MAXIMUM A POSTERIORI SUPER-RESOLUTION OF COMPRESSED VIDEO USING A NEW MULTICHANNEL IMAGE PRIOR

Stefanos P. Belekos[1,3]*, Nikolaos P. Galatsanos*[2]*, S. Derin Babacan*[3]*, and Aggelos K. Katsaggelos*[1,3]

[1] University of Athens-Faculty of Physics-15784 Athens, Greece-stefbel@phys.uoa.gr
[2]University of Patras-Department of Electrical and Computer Engineering-26500 Rio, Greece
[3]Northwestern University-Department of Electrical Engineering and Computer Science-Evanston, IL, 60208, USA

ABSTRACT

Super-resolution (SR) algorithms for compressed video aim at recovering high-frequency information and estimating a high-resolution (HR) image or a set of HR images from a sequence of low-resolution (LR) video frames. In this paper we present a novel SR algorithm for compressed video based on the maximum *a posteriori* (MAP) framework. We utilize a new multichannel image prior model, along with the state-of-the art image prior and observation models. Moreover, relationship between model parameters and the decoded bitstream are established. Numerical experiments demonstrate the improved performance of the proposed method compared to existing algorithms for different compression ratios.

Index Terms— Image restoration, MAP framework, multichannel prior, resolution enhancement, video coding.

1. INTRODUCTION

The SR problem is an inverse problem that requires a regularized solution. In most of the Bayesian formulations which have been used for this problem so far, single channel image priors have been adopted [1], whereas other works utilize both non-Bayesian and Bayesian [2] total variation (TV) techniques. As far as the imaging models are concerned, many techniques incorporate the motion field (MF) information provided by the HR data [3] into the model, whereas others do not.

In this paper we address the compressed video SR problem utilizing a MAP approach. The main contribution of this work is the use of a new multichannel prior that incorporates the registration information between frames. Such multichannel approaches have been used successfully in the past for compressed video reconstruction [4]. Nevertheless, they were deterministic and the multichannel idea was basically imposed by using regularization between frames. The efficacy of the multichannel prior has already been proved for uncompressed data [5]. However, the compressed bitstream introduces other departures into the SR problem, which are taken into account in this work.

This paper is organized as follows. Sec. 2 describes the appropriate mathematical background. In Sec. 3 we introduce a MAP problem formulation for the SR of compressed video regarding two existing models as well as the proposed one, along with the corresponding algorithms. Experimental results are illustrated in Sec. 4, indicating the benefits of the new prior. Finally, Sec. 5 concludes the paper.

This paper is part of the 03ED-535 research project, implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from E.U.-European Social Fund). The first author performed the work while at Northwestern University, Department of Electrical Engineering and Computer Science, Evanston, IL, 60208, USA.

2. MATHEMATICAL BACKGROUND

2.1. Observation Models

In this paper we use two different observation models. In the first one the relationship between an uncompressed LR observation frame $\mathbf{g}_i$ and its HR counterpart $\mathbf{f}_i$ is given in matrix-vector form by (all images are lexicographically ordered into vectors)

$$\mathbf{g}_i = \mathbf{A}\mathbf{H}\mathbf{f}_i + \mathbf{n}_i, \quad i = 1, 2 \dots P, \tag{1}$$

where $\mathbf{g}_i$ and $\mathbf{f}_i$ are of dimensions $MN \times 1$ and $LMLN \times 1$, respectively, $\mathbf{A}$ is the $MN \times LMLN$ down-sampling matrix, $\mathbf{H}$ is the $LMLN \times LMLN$ known blurring matrix, $\mathbf{n}_i$ of size $MN \times 1$, represents the additive white Gaussian (acquisition) noise (AWGN) term, P represents the total number of frames and L denotes the resolution enhancement factor.

