once the estimate for the motion field is found, then the high-resolution image is computed. For a fixed $\mathbf{D}_{\mathrm{TB,TF}}$, the minimization of (26) is expressed as

$$\begin{aligned}\hat{\mathbf{f}}_k^{n+1} = \hat{\mathbf{f}}_k^{n} - \alpha_f \Bigg\{ & \sum_{l=k-\mathrm{TB}}^{k+\mathrm{TF}} \mathbf{C}^T(\mathbf{d}_{l,k})\mathbf{H}^T\mathbf{A}^T \\ & \times \left[\mathbf{K}_{Q,l}^{-1}\left(\mathbf{y}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\hat{\mathbf{f}}_k^{n}\right)\right. \\ & \left. + \mathbf{K}_{\mathrm{MV},l}^{-1}\left(\mathbf{y}_l^{\mathrm{MV}} - \mathbf{AHC}(\mathbf{d}_{l,k})\hat{\mathbf{f}}_k^{n}\right)\right] \\ & + \lambda_1 \mathbf{Q}_1^T\mathbf{Q}_1\hat{\mathbf{f}}_k^{n} + \lambda_2 \mathbf{H}^T\mathbf{A}^T\mathbf{Q}_2^T\mathbf{Q}_2\mathbf{AH}\hat{\mathbf{f}}_k^{n} \Bigg\} \end{aligned} \tag{28}$$

where $\hat{\mathbf{f}}_k^{n+1}$ and $\hat{\mathbf{f}}_k^{n}$ are the enhanced frames at the $(n+1)$th and nth iteration, α_f is a relaxation parameter that determines the convergence and rate of convergence of the algorithm, and $\mathrm{C}^T(\mathbf{d}_{k,l})$ compensates an image backward along the motion vectors.

VI. Simulations

Assessing the performance of a super-resolution algorithm is difficult, as the solution quality depends on several tasks. These tasks include registration, interpolation, restoration, and (in the proposed approach) post-processing. In this section, we present a sequence of experiments to evaluate the proposed algorithm. A synthetically generated image sequence is considered first. Results quantify the accuracy of the motion estimates, as well as the resolution enhancement capability of the algorithm. In the second set of experiments, we consider the enhancement of a natural image sequence. This provides a more realistic gauge of resolution improvement.

In all of the presented results, the influence of the compression ratio on the super-resolution procedure is also considered. We utilize the MPEG-4 bitstream syntax, which describes a hybrid motion-compensation and block discrete cosine transform (DCT) compression system. The spatial dimension of the low-resolution sequence is 176×144 pixels, and the temporal rate is 30 frames per second. Besides the first image in the sequence, which is intra-coded, each frame is compressed as a P-frame. This restricts the reference frame for the motion vectors to be the temporally proceeding frame. The VM5+ rate control mechanism is utilized for bit allocation, and all frames are encoded. For both image sequences, the bit-rates of 256 kbps and 1 Mbps are considered. This corresponds to a "low" and "high" quality coding application, respectively.

Parameters in the bitstream define the covariance matrices $\mathbf{K}_{Q,l}$ and $\mathbf{K}_{\mathrm{MV},l}$, and we utilize the methods described in (13) and (16) for these experiments. Remaining parameters are chosen heuristically. Values are $\lambda_1 = \lambda_2 = 10^{-3}, \lambda_3 = 10^3$, and $\alpha_f = .125$ and $\alpha_d^{l,k} = 10^{-6}$. The matrices $\mathbf{Q}_1$ and $\mathbf{Q}_3$ are block circulant and denote circular convolution of the image with the discrete Laplacian. The matrix $\mathbf{Q}_2$ realizes a difference operator across the 8×8 block boundaries. In the iterative procedure that we utilize, given at iteration i the values of the high-resolution image $\hat{\mathbf{f}}_k^{i}$ and the displacement vector $\hat{\mathbf{d}}_{l,k}$, we fix $\mathbf{d}_{l,k}$ in (28) to $\hat{\mathbf{d}}_{l,k}$ and iterate that equation until the difference between two consecutive high-resolution image estimates, $\mathbf{f}^{\mathrm{new}}$ and $\mathbf{f}^{\mathrm{old}}$, satisfy $\|\mathbf{f}^{\mathrm{new}} - \mathbf{f}^{\mathrm{old}}\|^2 / \|\mathbf{f}^{\mathrm{old}}\|^2 < 10^{-6}$. Then, the new high-resolution image estimate $\hat{\mathbf{f}}_k^{i+1}$ is set to $\mathbf{f}^{\mathrm{new}}$. Using this new estimate of the original high-resolution image, we calculate the high-resolution displacements using (27) until the difference between two consecutive estimates satisfy $\|\hat{\mathbf{d}}_{l,k}^{\mathrm{new}} - \hat{\mathbf{d}}_{l,k}^{\mathrm{old}}\|^2 / \|\hat{\mathbf{d}}_{l,k}^{\mathrm{old}}\|^2 < 10^{-9}$. The entire process terminates when $\|\hat{\mathbf{f}}_k^{n+1} - \hat{\mathbf{f}}_k^{n}\|^2 / \|\hat{\mathbf{f}}_k^{n}\|^2 < 10^{-7}$.

In all experiments, the algorithm is initialized with the following procedure. First, the encoded images are bilinearly interpolated. This provides the initial estimate for the high-resolution intensities $\hat{\mathbf{f}}_k^{0}$. Displacement information is then calculated from the decoded image frames. The procedure in (27) is employed for this task and cast within a multiscale framework. Parameters

Fig. 1. Single frame from original high-resolution video sequence.

