Journal of Information Technology and Engineering
Vol. 3 No. 1 (June, 2018)

Super-Resolution Reconstruction of Compressed Video based on Various Noise Models

Zhong-Qiang XU and Xiu-Chang ZHU

Lab of Image Processing and Communication, Nanjing University of Posts & Telecommunications Nanjing, Jiangsu Province, P.R. China *E-mail: jstvxzq@sohu.com*

Received: 03rd June 2017 Revised: 14th August 2017 Accepted: 01st November 2017

Abstract: *Super-resolution (SR) technique is to estimate High-resolution (HR) images from a sequence of Low-resolution (LR) observations. This paper proposes a Bayesian SR reconstruction algorithm that models the process of video compression, DCT quantization noise and motion noise by exploiting the quantization step size and motion information embedded in the bit-stream. The proposed total noise term adaptively adjusts for different quantizers. With a Huber-Markov Random Field (HMRF) as the prior model, a Bayesian framework for MAP reconstruction and the gradient descent algorithm are presented and its performance are also analyzed. Simulation results show that proposed algorithm obtains better performances in both objective and subjective quality, which is applicable for compressed videos.*

Key words: *Super-Resolution, MAP Estimation, HMRF, DCT Quantization Noise, Motion Estimation*

1. INTRODUCTION

SR algorithms increase the resolution of an image by exploiting the underlying motion of a video sequence to provide multiple observations for each frame. The idea was first addressed by Tsai [1] and more realistic approaches can be divided into two distinct groups. One group uses deterministic methods, such as Projections onto Convex Sets (POCS) [2–3]. The second group of methods is based on a statistical formulation, such as a maximum likelihood (ML) or Maximum a Posteriori (MAP) probability estimate [4–5]. All of these methods are based on the assumption that the LR images are available in the spatial domain.

In this paper, we focus on SR from a video source that is available in a compressed format such as MPEG, H.263 or DV. These compression systems introduce several disparities to the SR approach. Firstly, the LR observations are compressed bitstream instead of a sequence of intensity images. This bitstream describes the original image sequence as a combination of quantized transform coefficients and motion vectors, and it introduces other departures into the SR problem. 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. There are several Bayesian algorithms that are designed for compressed video. In [6], the algorithm is designed to penalize any artifacts formed during the quantization process. In [7], the algorithm proposes to compute the joint statistics of the spatial quantization and additive noises. In [8], the quantization operator is incorporated into a SR procedure, in which all necessary displacement values are assumed known.

In this paper, we propose a Bayesian SR reconstruction technique by using MAP estimation with a HMRF as a prior model. We model the process of compression and exploit the quantization step size and motion vector information available in the bit-stream. The SR formulation relies on the Bayesian framework, and it incorporates both the transform coefficients and motion vectors from the compressed bit-stream. The proposed noise model automatically adjusts for different quantizers and it is a much more realistic model than assuming IID noise model throughout each frame. The method also uses the source statistics and additional reconstruction constraints, such as those that might aid in blocking artifact reduction and edge enhancement.

The paper is organized as follows. In Section 2, we present a general system model, including video compression, DCT quantization noise, motion estimation noise and a HMRF prior image model. Section 3 provides a Bayesian framework for the SR reconstruction and the gradient descent algorithm for its solution. Experimental results are presented in Section 4. Finally, we discuss conclusions and the future work in Section 5.

2. SYSTEM MODEL

An accurate system model is the key to the SR approach, which must be developed to exploit the information available in the compressed bit-stream. In this section, we formulate the system model of SR reconstruction of compressed video.

2.1 Video Compression Model

Assume that a sequence of L LR frames is available, with g_k designated as the reference frame. The idea is to extract additional knowledge about the unknown HR frame f_k from neighboring LR frames g_l, for $l \neq k$. The video observation model must consider motion occurring between g_l and the reference frame g_k. This model [9] is given as

$$g_l = \mathbf{W}_{l,k}\mathbf{f}_k + \mathbf{n}_{l,k} \tag{1}$$

where $\mathbf{W}_{l,k}$ is the motion-compensated sub-sampling matrix that taking into account the motion between frames, $\mathbf{n}_{l,k}$ is an independent and identically distributed (IID) Gaussian additive noise. When each HR image frame is of dimension $qM \times qN$, then g_l is of dimension $M \times N$ and $\mathbf{W}_{l,k}$ has dimension $qMqN \times qMqN$, where q represents the resolution enhancement factor.