The compressed observation of the ith LR frame $\mathbf{y}_i$ is therefore expressed as

$$\mathbf{y}_i = \mathbf{T}^{-1}\mathbf{Q}[\mathbf{T}(\mathbf{g}_i - \mathbf{M}(\mathbf{v}_{i,j})\mathbf{y}_j)] + \mathbf{M}(\mathbf{v}_{i,j})\mathbf{y}_j, j = 1 \dots P, \tag{2}$$

where $\mathbf{v}_{i,j}$ is the vector containing the transmitted motion vectors that predict frame i from a previously compressed frame j, $\mathbf{M}_{i,j} = \mathbf{M}(\mathbf{v}_{i,j})$ represents the 2-D matrix of size $MN \times MN$ which describes the mapping of frame $\mathbf{y}_j$ into frame $\mathbf{y}_i$,

$\mathbf{Q}[\cdot]$ denotes the quantization procedure, $\mathbf{T}$ and $\mathbf{T}^{-1}$ are the forward and inverse transform operations and $\mathbf{g}_i - \mathbf{M}(\mathbf{v}_{i,j})\mathbf{y}_j$ is the motion compensation error. Combining (1) and (2) the relationship between any LR and HR image becomes

$$\mathbf{y}_i = \mathbf{T}^{-1}\mathbf{Q}[\mathbf{T}(\mathbf{AHf}_i - \mathbf{y}_i^{MV} + \mathbf{n}_i)] + \mathbf{y}_i^{MV}, \quad (3)$$

where the motion compensated estimate of frame i is denoted by $\mathbf{y}_i^{MV} = \mathbf{M}(\mathbf{v}_{i,j})\mathbf{y}_j$.

Given that $\mathbf{Q}[\cdot]$ introduces compression noise into the decoded frames, which is dominant over the acquisition noise, we approximate the quantity $\mathbf{T}^{-1}\mathbf{Q}[\mathbf{T}(\mathbf{AHf}_i - \mathbf{y}_i^{MV} + \mathbf{n}_i)]$ by

$$\mathbf{AHf}_i - \mathbf{y}_i^{MV} + \mathbf{e}_i, \quad (4)$$

where $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{K}_i)$ is the quantization noise model and $\mathbf{K}_i$ is the covariance matrix in the spatial domain for the ith frame. Assuming an independent and identically distributed (IID) noise process, similarly to [1], $\mathbf{K}_i = \varepsilon_i^{-1}\mathbf{I}$, where $\mathbf{I}$ is the identity matrix of size $MN \times MN$ and ε_i is the inverse quantization noise variance (precision parameter). Combining (3) and (4), we obtain

$$\mathbf{y}_i = \mathbf{AHf}_i + \mathbf{e}_i. \quad (5)$$

The second imaging model used in this paper, is defined as

$$\mathbf{g}_i = \mathbf{AHD}(\mathbf{d}_{i,k})\mathbf{f}_k + \mathbf{w}_{i,k}, i = k-m, \ldots, k, \ldots, k+n, \quad (6)$$

with $\mathbf{w}_{i,k}$ a column vector of size $MN \times 1$ representing *the total contribution* of the noise term (registration and acquisition errors) which is again modelled as AWGN and n, m indicate respectively the number of frames used in the forward and backward temporal directions ($n + m + 1 = P$) with respect to the kth frame. Moreover, $\mathbf{D}(\mathbf{d}_{i,k})$ is the 2-D motion compensation matrix of size $LMLN \times LMLN$, mapping frame $\mathbf{f}_k$ into frame $\mathbf{f}_i$ with the use of $\mathbf{d}_{i,k}$ (displacements).

Following the previously analyzed steps, we state the relationship between any LR and HR image as

$$\mathbf{y}_i = \mathbf{AHD}_{i,k}\mathbf{f}_k + \mathbf{e}_{i,k}, \quad (7)$$

where $\mathbf{D}(\mathbf{d}_{i,k}) = \mathbf{D}_{i,k}$ and $\mathbf{e}_{i,k} \sim N(\mathbf{0}, \mathbf{K}_{i,k})$ is the quantization noise model with $\mathbf{K}_{i,k} = \varepsilon_{i,k}^{-1}\mathbf{I}$ representing the covariance matrix in the spatial domain for the ith frame, where $\varepsilon_{i,k}$ is the precision related to both quantization and registration noise components.