Fig. 2. Single frame from the (a) original low-resolution synthetic sequence. Frame after decoding the resulting (b) 256-kbps and (c) 1-Mbps MPEG-4 bitstreams. PSNR values for (b) and (c) are 29.8 dB and 37.4 dB, respectively.

for the motion estimate are the same as above, and the derivative in (27) is computed numerically. Throughout the experiments, this numerical procedure consists of simple difference operations. When expressed on a pixel-by-pixel basis, it is found with the equations $\mathbf{f}_k(x+\hat{d}^x_{l,k}(x,y)+1)-\mathbf{f}_k(x+\hat{d}^x_{l,k}(x,y)-1)$ and $\mathbf{f}_k(y+\hat{d}^y_{l,k}(x,y)+1)-\mathbf{f}_k(y+\hat{d}^y_{l,k}(x,y)-1)$. Pre-filtering the intensity data with a Gaussian filter mitigates inaccuracies from this simple difference estimate, and it also addresses any large-scale displacements. The variance of the filter is defined as $3S^2/2$, where S is the sample factor. During initialization, displacements are first found on the coarse 11×9 pixel grid ($S=16$), and the results upsampled and scaled by a factor of two with bilinear interpolation. The interpolated values serve as the initial condition for motion estimation at the finer scale. In the procedure, we utilize the scale factors 16, 8, 4, and 1 and initialize the displacement information to be zero at the coarsest scale. Once the displacements are found for the decoded image frames, the information is bilinearly interpolated and scaled to the dimensions of the high-resolution data. This serves as the initial estimate for $\hat{\mathbf{d}}^0_{l,k}$.

A. Synthetic Simulations

For the first set of experiments, we utilize a single frame from the "mobile" image sequence. The spatial dimension of the frame is 704×576 pixels, though we restrict our processing to the central 352×288 pixel region to reduce computational expense. The resulting image is shown in Fig. 1. Having extracted the image, we synthetically create an image sequence by shifting the frame according to

$$\mathbf{f}_k = C\left(\mathbf{d}_{\mathrm{mod}(k,4),E}\right)\mathbf{f}_E \tag{29}$$

where $\mathbf{f}_E$ is the extracted image, $\mathrm{mod}(k,4)$ is the modulo operator that divides k by 4 and returns the remainder, and $\mathbf{d}_{0,E},\mathbf{d}_{1,E},\mathbf{d}_{2,E}$, and $\mathbf{d}_{3,E}$ represent, respectively, no displacement, a horizontal pixel displacement, a vertical pixel displacement and a pixel displacement in both the horizontal and vertical directions.

The image sequence defined by (29) contains two important attributes. First, displacements between the frames are completely known. This facilitates a quantitative evaluation of the displacement estimate. Second, since the low-resolution image sequence is constructed by discarding every other pixel in the horizontal and vertical directions, each pixel in the original image is observable within four frames of the image sequence. This serves as the best-case scenario for resolution enhancement. Thus, we can evaluate the algorithm under ideal enhancement conditions.

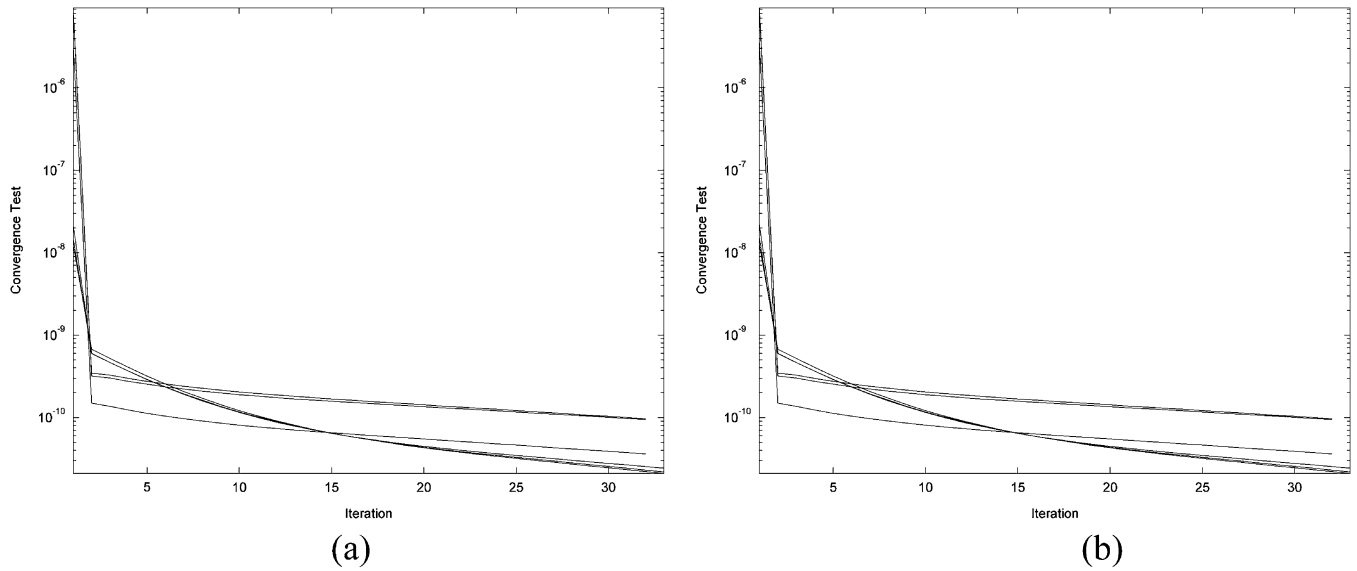

Fig. 3. Convergence plots for nonsynthetic experiments: (a) subpixel displacement and (b) high-resolution intensity values. Both converge for the 256-kbps and 1-Mbps estimates. The curves terminating at iteration 33 and 32 correspond to the 256-kbps and 1-Mbps experiments, respectively.

