BAYESIAN RESOLUTION-ENHANCEMENT FRAMEWORK FOR TRANSFORM-CODED VIDEO

Bahadir K. Gunturk, Yucel Altunbasak, and Russell Mersereau

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, GA 30332-0250

ABSTRACT

Resolution enhancement for video sequences has always been an attractive application in multimedia signal processing. "Superresolution" methods, that combine non-redundant information from a set of low-resolution images, are beginning to be applied to the most popular video compression standard, MPEG. Bayesian approaches, which are very successful for raw video, largely fail for MPEG video, since they do not incorporate the compression process into their models. This compression process introduces quantization noise, which is comparable to the additive noise that is used in the Bayesian models. In this paper we present an analytical derivation that combines the quantization and additive noises in a stochastic framework for MPEG-compressed video. This is a general framework in the sense that different video acquisition models, source statistics, implementation techniques can be used with it.

1. INTRODUCTION

The goal of superresolution is to increase the available resolution for a single frame by using the neighboring frames. There have been many methods proposed to address this problem. Iterative methods [1, 2], Bayesian methods [3, 4], hybrid methods [5], and adaptive filtering [6] are among these methods.

In contrast to the abundance of methods proposed to enhance the raw video, there are only a few methods that have been proposed for MPEG-compressed video. Chen and Schultz [7] advocate decompressing the MPEG video and using the uncompressed-video algorithm given in [4]. The drawback with this method is that decompression discards information about quantization noise that is introduced in the compression process. Patti and Altunbasak [8] suggested another solution to overcome this problem. Their method simply extends the model given in [1] by adding the MPEG stages, and uses the quantization information as the basis for a projections onto convex sets (POCS) algorithm that operates in the compressed domain. However, in this approach all sources of error except for the quantization noise are ignored, which is not a good assumption at medium to high bit rates. It is also difficult with the POCS approach to impose additional enhancement constraints on the reconstructed frame.

In this paper we propose a novel Bayesian framework that addresses all of these problems. Designed for MPEG-compressed video sequences, it takes both the quantization and the additive noise into account. It is cast in a stochastic framework that allows the use of both source statistics and additional reconstruction constraints, such as those that may aid in blocking artifact reduction and edge enhancement. Our procedure uses a maximum likelihood (ML) estimator. It can be extended to maximum *a posteriori* probability (MAP) estimators by simply using a prior image model for the high-resolution image.

In Section II, we briefly develop a general imaging model and extend it by adding stages that model the MPEG compression. Section III introduces our Bayesian framework that can be used with either ML or MAP estimators. Simulation results are given in Section IV. Finally, Section V concludes the paper.

This work was supported by the Office of Naval Research (ONR) under the award N000140110619.

2. IMAGING MODEL

This section extends a general video acquisition model to accommodate MPEG compression. The result is a linear set of equations that relate the high-resolution (HR) image to the quantized DCT coefficients of low-resolution (LR) frames. We use this set of equations to establish the Bayesian framework in the next section.

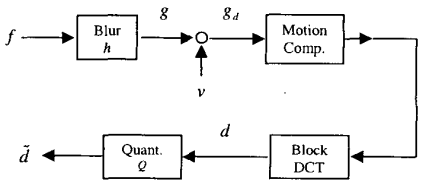

Fig. 1. Imaging model

The starting point is the video capture model shown in Figure 1. In that figure, f denotes the high-resolution scene. The acquisition process is modeled using a linear shift-varying (LSV) operator h plus an additive noise v. The result is the discrete low-resolution frame g_d. The blurring process h may incorporate the motion blur (caused by the relative movement of the LR camera or the changes in the scene), the nonzero sensor aperture time, the nonzero physical dimensions of the individual sensor elements, and the blur caused by the imaging optics [1]. The additive noise v is assumed to be independent of the signal f, which is a fair assumption for most imaging models. The discrete LR frame can be written as:

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

where $(\mathbf{l}, k)$ are the spatio-temporal indices representing the discrete spatial location $\mathbf{l} \equiv (l_1, l_2)$ of the k^{th} frame. h is the blurring function and $f(\mathbf{n}, t)$ is the value of the HR frame at spatial coordinate $\mathbf{n} \equiv (n_1, n_2)$ at time t. $g_d(\mathbf{l}, k)$ is then motion compensated and transformed using a series of 8×8 block-DCTs to result in the DCT coefficients $d(\mathbf{m}, k)$. The index $\mathbf{m} \equiv (m_1, m_2)$ denotes the spatial coordinate in the DCT domain. Defining $\hat{g}(\mathbf{l}, k)$ as the prediction frame and $g(\mathbf{l}, k) \equiv \sum_{\mathbf{n}} f(\mathbf{n}, t) h(\mathbf{n}, t; \mathbf{l}, k)$, we will write:

$$d(\mathbf{m}, k) = DCT\{\, g(\mathbf{l}, k) - \hat{g}(\mathbf{l}, k) + v(\mathbf{l}, k)\}\,, \qquad (2)$$

where $DCT\{\cdot\}$ represents the 8×8 block-DCT. Denoting G, $\hat{G}$, and V as the block-DCTs of g, $\hat{g}$, and v respectively, we will rewrite Equation 2 as:

$$d(\mathbf{m}, k) = G(\mathbf{m}, k) - \hat{G}(\mathbf{m}, k) + V(\mathbf{m}, k). \qquad (3)$$

Using the definition of g, G can be written as a function of $f(\mathbf{n}, t)$:

$$\begin{aligned} G(\mathbf{m}, k) &= DCT\left\{\sum_{\mathbf{n}} f(\mathbf{n}, t) h(\mathbf{n}, t; \mathbf{l}, k)\right\} \\ &= \sum_{\mathbf{n}} f(\mathbf{n}, t) h_{DCT}(\mathbf{n}, t; \mathbf{m}, k) \end{aligned} \qquad (4)$$

where $h_{DCT}(\mathbf{n}, t; \mathbf{m}, k)$ is the 8×8 block DCT of $h(\mathbf{n}, t; \mathbf{l}, k)$.

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 $\mathcal{Q}\{\cdot\}$:

$$\begin{aligned} \tilde{d}(\mathbf{m}, k) &= \mathcal{Q}\{d(\mathbf{m}, k)\} \\ &= \mathcal{Q}\left\{G(\mathbf{m}, k) - \hat{G}(\mathbf{m}, k) + V(\mathbf{m}, k)\right\}. \end{aligned} \qquad (5)$$

In MPEG compression, quantization is realized by dividing each DCT coefficient by a quantization step size followed by rounding to the nearest integer. The quantization step size is determined by the location of the DCT coefficient, the bit rate, and the macroblock mode [9]. The quantized DCT coefficients $\tilde{d}(\mathbf{m}, k)$ and the corresponding step sizes are available at the decoder, *i.e.*, they are either embedded in the compressed bit-stream or specified as part of the coding standard.

Equation 5 is the fundamental equation that represents the relation between the HR image $f(\mathbf{n}, t)$ and the quantized DCT coefficients $\tilde{d}(\mathbf{m}, k)$. In the next section, we discuss how to make use of this equation to establish a Bayesian framework for resolution enhancement.

3. BAYESIAN FRAMEWORK

With the use of Bayesian estimator, not only the source statistics, but also various regularizing constraints can be incorporated in the solution. Bayesian estimators have been used frequently for superresolution [3, 4, 7]. However, in those approaches either the video source is assumed to be available in uncompressed form, or it is simply decompressed without considering the quantization process. Our model takes both the quantization noise and the additive noise into account. In this section we will model the noise statistics and establish the MAP solution, which generalizes the ML solution.

In the MAP formulation, the noise, the quantized DCT coefficients $\tilde{d}(\mathbf{m}, k)$, and the original frame $f(\mathbf{n}, t)$ are all assumed to be random processes. The MAP estimate $\hat{f}$ is given by:

$$\hat{f} = \arg\max_{f} \left\{p_{f|\tilde{d}}(f|\tilde{d})\right\} \qquad (6)$$

Using the Bayes' rule, Equation 6 can be rewritten as:

$$\begin{aligned} \hat{f} &= \arg\max_{f} \left\{\frac{p_{\tilde{d}|f}(\tilde{d}|f) p_f(f)}{p_{\tilde{d}}(\tilde{d})}\right\} \\ &= \arg\max_{f} \left\{p_{\tilde{d}|f}(\tilde{d}|f) p_f(f)\right\}, \end{aligned} \qquad (7)$$

where we have used the fact that $p_{\tilde{d}}(\tilde{d})$ is not a function of f.