Moreover, in both these observation models the noise component by the motion vectors provided in the compressed bitstream should be incorporated, which is modelled as

$$\mathbf{y}_i^{MV} = \mathbf{M}(\mathbf{v}_{i,j})\mathbf{y}_j = \mathbf{AHD}_{i,j}\mathbf{f}_j + \mathbf{w}_{ij,MV}, \quad (8)$$

where based on the assumptions stated in [1] $\mathbf{w}_{ij,MV} \sim N(\mathbf{0}, \mathbf{K}_{ij,MV})$ and $\mathbf{K}_{ij,MV} = \delta_{i,j}^{-1}\mathbf{I}$ is the respective covariance matrix in the spatial domain for the ith frame ($\delta_{i,j}$ is the precision related to the displaced frame difference (DFD)). Clearly, in the case of the second observation model j=k.

2.2. Image Prior Models

In this work we consider two prior models in order to penalize compression errors. The first one, is the new *multichannel* prior proposed in [5], which takes into account both within frame smoothness and between-frame smoothness incorporated through the motion (MF) information. More specifically, the multichannel prior pdf is expressed as

$$p(\tilde{\mathbf{f}}; \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \propto \prod_{i=k-m}^{k+n} \prod_{\substack{j=k-m \\ j\neq i}}^{k+n} p(\mathbf{f}_i|\mathbf{f}_j; \beta_{i,j}) p(\mathbf{f}_j; \alpha_j), \quad (9)$$

where $\tilde{\mathbf{f}} = [\mathbf{f}_{k-m}^T, \ldots, \mathbf{f}_k^T, \ldots, \mathbf{f}_{k+n}^T]^T$ and T denotes the transpose of a matrix or vector. Moreover, $p(\mathbf{f}_i|\mathbf{f}_j; \beta_{i,j}) \propto \exp(-\frac{\beta_{i,j}}{2}\|\mathbf{f}_i - \mathbf{D}_{i,j}\mathbf{f}_j\|^2)$, with $\mathbf{D}_{j,i} = (\mathbf{D}_{i,j})^T = \mathbf{D}(\mathbf{d}_{j,i})$ and matrix $(\mathbf{D}_{i,j})^T$ represents the operation of backward motion compensation along the motion vectors. The second image prior $p(\mathbf{f}_j; \alpha_j) \propto \exp(-\frac{\alpha_j}{2}\|\mathbf{Q}\mathbf{f}_j\|^2)$ is related to the smoothness within each frame, and is modelled by a Simultaneously Autoregressive (SAR) distribution, where $\mathbf{Q}$ represents a linear high-pass convolutional operator of size $LMLN \times LMLN$, $\|\cdot\|$ denotes the l_2 norm and the parameter α_j accounts for the within channel precision. Thus, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\alpha}}$ are the column vectors that contain the parameters $\beta_{i,j}$ and α_j, respectively ($\beta_{i,j}$ represents the precision of the motion compensation error between the HR frames i and j).

The prior in (9) can be rewritten using $m = n$ as [5]

$$p(\tilde{\mathbf{f}}; \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \propto \exp[-\frac{1}{2}\tilde{\mathbf{f}}^T\tilde{\boldsymbol{\Omega}}\tilde{\mathbf{f}}], \quad (10)$$

where $\tilde{\boldsymbol{\Omega}}$ is not given in closed form due to space limitations.

3. MAP PROBLEM FORMULATION AND PROPOSED ALGORITHM

In this section we present a MAP problem formulation for the SR of compressed video, based on the proposed algorithm and on two already existing ones.

3.1. Proposed Model

In the proposed model, the observation term described by (5) and (8), is combined with the new *multichannel* prior expressed by (10). Following [1] (taking into account all possible combinations of the HR motion fields and the compressed LR motion fields) the objective function becomes $J_{MAP}(\tilde{\mathbf{f}}|\tilde{\mathbf{y}}, \tilde{\mathbf{v}}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \propto -2\log[p(\tilde{\mathbf{y}}, \tilde{\mathbf{v}}|\tilde{\mathbf{f}}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\varepsilon}})p(\tilde{\mathbf{f}}; \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})]$,where $\tilde{\mathbf{y}}$, $\tilde{\mathbf{v}}$, $\tilde{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\varepsilon}}$ are the column vectors containing the decoded observations, all possible combinations of the motion vectors provided by the compressed bitstream, all $\delta_{i,j}$ parameters and all ε_i parameters respectively, and its minimization results in

$$\alpha_j = \frac{(LMLN-1)}{\left\|\mathbf{Q}\mathbf{f}_j\right\|^2}, \beta_{i,j} = \frac{LMLN}{\left\|\mathbf{f}_i - \mathbf{D}_{i,j}\mathbf{f}_j\right\|^2}, \quad (11)$$