After extracting the high-resolution image sequence, it is decimated by a factor of two in the horizontal and vertical directions. The low-resolution image sequence is then processed with an MPEG-4 encoder operating at 256 kbps and 1 Mbps. A frame from each sequence appears in Fig. 2, where the amount of degradation is quantified by the peak signal-to-noise ratio (PSNR)

$$\mathrm{PSNR} = 255^2 \cdot \mathrm{MN}/\|\mathbf{e}\|^2 \tag{30}$$

where MN is the number of pixels in the low-resolution image. For the 256-kbps and 1-Mbps images in Fig. 2, the PSNR is 29.8 and 37.4 dB, respectively. Both bitstreams are then processed by the proposed algorithm with $\mathbf{TF} = 1$ and $\mathbf{TB} = 2$, and the algorithm converges to a joint estimate for the high-resolution frame and subpixel displacements. Convergence plots appear in Fig. 3.

Evaluation of the displacement estimate is our first task. We calculate the Euclidean distance between the estimated and known displacement value for each pixel. (Note that there are three error measurements per pixel since $\mathbf{TF} = 1$ and $\mathbf{TB} = 2$, so the number of displacement errors is three times the image size.) These errors are then pooled over the entire frame by computing the average squared value. The resulting values are 6.34×10^{-2} pixels^2 and 4.85×10^{-2} pixels^2 for the 256-kbps and 1-Mbps bitstreams, respectively. This indicates a motion estimate that is closely aligned with the actual values.

Evaluating the displacement estimates is the first step in assessing the proposed algorithm. Next, we consider the super-resolution estimates. We interpolate the decoded frames using the traditional bilinear and bicubic methods, and compare the results to the proposed method. PSNR values for the three high-resolution images provide a method for comparison. For the low bit-rate sequence, the bilinear, bicubic and super-resolution approaches result in a PSNR of 28.2 dB, 28.4 dB, and 28.5 dB, respectively. This is an interesting result, as it illustrates a peculiarity of the super-resolution for compressed video problem. While the motion estimates are known to be reliable here, redundancies between frames have been removed during encoding to lower the bit-rate. Thus, the use of additional frames for resolution enhancement (by the proposed method) provides little gain over the single frame interpolation methods.

Encoding the synthetic sequence at 1-Mbps provides a bitstream more amendable to resolution improvement. PSNR values for the bilinear, bicubic, and proposed methods are 30.1 dB, 30.3 dB, and 35.8 dB, respectively. This corresponds to an improvement of over 5.7 dB and 5.5 dB, when compared to the bilinear and bicubic methods, respectively. The improvement is easily visible. The bilinear, bicubic and super-resolution estimates for a frame from the high-resolution sequence are presented in Fig. 4(a)–(c), respectively. By inspecting the images, it should be apparent that the proposed algorithm increases the resolution. This is evident in a number of locations in the image. For example, notice the legibility of the text in the calendar. The numbers are intelligible and well formed, as is the "January" month heading. For a second example, inspect the vertical and diagonal strips in the central top part of the frame. The pattern is quite degraded in the bilinear and bicubic estimates. However, it is visible and sharp in the super-resolution result.

B. Nonsynthetic Experiments

In the second set of experiments, we decimate and encode the entire "mobile" image sequence. This set of images differs from the synthetic set in two important ways. As a first difference, subpixel displacements between image frames are no longer defined explicitly. The displacements now correspond to inherent motion within the scene, as introduced by the camera and objects. As a second difference, all pixels in the high-resolution frame may not appear in the low-resolution sequence. Instead, a pixel may never be observable (if there is no motion within the scene), or it may only be observable in temporally distant observations. Fusing additional frames into the high-resolution estimate somewhat mitigates this problem.