We now add the MPEG compression stages to this model [10]. The LR frame is motion compensated (i.e., the prediction frame is computed and subtracted from the original to get a residual image), and the residual is transformed using a series of block-DCT to produce the DCT coefficients. Defining D as the DCT matrix and $\hat{g}_l$ as the prediction frame which is obtained using neighboring frames (except for the case of intra-coded frames, where the prediction frame is zero), we write

$$\mathbf{y}_l = \mathbf{D}\mathbf{W}_{l,k}\mathbf{f}_k - \mathbf{D}\hat{g}_l + \mathbf{D}\mathbf{n}_{l,k} \tag{2}$$

$\mathbf{y}_l$ are then quantized to produce the quantized DCT coefficients. 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. The quantization operation is a nonlinear process that will be denoted by the operator Q

$$\mathbf{y}_l^q = Q\left\{\mathbf{D}\mathbf{W}_{l,k}\mathbf{f}_k - \mathbf{D}\hat{g}_l + \mathbf{D}\mathbf{n}_{l,k}\right\} \tag{3}$$

The quantized DCT coefficients $\mathbf{y}_l^q$ 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.

2.2 DCT Quantization Noise

The quantization error is a deterministic quantity that is defined as the difference between $\mathbf{y}_l^q$ and $\mathbf{y}_l$, but it can also be treated as a stochastic vector. There have been a number of studies directed toward modeling the statistical distribution of the quantization error [11]. Since the quantization takes place in the transform domain, the natural way to exploit this information is to use it in the DCT domain without reverting back to the spatial domain. Denoting $\mathbf{e}_l$ as the quantization error, we can write

$$\mathbf{y}_l^q = \mathbf{y}_l + \mathbf{e}_l^q \tag{4}$$

where $$\mathbf{e}_l^q \sim N(0, \mathbf{K}_{Q,l}) \tag{5}$$

and $K_{Q,l}$ is the covariance matrix of the quantization noise in the spatial domain at frame l. Defining the covariance matrix, $K_{Q,l}$ 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

$$\mathbf{K}_{Q,l} = E\left[\mathbf{e}_l^q\left(\mathbf{e}_l^q\right)^t\right] = \mathbf{D}^{-1}\mathbf{K}_{T,l}\mathbf{D} \tag{6}$$

where $\mathbf{K}_{T,l}$ is the autocovariance matrix of the noise in DCT domain.

Due to the de-correlating properties of the DCT for typical images, the autocovariance matrix of $\mathbf{y}_l^q$ is approximately diagonal, $\mathbf{K}_{T,l}$ can be shown to be diagonal as well. The diagonal elements of $\mathbf{K}_{T,l}$ represent the variances of the DCT-domain quantization errors. 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 [12]. Whenever the assumption of a uniform distribution is viable, the noise variance for transform index is defined completely by the bitstream and expressed as

$$\sigma_l^2 = \frac{q_l^2}{12} \tag{7}$$

where q_l is the quantizer step-size. Thus the matrix $\mathbf{K}_{T,l}$ is easily constructed based on the quantization limits defined by the quantizer. It is interesting to note that when the diagonal entries of $\mathbf{K}_{T,l}$ are equal, the resulting covariance matrix describes an IID noise process in the transform domain. For the large class of linear transforms, 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 $\mathbf{K}_{T,l}$ 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 DCT noise in this case is IID in both the transform and spatial domains.

From Eq. (6), each spatial-domain noise term is a linear combination of independently distributed uniform random variables, allowing the spatial-domain quantization noise to be approximated as a 0-mean Gaussian random vector with autocovariance matrix $\mathbf{K}_{Q,l}$ as following

$$p\left(\mathbf{e}_l^q\right) = p\left(\mathbf{y}_l^q \middle| \mathbf{f}_l\right) = \frac{1}{(2\pi)^{\frac{MN}{2}} \left|\mathbf{K}_{Q,l}\right|} \exp\left\{-\frac{1}{2}\left(\mathbf{e}_l^q\right)^t \mathbf{K}_{Q,l}^{-1} \mathbf{e}_l^q\right\} \quad (8)$$