Clearly we need to model the conditional PDF $p_{\tilde{d}|f}(\tilde{d}|f)$ and the prior PDF $p_f(f)$ in order to find the MAP estimate $\hat{f}$. The conditional PDF $p_{\tilde{d}|f}(\tilde{d}|f)$ is derived in this section to establish a general framework that can be used with any prior image models.

Given the frame f, the only random variable on the right-hand side of Equation 5 is the DCT of the additive noise, $V(\mathbf{m}, k)$. In order to find the PDF of $V(\mathbf{m}, k)$, we first need to know about the statistics of the additive noise process $v(\mathbf{m}, k)$. A reasonable assumption for $v(\mathbf{m}, k)$ is that of a zero-mean independent, identically distributed (IID) Gaussian process [4]. Under this assumption the PDF of the additive noise at each location $(\mathbf{m}, k)$ is:

$$p_{v(\mathbf{m},k)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}, \qquad (8)$$

where σ^2 denotes the noise variance. It is straightforward to show that the PDF of $V(\mathbf{m}, k)$ is also a zero-mean IID Gaussian random variable with the same variance σ^2. (A sketch of the proof is provided in the Appendix.) Therefore,

$$p_{V(\mathbf{m},k)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}. \qquad (9)$$

From Equation 3, it follows that the PDF of $d(\mathbf{m}, k)$ is also an IID Gaussian process, but in this case one with mean $G(\mathbf{m}, k) - \hat{G}(\mathbf{m}, k)$. That is,

$$p_{d(\mathbf{m},k)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x - G(\mathbf{m},k) + \hat{G}(\mathbf{m},k))^2/2\sigma^2}. \qquad (10)$$

To compute the PDF of the quantized DCT coefficients $\tilde{d}(\mathbf{m}, k)$ we will borrow from quantization theory. It is shown in [10] that, under some conditions, the PDF of the quantized signal can be computed using a two-step process: first, we evaluate the convolution of the input signal PDF and a rectangular pulse (with an area of one and a width equal to the quantization step size), and then we multiply the result of this convolution with an impulse train. Thus,

$$\begin{aligned} &p_{\tilde{d}(\mathbf{m},k)|f}(x) = \\ &\quad \left(p_{n(\mathbf{m},k)}(x) * p_{d(\mathbf{m},k)}(x)\right) \sum_{r=-\infty}^{\infty} \delta\left(x - r\Delta(\mathbf{m}, k)\right), \end{aligned} \qquad (11)$$

where $\Delta(\mathbf{m}, k)$ is the quantization step size, and $p_{n(\mathbf{m},k)}(x)$ is the rectangular pulse:

$$p_{n(\mathbf{m},k)}(x) = \begin{cases} 1/\Delta(\mathbf{m}, k), & |x| < \Delta(\mathbf{m}, k)/2 \\ 0, & otherwise \end{cases}. \qquad (12)$$

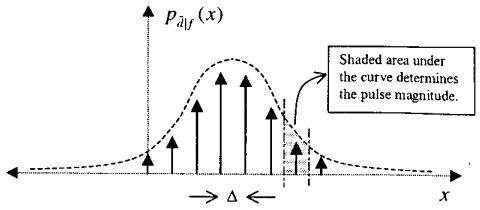

Fig. 2. Conditional PDF.

Equation 11 is valid under the conditions that the quantizer has a uniform step size, and there is no saturation. The former condition is valid for MPEG as well as many other compression standards. For each DCT coefficient, there is a uniform step size that is determined by the location of the DCT coefficient, and the compression rate of the encoder. The latter condition is also satisfied for all DCT coefficients.

Substituting Equations 10 and 12 into Equation 11, we get:

$$p_{\tilde{d}(\mathbf{m},k)|f}(x) = \left\{ \frac{1}{\sqrt{2\pi}\sigma\Delta} \int_{x-\Delta/2}^{x+\Delta/2} e^{\frac{-\left(u-G+\hat{G}\right)^2}{2\sigma^2}} du \right\} \sum_{r=-\infty}^{\infty} \delta(x - r\Delta), \quad (13)$$

where the dependence on $(\mathbf{m}, k)$ is dropped from Δ, G, and $\hat{G}$ to simplify the notation.