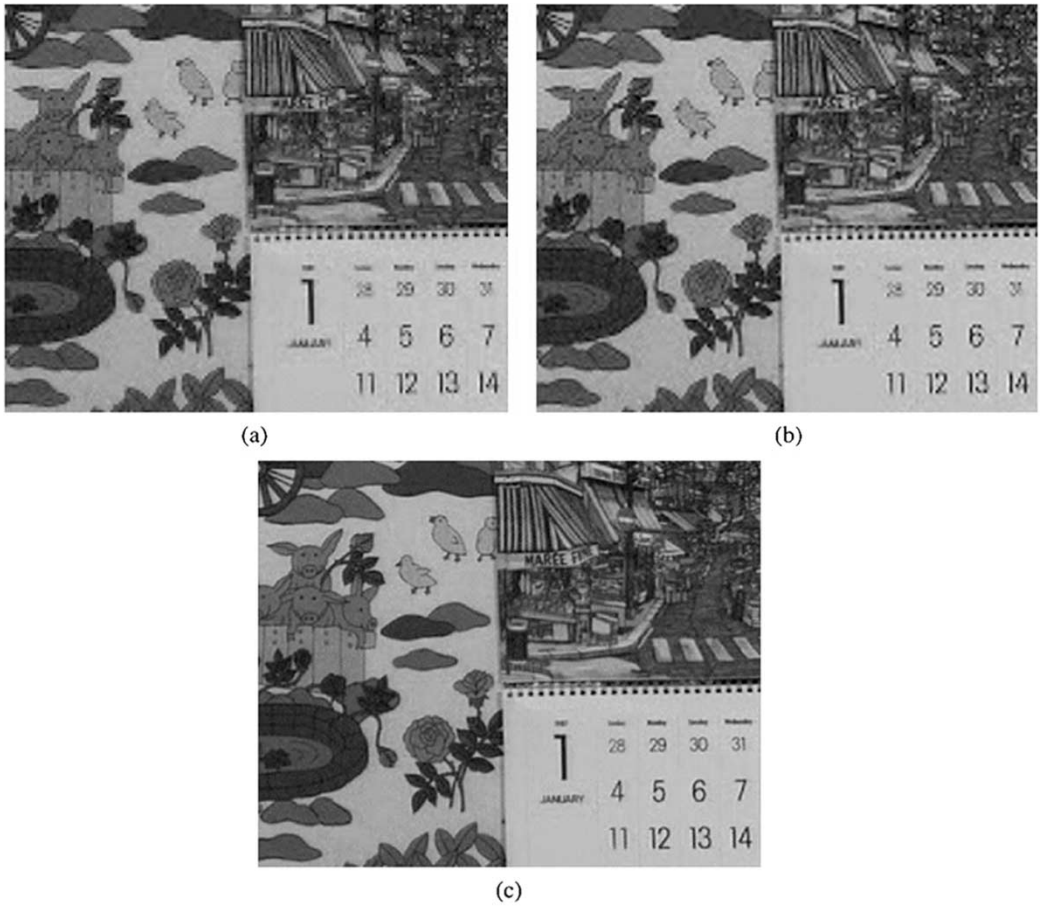

Fig. 4. High-resolution estimates from 1-Mbps synthetic experiment. Result after (a) bilinear interpolation, (b) bicubic interpolation, and (c) the proposed algorithm. The proposed method enhances the resolution throughout the image. Notice the sharpness of the numbers and text as well as the vertical features in the central-upper part of the frame. PSNR values for the frames are (a) 30.1 dB, (b) 30.3 dB, and (c) 35.8 dB.

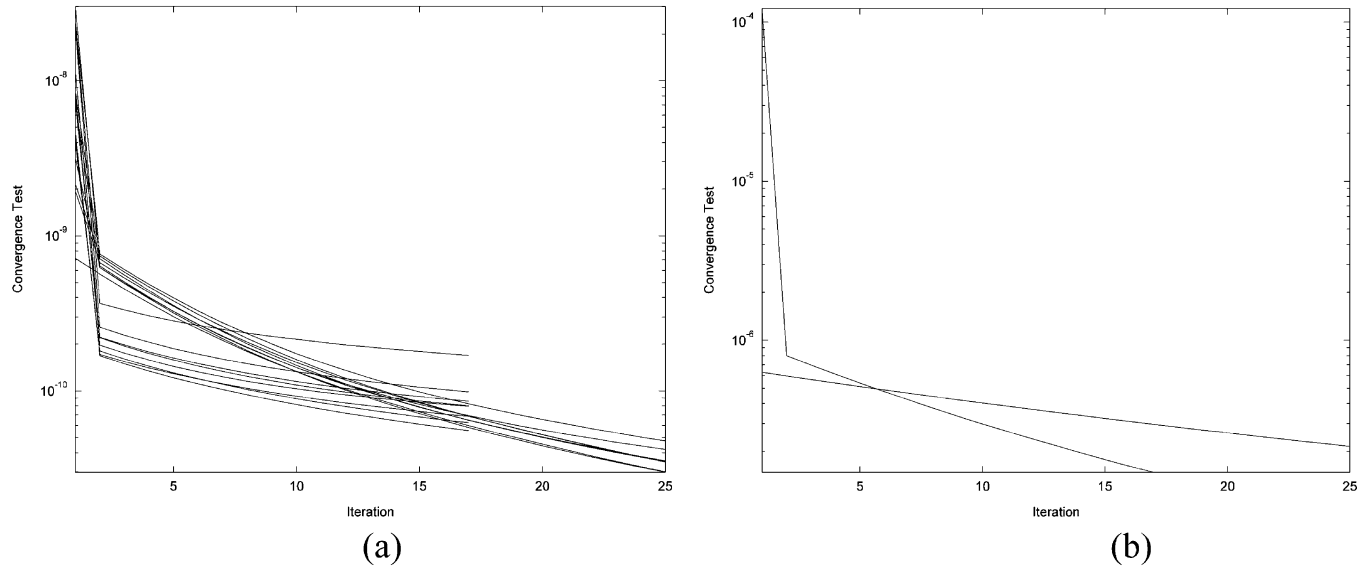

Fig. 5. Convergence plots for nonsynthetic experiments: (a) subpixel displacement and (b) high-resolution intensity values. Both converge for the 256-kbps and 1-Mbps estimates. The curves terminating at iteration 25 and 17 correspond to the 256-kbps and 1-Mbps experiments, respectively.

After extracting the high-resolution image sequence, it is decimated by a factor of two in the horizontal and vertical directions. The low-resolution image sequence is then processed with an MPEG-4 encoder operating at 256 kbps and 1 Mbps. These are the same bit-rates considered previously. However, please note that this does not assure a similar corrupting process. The compression process varies the level of degradation according to the composition of the images and underlying motion. Frames