2.3 Motion Estimation Noise

In addition to the quantization intervals, information about the sub-pixel displacements also appears in the compressed bit-stream. This data is encapsulated in the motion vectors that provide a noisy observation of the original displacements [13]. We define the relationship between the motion compensated prediction and the reference image frame as

$$\mathbf{y}_l = \mathbf{y}_k + \mathbf{e}_{l,k}^m \quad (9)$$

where $\mathbf{e}_{l,k}^m$ is the error that accounts for the uncertainty in estimating $\mathbf{y}_l$ from $\mathbf{y}_k$. It is assumed that the motion noise $\mathbf{e}_{l,k}^m$ is IID Gaussian noise [4] Gaussian with autocovariance matrix $\mathbf{K}_{l,k}^M$, which for the case at hand is rather limited—IID noise implies that each observation is equally reliable, and does not take into consideration spatially-varying errors in motion estimation such as appearing/disappearing objects, lighting changes, or incorrect motion vectors. Here, we propose a spatially-varying motion noise model. Suppose first that $\mathbf{K}_{l,k}^M$ is diagonalized by the transformation matrix $\mathbf{U}$, such that

$$\mathbf{K}_{l,k}^M = \mathbf{U}^T \mathbf{K}^M \mathbf{U} \quad (10)$$

where $\mathbf{K}^M$ is a diagonal matrix. Thus, with knowledge of $\mathbf{U}$ the autocovariance of $\mathbf{e}_{l,k}^m$ is represented by the N elements of $\mathbf{K}^M$ rather than the N^2 elements of $\mathbf{K}_{l,k}^M$, where N is the number of pixels in the image. Two-dimensional signals that are well-modeled by first-order prediction models with high one-step correlation parameters have autocovariance matrices that are approximately diagonalized by the DCT [14]. Assuming such a signal leads to the approximation.

$$\mathbf{K}_{l,k}^M = \mathbf{D}^T \mathbf{K}^M \mathbf{D} \quad (11)$$

where $\mathbf{D}$ is the block-DCT, which is the DCT applied on a block-by-block basis. It is argued that for most situations using the BDCT provides a more accurate description of the noise than an IID assumption. Furthermore, when performing at relatively low bit rates, motion-compensation errors that are not approximately de-correlated by the BDCT—where the motion error is quite small—will be dominated by the BDCT quantization noise.

The variances $\sigma_{l,k}^2$ that compose the diagonal matrix $\mathbf{K}^M$ must be determined. Assuming it is proportional to the distance between frames [15], the required terms are found here as

$$\sigma_{l,k}^2 = |l-k| / \beta \quad (12)$$

where β is a parameter.

2.4 Image Prior Model

We now need to model the prior distribution to complete the MAP formulation. The prior image model is chosen as a HMRF model, which has been used extensively in image and video processing [16]. The Huber function is a convex function that has edge-preserving properties relative to a simple quadratic. The HMRF is given here without excessive discussion,

$$p(\mathbf{f}) = \frac{1}{G} \exp\left\{-\lambda \sum \rho_\alpha\left(\mathbf{d}_x^t \mathbf{f}\right)\right\} \quad (13)$$

where G is a normalizing constant known as the partition function, λ is a regularization parameter, and x is a local group of pixels contained within the set of all image cliques $\mathbf{X}$. The quantity $\mathbf{d}_x^t \mathbf{f}$ is an activity measure, with a small value in smooth image locations and a large value near edges. Four spatial activity measures are computed at each pixel in the HR image, given by the following second-order finite differences:

$$\begin{aligned}
\mathbf{d}_{x,1}^t \mathbf{f} &= \mathbf{f}_{x_1,x_2-1} - 2\mathbf{f}_{x_1,x_2} + \mathbf{f}_{x_1,x_2+1} \\
\mathbf{d}_{x,2}^t \mathbf{f} &= 0.5\mathbf{f}_{x_1+1,x_2} - \mathbf{f}_{x_1,x_2} + 0.5\mathbf{f}_{x_1-1,x_2+1} \\
\mathbf{d}_{x,3}^t \mathbf{f} &= \mathbf{f}_{x_1-1,x_2} - 2\mathbf{f}_{x_1,x_2} + \mathbf{f}_{x_1+1,x_2} \\
\mathbf{d}_{x,4}^t \mathbf{f} &= 0.5\mathbf{f}_{x_1-1,x_2-1} - \mathbf{f}_{x_1,x_2} + 0.5\mathbf{f}_{x_1+1,x_2+1}
\end{aligned}$$