$$\varepsilon_i = \frac{MN}{\|\mathbf{y}_i - \mathbf{AHf}_i\|^2}, \delta_{i,j} = \frac{MN}{\left\|\mathbf{y}_i^{MV} - \mathbf{AHD}_{i,j}\mathbf{f}_j\right\|^2}, \quad (12)$$

$$(\tilde{\mathbf{G}} + \tilde{\mathbf{\Omega}})\hat{\tilde{\mathbf{f}}} = \tilde{\mathbf{\Lambda}}\tilde{\mathbf{y}}, \quad (13)$$

where
$\tilde{\mathbf{G}} = diag\{(\Gamma_{k-m} + \Theta_{k-m}), ..., (\Gamma_{k+m} + \Theta_{k+m})\}$
and
$\tilde{\mathbf{\Lambda}} = diag\{(\Delta_{k-m} + \Phi_{k-m}), ..., (\Delta_{k+m} + \Phi_{k+m})\}$,
with
$\Gamma_j = \varepsilon_j \mathbf{H}^T\mathbf{A}^T\mathbf{AH}$, $\Theta_j = \sum_{i=k-m}^{k+m} \delta_{i,j}\mathbf{D}_{j,i}\mathbf{H}^T\mathbf{A}^T\mathbf{AHD}_{i,j}$,
$\Delta_j = \varepsilon_j \mathbf{H}^T\mathbf{A}^T$, $\Phi_j = \sum_{i=k-m}^{k+m} \delta_{i,j}\mathbf{D}_{j,i}\mathbf{H}^T\mathbf{A}^T\mathbf{M}_{i,j}$.

In this model the (HR) MF information is also taken into account through the prior, while *simultaneous SR (and restoration) of all the HR frames is taking place* which is not the case in models 1 and 2, which are presented next. Clearly the proposed model is specifically designed for compressed video, given that as can be seen in (12) the motion vector information provided by the compressed bitstream is used, the precision related to the quantization noise is also estimated and all these are incorporated in (13).

3.2. Model 1

The simplest approach to the studied problem is to use a single decoded channel to obtain the observation model of (5). In this case no motion information is used and the video frames are super resolved without using any of the adjacent channels. Combining (5) with the SAR prior model the MAP estimate results in [5]

$$\alpha_i = \frac{(LMLN - 1)}{\|\mathbf{Qf}_i\|^2}, \varepsilon_i = \frac{MN}{\|\mathbf{y}_i - \mathbf{AHf}_i\|^2}, \quad (14)$$

$$\left(\mathbf{H}^T\mathbf{A}^T\mathbf{AH} + \frac{\alpha_i}{\varepsilon_i}\mathbf{Q}^T\mathbf{Q}\right)\hat{\mathbf{f}}_i = \mathbf{H}^T\mathbf{A}^T\mathbf{y}_i, \quad (15)$$

where $\mathbf{A}^T$ defines the up-sampling operation.

3.3. Model 2

This model is based on [1], where the observation model is now described by (7) and (8) (for j=k) and the image prior consists of the within frame SAR prior, as in model 1. Therefore, motion information between video frames is utilized only by the observation model. Consequently, the respective MAP estimate yields

$$\varepsilon_{i,k} = \frac{MN}{\|\mathbf{y}_i - \mathbf{AHD}_{i,k}\mathbf{f}_k\|^2}, \delta_{i,k} = \frac{MN}{\left\|\mathbf{y}_i^{MV} - \mathbf{AHD}_{i,k}\mathbf{f}_k\right\|^2}, \quad (16)$$

$$(\tilde{\mathbf{J}} + \alpha_k\mathbf{Q}^T\mathbf{Q})\hat{\mathbf{f}}_k = \tilde{\mathbf{Z}}, \quad (17)$$

where $\tilde{\mathbf{J}} = \sum_{i=k-m}^{k+m} [(\varepsilon_{i,k} + \delta_{i,k})\mathbf{D}_{k,i}\mathbf{H}^T\mathbf{A}^T\mathbf{AHD}_{i,k}]$ and
$\tilde{\mathbf{Z}} = \sum_{i=k-m}^{k+m} [\varepsilon_{i,k}\mathbf{D}_{k,i}\mathbf{H}^T\mathbf{A}^T\mathbf{y}_i + \delta_{i,k}\mathbf{D}_{k,i}\mathbf{H}^T\mathbf{A}^T\mathbf{M}_{i,k}\mathbf{y}_k]$,
whereas the estimation of parameter α_k is given by the left hand side term of (14) for $i = k$ and $\mathbf{M}_{i,k} = \mathbf{M}(\mathbf{v}_{i,k})$.