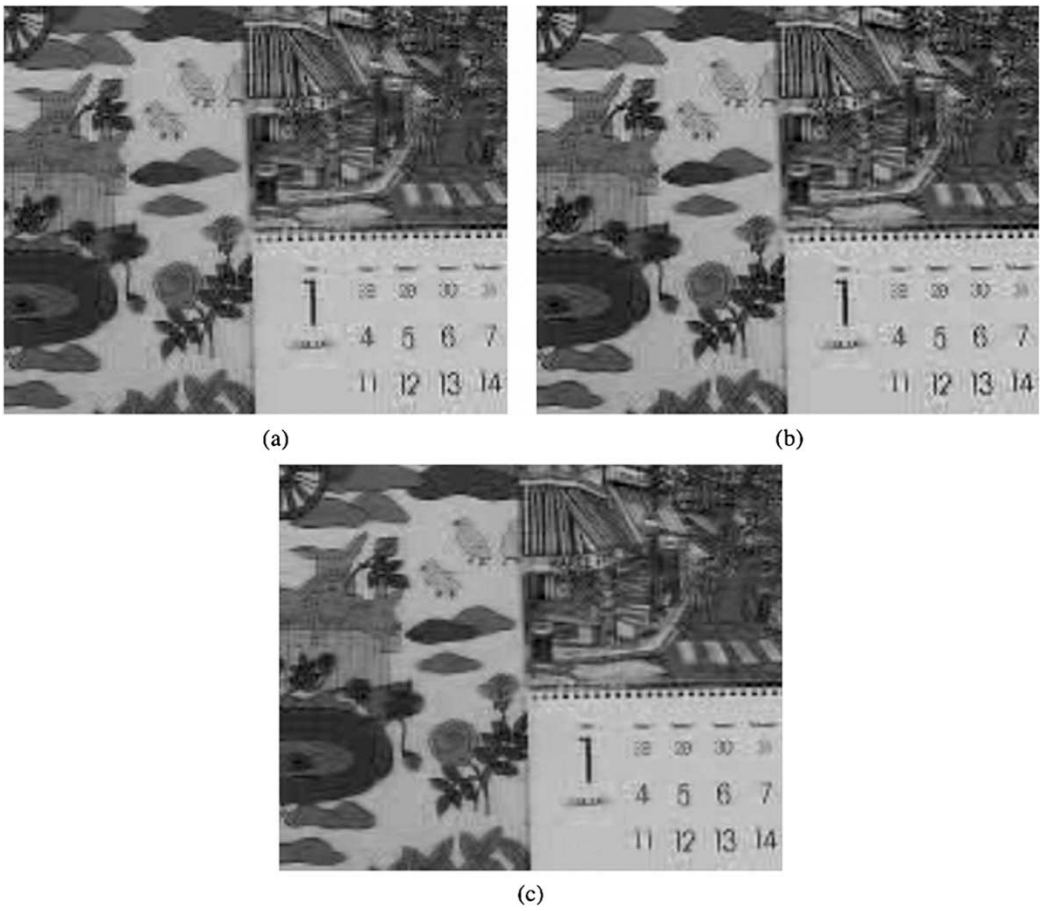

Fig. 6. High-resolution estimates from 256-kbps nonsynthetic experiment. Result after (a) bilinear interpolation, (b) bicubic interpolation, and (c) the proposed algorithm. As in the synthetic experiments, frames in the compressed sequence are highly similar and preclude resolution-enhancement. PSNR values for the frames are (a) 28.5 dB, (b) 28.6 dB, and (c) 29.0 dB.

from the 256-kbps and 1-Mbps nonsynthetic sequence are similar to the synthetic simulations, though PSNR values of 30.2 and 39.9 dB indicate less degradation. This suggests more redundancy between frames in the original image sequence (and less potential for resolution improvement).

Evaluating the simulations begins with the displacement values. To measure this error, we process a frame from both bitstreams with the proposed algorithm. Parameters in the bitstream define the covariance matrices $\mathbf{K}_{Q,l}$ and $\mathbf{K}_{\mathrm{MV},l}$, as before, and we integrate nine frames into the result by setting $\mathrm{TB} = \mathrm{TF} = 4$. (Convergence plots appear in Fig. 5.) This produces eight displacement estimates. The original high-resolution frame is then compensated with each of the estimates and compared to the actual frame for that time instant. Histograms of the errors are computed, and the variance of the errors is 61.6 and 72.3 intensity values^2 for the 256-kbps and 1-Mbps bitstreams, respectively. This reflects a number of inaccuracies within the displacement estimate, which include the occlusion of objects entering and leaving the frame as well as the temporal noise present in the acquired sequence.

With larger residual errors, it is important to assess the resulting decrease in the amount of resolution recovery. This is the second point of the experiment, and we begin with the 256-kbps bitstream. PSNR values for the bilinear, bicubic and super-resolution methods are computed as before, and they are equal to 28.5, 28.6, and 29.0 dB, respectively. These values are similar to the synthetic results, and it is observed that the encoder introduces redundancies between frames. However, the proposed method is still able to impart a benefit. Visual results are provided in Fig. 6, and inspection of the image shows an improvement within the super-resolved estimate. This improvement is best described as a suppression of high-frequency artifacts, which is a form of post-processing.

The 1-Mbps bitstream provides the information necessary for resolution recovery. PSNR values for the bilinear, bicubic and proposed algorithm are 30.3, 30.5, and 33.2 dB, respectively. This suggests a marked amount of resolution improvement by the proposed method, and visual results support this conclusion. An example frame appears in Fig. 7(a)–(c), which shows the results of bilinear interpolation, bicubic interpolation, and the proposed super-resolution method, respectively. As can be seen from the figure, several areas benefit from the proposed method. For example, the numbers in the calendar are sharper in the super-resolved estimate. Additionally, the vertical and diagonal stripes at the central part of the frame are improved. A second example appears in Fig. 8, where we show results from processing a different frame region encoded at 1 Mbps. The motion within this image is more complicated, and it is therefore encoded with more redundancy. Nevertheless, we still observe improvements in the super-resolved estimate. In the figure, the