Equation 13 implies that the conditional PDF is an impulse train with magnitudes determined by the *areas* (of the Gaussian function) within the $\Delta/2$ neighborhood of the impulse locations. (See Figure 2.) This can also be considered as *area sampling* [10].

Since the quantization noise and the additive noise (hence the overall noise) corresponding to different coefficients are assumed to be independent, the joint PDF will be the product of the individual PDFs. This provides an analytical relation that could be used with different prior image models for Bayesian estimation. Although computationally complex, it provides an optimal way of combining the additive and the quantization noises in a Bayesian framework.

4. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the proposed method we performed the following experiment. The test image AERIAL (Figure 3) was blurred, corrupted by additive noise with gaussian distribution, jittered, and downsampled to create multiple low-resolution frames. The low-resolution frames were then MPEG compressed to produce the low resolution input with a video sequence. Figure 4 shows one of those low resolution frames. The proposed Bayesian framework is implemented using the Iterated Conditional Modes technique [11] for uniform image prior. For the blurring process h, we adopted the model given in [1], which is one the most comprehensive models for the video acquisition process. (The details of the implementation will be presented in a forthcoming paper.) The reconstructed frame is given in Figure 5.

5. CONCLUSION AND FUTURE WORK

In this paper we presented a multi-frame resolution-enhancement framework that incorporates both the quantization and the additive noises in a stochastic framework for transform-coded video. Although the MPEG compression standard was emphasized throughout the paper, the framework is still valid for all video coding standards where the transform utilized is linear. This framework can be used with different video acquisition models, source statistics, implementation techniques.

However, there are still open issues for superresolution reconstruction. For software implementations, fast, but perhaps suboptimal, solutions need to be investigated. For hardware implementations, algorithm parallelization needs to be examined. Since accurate motion estimation is a crucial requirement SR algorithms, we also need to deal with the regions that have inaccurate motion estimates because of motion model failure.

6. APPENDIX

In this Appendix we provide a sketch of the proof that the block-DCT of a zero-mean IID Gaussian random vector is also a zero-mean IID Gaussian random vector with the same variance.

Let $v_1, \ldots, v_n$ be a zero-mean IID Gaussian vector with each component having the variance σ^2. n is the size of the vector, which is 64 for an 8 × 8 block. Thus,

$$p_{v_i}(v_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-v_i^2/2\sigma^2}, \quad i = 1, \ldots, n. \quad (14)$$

Similarly, let $V_1, \ldots, V_n$ be the DCT coefficients of that block, obtained by taking the 8 × 8 block-DCT. Each DCT coefficient V_i will be a function of $v_1, \ldots, v_n$:

$$V_i = F_i(v_1, v_2, \ldots, v_n), \quad i = 1, \ldots, n, \quad (15)$$

where F_i calculates V_i as a linear combination of $v_1, \ldots, v_n$ according to the block-DCT transform.

The joint PDF of the DCT coefficients $V_1, V_2, \ldots, V_n$ can be found by

$$p_{V_1,V_2,\ldots,V_n}(V_1, V_2, \ldots, V_n) = p_{v_1,v_2,\ldots,v_n}(v_1, v_2, \ldots, v_n)\,|\mathbf{J}|^{-1}, \quad (16)$$

where $\mathbf{J}$ is the Jacobian defined as:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial F_1}{\partial v_1} & \cdots & \frac{\partial F_1}{\partial v_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial v_1} & \cdots & \frac{\partial F_n}{\partial v_n} \end{bmatrix}. \quad (17)$$

Since $|\mathbf{J}|$ is constant (=1) for the DCT, the joint PDF of $V_1, \ldots, V_n$ will be:

$$p_{V_1,V_2,\ldots,V_n}(V_1, V_2, \ldots, V_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(v_1^2+v_2^2+\cdots+v_n^2)/2\sigma^2}. \quad (18)$$

Next, we should write $v_1, v_2, \ldots, v_n$ in terms of $V_1, V_2, \ldots, V_n$. Since the DCT is a unitary transformation, $v_1^2 + v_2^2 + \cdots + v_n^2 = V_1^2 + V_2^2 + \cdots + V_n^2$. As a result, the joint PDF of V is:

$$p_{V_1,V_2,\ldots,V_n}(V_1, V_2, \ldots, V_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(V_1^2+V_2^2+\cdots+V_n^2)/2\sigma^2}. \quad (19)$$