Finally, (13), (15) and (17) can not be solved in closed form, due to the non-circulant nature of the matrices $\mathbf{A}, \mathbf{A}^T$, $\mathbf{D}_{i,j}$ and $\mathbf{M}_{i,j}$. Thus, we resorted to a numerical solution using a *conjugate-gradient (CG) algorithm.*

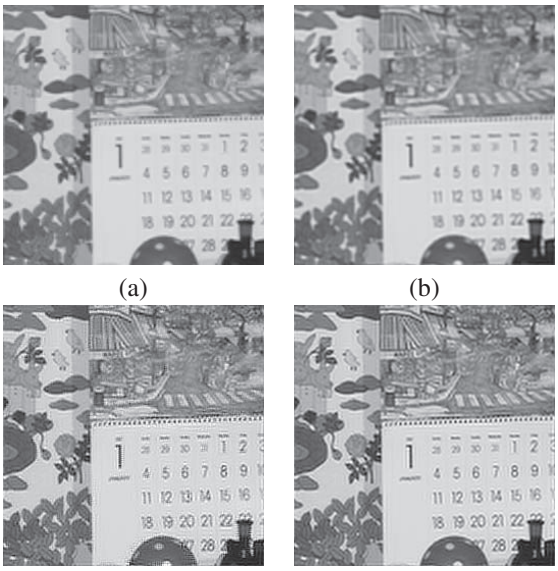
4. EXPERIMENTAL RESULTS

In this section we present numerical experiments to evaluate the proposed method and also justify the benefits provided in compressed video by the use of the new multichannel prior based on (10). In all presented results, the influence of the compression ratio on the SR procedure is also considered, utilizing the H.264 / AVC. The central 256×256 region of first five frames of the sequence "Mobile" (CIF format) was selected for the HR intensities. The temporal rate for the LR sequence is 30 frames per second. Except for the first image in the sequence, which is intra-coded, each frame is compressed as a P-frame. We experimented with two bit-rates 1.63 Mbps and 564 kbps, corresponding to a "high" and "low" quality coding application respectively.

In both bit-rate scenarios the quantization parameter was selected equal to 16. In the "high" quality scenario the input sequence to the encoder was not blurred ($\mathbf{H} = \mathbf{I}$), whereas in the "low" quality case uniform 9×9 blur was used. After blurring, the frames are down-sampled by a factor of two ($L = 2$).

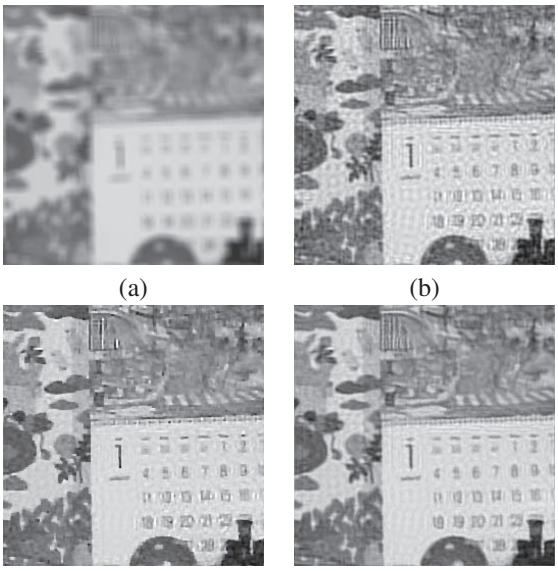
The metrics used to quantify performance was the improvement in signal-to-noise ratio (ISNR) defined (in dB) as $ISNR = 10\log_{10}\left(\|\mathbf{f}_i - \mathbf{y}_{i,I}\|^2 / \left\|\mathbf{f}_i - \hat{\mathbf{f}}_i\right\|^2\right)$, where $\mathbf{y}_{i,I}$ denotes the bicubic interpolation of the ith LR observation and the visual information fidelity (VIF) defined within the range [0,1] [6].