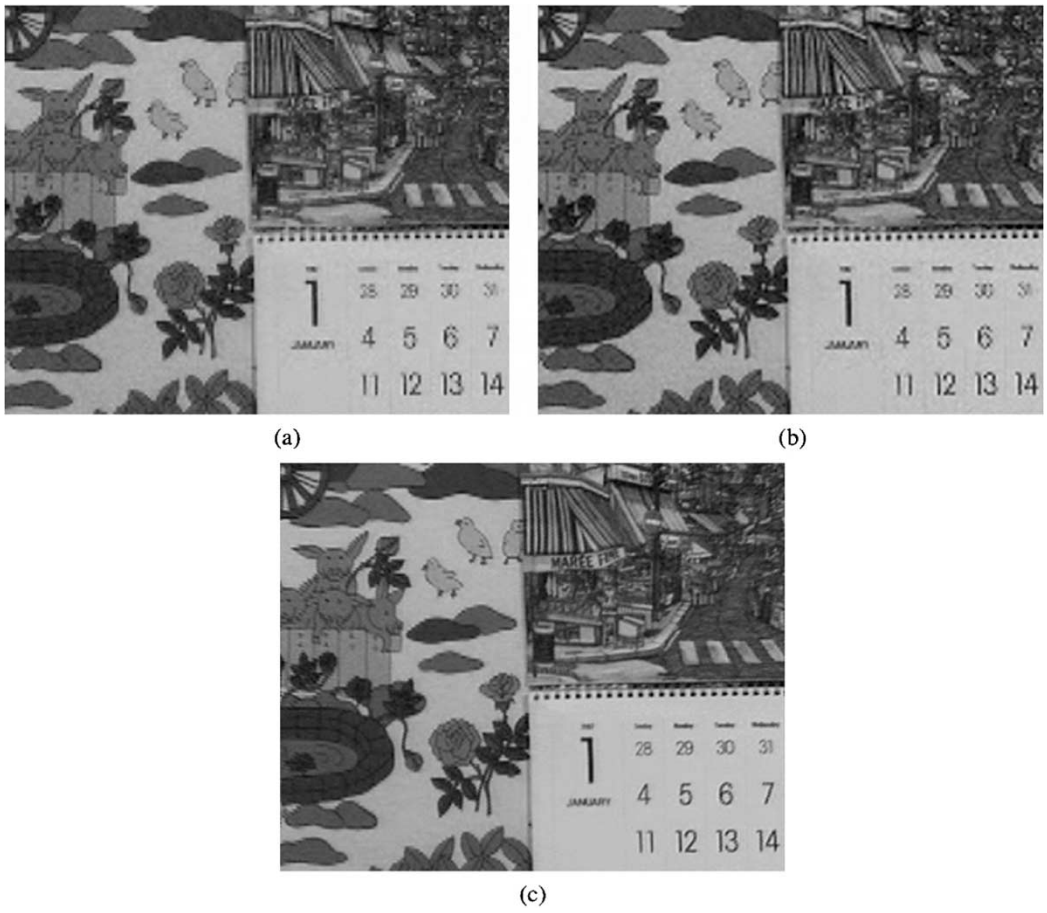

Fig. 7. High-resolution estimates from 1-Mbps nonsynthetic experiment. Result after (a) bilinear interpolation, (b) bicubic interpolation, and (c) the proposed algorithm. The proposed algorithm improves the overall quality of the frame by enhancing resolution and attenuating coding noise. PSNR values for the frames are (a) 30.3 dB, (b) 30.5 dB, and (c) 33.2 dB, respectively.

circles on the ball are better defined while the tip of the train is less jagged.

Comparing the super-resolution estimates from the synthetically generated and actual image sequence leads to several conclusions. As a first observation, we see that the synthetic images suffer more corruption during compression. This is attributed to a lack of redundancy between frames, which requires more bits to represent, and is apparent in a .2 dB decrease in the bilinear and bicubic interpolation estimates. At the higher compression ratio though, the encoder introduces additional redundancy into the sequence. Thus, we observe little resolution improvement and perceive the result of post-processing. Interestingly, the PSNR improvement for the actual image sequence is higher than the synthetic result. We credit this increase to the use of an additional five frames in the recovery procedure.

The 1-Mbps sequence provides a good example for resolution recovery. In the synthetic simulations, we see visual evidence of resolution improvement and a PSNR increase of 5.7 and 5.5 dB above the bilinear and bicubic interpolation methods, respectively. This sequence is biased toward the super-resolution algorithm, as frames in the sequence exhibit minimal redundancy and the displacements contain little entropy. Thus, the PSNR differences suggest a measure of maximum resolution improvement. Simulations with the actual image sequence realize an improvement below the synthetic experimental value. The improvement in PSNR is 2.9 and 2.7 dB when compared to the bilinear and bicubic results, respectively. This corresponds to approximately half of the improvement realized during the synthetic experiments. It also provides a discernible amount of resolution enhancement, increasing the legibility of text and the visibility of objects.

VII. Conclusion

Super-resolution from compressed video requires a solution to the registration, interpolation, restoration and post-processing tasks. In this paper, we have presented a methodology for solving these problems concurrently. This is especially appealing when considering hybrid motion-compensation and transform based coding methods, as the techniques provide an observation for both the necessary registration and intensity parameters. Our algorithm utilizes the Bayesian framework to incorporate information from the bitstream as well as to model synthetic coding artifacts, and it relies on a cyclic coordinate descent optimization for realization. Relationships between the algorithm parameters and information in the bitstream are also considered.

Simulations explore the efficacy of the algorithm for the resolution enhancement task. A synthetic sequence (composed