Since the joint PDF is a separable function of V_i, each coefficient will be independent of the others, and have a PDF equal to:

$$p_{V_i}(V_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-V_i^2/2\sigma^2}, \quad i = 1, \cdots, n. \quad (20)$$

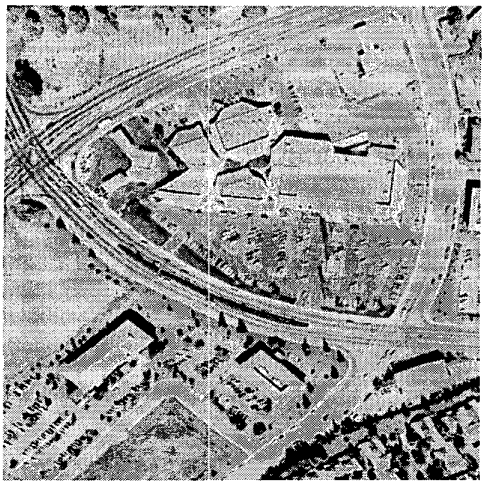

Fig. 3. Original AERIAL image.

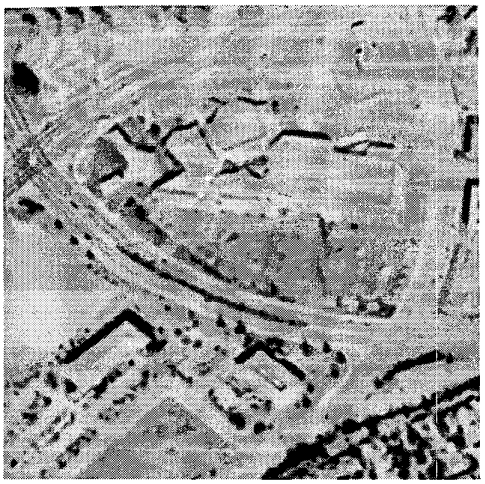

Fig. 4. A frame from the synthetic AERIAL video. (Bilinearly interpolated.)

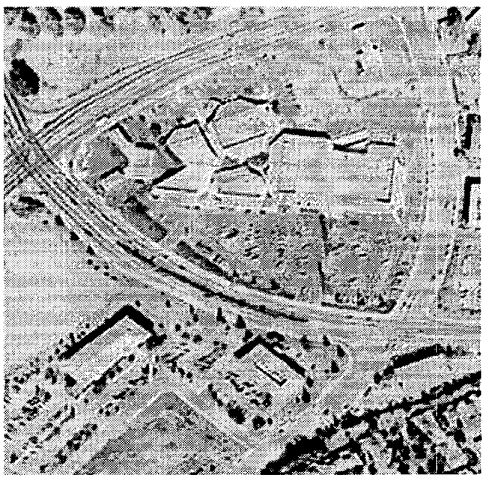

Fig. 5. Reconstructed frame

7. REFERENCES

[1] 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, no. 8, pp. 1064–1076, August 1997.

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

[3] P.Cheeseman, B.Kanefsky, and R.Hanson, "Super-resolved surface reconstruction from multiple images," *Tech. Rep., NASA*, Jan. 1993.

[4] R.R.Schultz and R.L.Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.

[5] M.Elad and A.Feuer, "Restoration of a single superresolution image from several blurred, noisy and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.

[6] M.Elad and A.Feuer, "Superresolution restoration of an image sequence: adaptive filter approach," *EEE Trans. Image Processing*, vol. 8, no. 3, pp. 387–395, Mar. 1999.

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

[8] A.J.Patti and Y.Altunbasak, "Super-resolution image estimation for transform coded video with application to mpeg," *Proc. IEEE Int. Conf. Image Processing*, vol. 3, pp. 179–183, 1999.

[9] A. M. Tekalp, *Digital Video Processing*, Prentice Hall, 1995.

[10] B.Widrow, I.Kollar, and M.C.Liu, "Statistical theory of quantization," *IEEE Trans. Instrumantation and Measurement*, vol. 45, no. 2, pp. 353–361, April 1996.

[11] J.Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.