In model 1, denoted by *m1*, the bicubically interpolated LR observations served as initial conditions for the CG algorithm and for the estimation of the parameters. The algorithm in model 2 (denoted by *m2*) and in the proposed model (denoted by *pm*) is initialized by model 1 from which the motion estimation was also performed using a 3-level hierarchical block matching algorithm with integer pixel accuracy at each level. For the CG algorithm implementation, matrices $\mathbf{D}_{i,j}$ were initially estimated and remained fixed (the same holds for the estimated precision parameters). The aforementioned procedure was also used to estimate matrices $\mathbf{M}_{i,j}$ given the decoded LR observations in order to get better estimates for the motion vectors $\mathbf{v}_{i,j}$ with respect to their values given by the compressed bitstream. In all models, the CG iterations are terminated when $\left\|\mathbf{f}_k^{new} - \mathbf{f}_k^{old}\right\|^2 / \left\|\mathbf{f}_k^{old}\right\|^2 < 10^{-6}$.



(a) (b) (c) (d)

Fig. 1. SR results for 1.63Mbps. Result after (a) bicubic interpolation of the LR observation, (b) *m1*, (c) *m2*, (d) *pm*.



(a) (b) (c) (d)

Fig. 2. SR results for 564kbps. Result after (a) bicubic interpolation of the LR observation, (b) *m1*, (c) *m2*, (d) *pm*.

Table 1 shows the ISNR results along with the VIF values among all 3 models for both bit-rate cases, whereas in table 2 the ISNR results are given for all 3 models when uncompressed data is used. All results pertain to the middle frame.

Table 1. ISNR (in dB) and VIF comparison

$Bit-rate$: $1.63Mbps/564kbps$	$ISNR$	VIF
$Model1$	$0.69/2.28$	$0.26/0.19$
$Model2$	$2.21/2.32$	$0.33/0.20$
$Proposed\ Model$	$3.33/2.71$	$0.38/0.23$

Table 2. ISNR (in dB) comparison for uncompressed data

$No\ blur\ case/Un.\ blur\ 9\times 9\ case$	$ISNR$
$Model1$	$0.59/0.95$
$Model2$	$1.89/1.23$
$Proposed\ Model$	$2.83/1.36$

A noteworthy observation is the robustness of the proposed model in terms of both ISNR and VIF in the cases where it is used for compressed data compared to its respective utilization for uncompressed data and the extra gain is due to the removal of the artifacts which are introduced by the compression procedure.

Examples of recovered HR frames are shown in Figures 1 and 2. Clearly, the proposed algorithm outperforms the other two already existing methods. The numbers are sharper, stripes are also improved, while in the high bit-rate scenario the circles on the ball are better defined and the tip of the train is less jagged given the removal of compression artifacts.

5. CONCLUSIONS

In this paper, we presented a MAP approach of a new multichannel image prior applied to the compressed video SR problem, along with 3 algorithms and their comparative study. The experimental results show in all cases that the proposed algorithm performs better than previous ones in terms of both ISNR and VIF, as well as visual quality. Moreover, the comparison between model 2 and the proposed one strongly indicates that the use of MF in the prior term is much more efficient than its use in the observation term, in terms of both restoration capability and resolution enhancement.

6. REFERENCES

[1] C.A. Segall, A.K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Processing*, vol. 13, no. 7, pp. 898–911, 2004.

[2] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Total variation super resolution using a variational approach," in *Proc. ICIP 2008*, pp. 641–644.

[3] S. Farsiu, MD D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[4] M.G. Choi, Y. Yang, and N.P. Galatsanos, "Multichannel regularized recovery of compressed video sequences," *IEEE Trans. Circuits Syst. II: Analog and Digital Signal Processing*, vol. 48, no. 4, pp. 376–387, 2001.

[5] S.P. Belekos, N.P. Galatsanos, and A.K. Katsaggelos, "Maximum a posteriori video super-resolution with a new multichannel image prior," in *Proc. EUSIPCO 2008*, Lausanne, Switzerland, August 25-29, 2008.

[6] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.