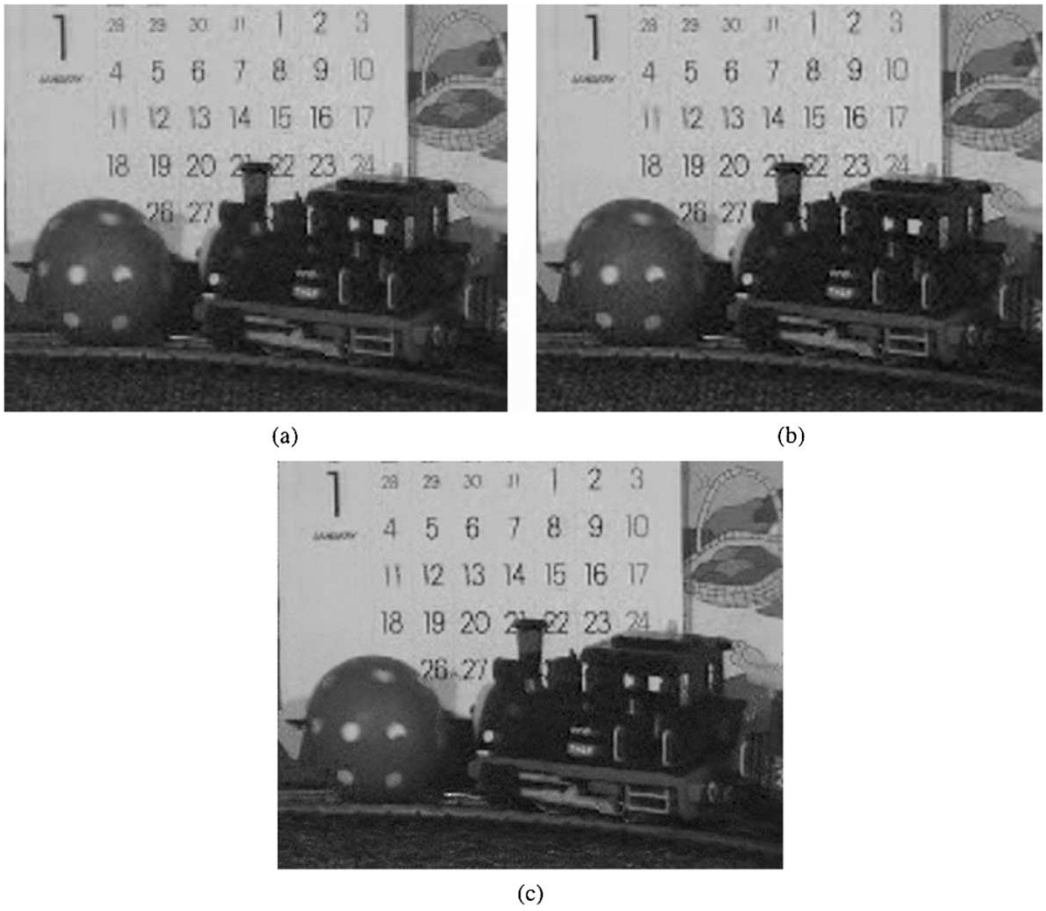

Fig. 8. High-resolution estimates from 1-Mbps nonsynthetic experiment. Result after (a) bilinear interpolation, (b) bicubic interpolation, and (c) the proposed algorithm. The proposed algorithm improves the overall quality of the frame by enhancing resolution and attenuating coding noise. PSNR values for the frames are (a) 31.0 dB, (b) 31.1 dB, and (c) 31.5 dB, respectively.

of shifted versions of a single frame) is evaluated first, and it illustrates the impact of compression on the super-resolution problem. Severely compressed sequences are shown to be poor candidates for resolution improvement, as the compression process increases the similarities between observations. More moderate bit rates provide a discernible improvement in signal quality. A second set of experiments processes an actual image sequence. Here, the motion is unknown and must be estimated from the decoded frames. When compared to the synthetic experiments, this decreases the level of resolution enhancement. However, results from the proposed method are still improved when compared to traditional bilinear and bicubic interpolation methods.

REFERENCES

[1] *Video Coding for Low Bitrate Communications*, ITU-T Recommendation H.263, Feb. 1998.

[2] *Video Codec for Audio Visual Services at* $p \times 64$ *kbits/s*, ITU-T Recommendation H.261, Mar. 1993.

[3] *Information Technology—Generic Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1.5 Mbps—Part 2: Video*, ISO/IEC JTC1/SC29 International Standard 11 172-2, 1993.

[4] *Information Technology—Generic Coding of Moving Pictures and Associated Audio Information: Video*, ISO/IEC JTC1/SC29 International Standard 13 818-2, 1995.

[5] *Information Technology—Generic Coding of Audio-Visual Objects: Visual*, ISO/IEC JTC1/SC29 International Standard 14 496-2, 1999.

[6] *Information Technology—Generic Coding of Audio-Visual Objects: Visual*, ISO/IEC JTC1/SC29 International Standard 14 496-2AM1, 2000.

[7] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, T. S. Huang, Ed. Greenwich, CT: JAI Press, 1984, pp. 317–339.

[8] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multi-frames," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 18, pp. 1013–1027, June 1990.

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

[10] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

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

[12] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, pp. 1621–1633, Dec. 1997.

[13] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, "Super-resolved image reconstruction from multiple images," NASA Ames Research Center, Moffet Field, CA, Tech. Rep. FIA-94-12, 1994.

[14] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," in *Proc. IEEE Int. Conf. Image Processing*, Washington, DC, 1995.

[15] S. Chaudhuri, Ed., *Super-Resolution Imaging*. Boston, MA: Kluwer Academic, 2001.
[16] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, no. 4, pp. 324–335, 1993.
[17] 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.
[18] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Super-resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, pp. 1064–1076, Aug. 1997.
[19] N. R. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Trans. Image Processing*, vol. 8, pp. 879–885, June 1999.
[20] P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object based high resolution image reconstruction from low resolution video," *IEEE Trans. Image Processing*, vol. 6, pp. 1446–1451, Oct. 1997.
[21] B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Processing*, vol. 10, pp. 278–287, Feb. 2001.
[22] H. C. Reeves and J. S. Lim, "Reduction of blocking effects in image coding," *Opt. Eng.*, vol. 23, no. 1, pp. 34–37, 1984.
[23] B. Ramamurthi and A. Gersho, "Nonlinear space variant post-processing of block coded images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1258–1267, May 1986.
[24] K. Sauer, "Enhancement of low bit-rate coded images using edge detection and estimation," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 1, pp. 52–62, 1991.
[25] C. J. Kuo and R. J. Hsich, "Adaptive postprocessor for block encoded images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 322–336, Aug. 1995.
[26] S.-W. Wu and A. Gersho, "Improved decoder for transform coding with application to the JPEG baseline system," *IEEE Trans. Commun.*, vol. 40, pp. 251–254, Feb. 1992.
[27] J. Luo, C. W. Chen, K. J. Parker, and T. Huang, "Artifact reduction in low bit rate DCT-based image compression," *IEEE Trans. Image Processing*, vol. 5, pp. 1363–1368, 1996.
[28] T. P. O'Rourke and R. L. Stevenson, "Improved image decompression for reduced transform coding artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 490–499, Dec. 1995.
[29] T. Ozcelik, J. C. Brailean, and A. K. Katsaggelos, "Image and video compression algorithms based on recovery techniques using mean field annealing," *Proc. IEEE*, vol. 83, pp. 304–316, Feb. 1995.
[30] Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images," *IEEE Trans. Image Processing*, vol. 4, pp. 896–908, July 1995.
[31] Y. Yang and N. P. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," *IEEE Trans. Image Processing*, vol. 6, pp. 1345–1357, Oct. 1998.
[32] S. J. Reeves and S. I. Eddins, "Comments on iterative procedures for reduction of blocking effects in transform coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 439–440, Dec. 1993.
[33] R. Rosenholtz and A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 91–94, Jan. 1992.
[34] Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 421–432, Dec. 1993.
[35] B. K. Gunturk, Y. Altunbasak, and R. Mersereau, "Bayesian resolution-enhancement framework for transform-coded video," in *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, 2001.
[36] A. J. Patti and Y. Altunbasak, "Super-resolution image estimation for transform coded video with application to MPEG," in *Proc. IEEE Int. Conf. Image Processing*, Kobe, Japan, 1999.
[37] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Processing*, vol. 3, pp. 233–242, Mar. 1994.
[38] J. Mateos, A. K. Katsaggelos, and R. Molina, "Simultaneous motion estimation and resolution enhancement of compressed low resolution video," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, 2000.
[39] ——, "Resolution enhancement of compressed low resolution video," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, 2000.
[40] C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian high-resolution reconstruction of low-resolution compressed video," in *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, 2001.
[41] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Super-resolution from compressed video," in *Super-Resolution Imaging*, S. Chaudhuri, Ed. Boston, MA: Kluwer, 2001, pp. 211–242.
[42] C. A. Segall and A. K. Katsaggelos, "Enhancement of compressed video using visual quality metrics," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, 2000.
[43] J. Mateos, A. K. Katsaggelos, and R. Molina, "A Bayesian approach for the estimation and transmission of regularization parameters for reducing blocking artifacts," *IEEE Trans. Image Processing*, vol. 9, pp. 1200–1215, July 2000.
[44] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurry, noisy, and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec. 1997.
[45] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984.

C. Andrew Segall received the B.S. and M.S. degrees in electrical engineering from Oklahoma State University, Stillwater, in 1995 and 1997, respectively, and the Ph.D. degree in electrical engineering from Northwestern University, Evanston, IL, in 2002.

He was a Murphy Fellow and Research Assistant at Northwestern University. He is currently a Senior Engineer at Pixcise, Inc., Palo Alto, CA., where he is developing scalable compression methods for high-definition video. His research interests are in image processing and include recovery problems for compressed video, scale-space theory, and nonlinear filtering.

Dr. Segall is a member of Phi Kappa Phi and Eta Kappa Nu.

Aggelos K. Katsaggelos (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical and Computer Engineering at Northwestern University, where he is currently a Professor, holding the Ameritech Chair of Information Technology. He is also the Director of the Motorola Center for Communications and a member of the Academic Affiliate Staff, Department of Medicine, at Evanston Hospital. He is the editor of *Digital Image Restoration* (New York: Springer-Verlag, 1991), co-author of *Rate-Distortion Based Video Compression* (Norwell, MA: Kluwer 1997), and co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Norwell, MA: Kluwer 1998), and the co-inventor of eight international patents

Dr. Katsaggelos is a member of the Publication Board of the IEEE PROCEEDINGS, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, the Editorial Board of Academic Press, Marcel Dekker: Signal Processing Series, *Applied Signal Processing*, and *Computer Journal*. He has served as Editor-in-Chief of the *IEEE Signal Processing Magazine* (1997–2002), member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), Area Editor for the journal *Graphical Models and Image Processing* (1992–1995), member of the Steering Committees of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1992–1997) and the IEEE TRANSACTIONS ON MEDICAL IMAGING (1990–1999), member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992–1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). He is the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), and an IEEE Signal Processing Society Best Paper Award (2001).

Rafael Molina was born in 1957. He received the degree in mathematics (statistics) in 1979 and the Ph.D. degree in optimal design in linear models in 1983.

He became a Professor of Computer Science and Artificial Intelligence at the University of Granada, Granada, Spain, in 2000. His areas of research interest are image restoration (applications to astronomy and medicine), parameter estimation in image restoration, low to high image and video, and blind deconvolution.

Dr. Molina is a member of SPIE, the Royal Statistical Society, and the Asociación Española de Reconocimiento de Formas y Análisis de Imágenes (AERFAI).

Javier Mateos was born in Granada, Spain, in 1968. He received the degree in computer science in 1991 and the Ph.D. degree in computer science in 1998, both from the University of Granada.

He was an Assistant Professor at the Department of Computer Science and Artificial Intelligence, University of Granada, from October 1992 to March 2001, and then became a permanent Associate Professor. He is conducting research on image and video processing, including image restoration, image and video recovery, and compression and super-resolution from (compressed) stills and video sequences.

He is a member of the Asociación Española de Reconocimento de Formas y Análisis de Imágenes (AERFAI) and International Association for Pattern Recognition (IAPR).