The Huber function $\rho_\alpha(\mu)$ is defined as

$$\rho_\alpha(\mu) = \begin{cases} \mu^2 & |\mu| \le \alpha \\ \alpha^2 + 2\alpha\left(|\mu| - \alpha\right) & |\mu| > \alpha \end{cases}$$

Using the HMRF encourages smoothness in the final SR reconstruction, since the probability in Eq. (13) is higher for smoother images. The Huber function penalizes differences less than ± quadratically; however, differences larger than ± are only penalized linearly, which helps prevent oversmoothing image edges. In this work, the HMRF smoothes compression artifacts introduced from the received frames, preventing their presence in the reconstructed frames. The prior image model also fills in gaps in the image that may result when no motion information is present.

3. BAYESIAN SR RECONSTRUCTION

In the MAP formulation, the quantized DCT coefficients, motion estimation noise and the original HR frame are all assumed to be random processes. Denoting $p\left(\mathbf{f}\left|\mathbf{y}^q\right.\right)$ as the conditional probability density function, the MAP estimate is given by

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}}\left\{p\left(\mathbf{f}\left|\mathbf{y}^q\right.\right)\right\} \tag{16}$$

Using the Bayes rule, Eq. (16) can be rewritten as

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}}\left\{p\left(\mathbf{f}\left|\mathbf{y}^q\right.\right)p\left(\mathbf{f}\right)\right\} \tag{17}$$

where we use the fact that $p\left(\mathbf{y}^q\right)$ is independent of $p\left(\mathbf{f}\right)$.

In section 2, we have modeled $p\left(\mathbf{f}\left|\mathbf{y}^q\right.\right)$ and $p\left(\mathbf{f}\right)$, thus the final optimization problem to be solved becomes

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}}\left\{\lambda\mu\left(\mathbf{f}\right)+v\left(\mathbf{f}\right)\right\} \tag{18}$$

where

$$\mu\left(\mathbf{f}\right)=\sum_{x\in\mathbf{X}}\rho_\alpha\left(\mathbf{d}_x^t\mathbf{f}\right) \tag{19}$$

$$v\left(\mathbf{f}\right)=\frac{1}{2}\sum_{l\in L}\mathbf{e}_l^t\mathbf{K}_l^{-1}\mathbf{e}_l=\sum_{l\in L}\left(\mathbf{W}_l\mathbf{f}-\mathbf{y}_l^q\right)^{-1}\mathbf{K}_l^{-1}\left(\mathbf{W}_l\mathbf{f}-\mathbf{y}_l^q\right) \tag{20}$$

As discussed above, it is known that $\mathbf{y}_l^q=\mathbf{y}_l+\mathbf{e}_l$, combining with Eq. (9), we can write

$$\mathbf{y}_l^q=\mathbf{y}_l+\mathbf{e}_l^q=\mathbf{y}_k+\mathbf{e}_{l,k}^m+\mathbf{e}_l^q=\mathbf{y}_k+\mathbf{e}_l \tag{21}$$

where $\mathbf{e}_l=\mathbf{n}_{l,k}^m+\mathbf{e}_l^q$ consists of both DCT quantization noise and motion estimation noise, thus $\mathbf{K}_l=\mathbf{K}_{Q,l}+\mathbf{K}_{l,k}^M$. After substituting for $\mathbf{K}_l$, the gradients of the individual terms are given as

$$\nabla\mu\left(\mathbf{f}\right)=\sum_{x\in\mathbf{X}}\mathbf{d}_x\rho_\alpha^{'}\left(\mathbf{d}_x^t\mathbf{f}\right) \tag{23}$$

$$\nabla v\left(\mathbf{f}\right)=\sum_{l\in\mathbf{L}}\mathbf{W}_l^t\mathbf{D}^t\left(\mathbf{K}_{T,l}+\mathbf{K}_l^m\right)^{-1}\mathbf{D}\left(\mathbf{W}_l\mathbf{f}-\mathbf{y}_l^q\right) \tag{24}$$

The gradient term in Eq. (22) has larger values for large differences in $\mathbf{d}_x^t\mathbf{f}$, thus encouraging smoothness. Note, however, that the first derivative of the Huber function has a maximum magnitude of 2α, which effectively prevents the gradient from becoming too large, and thus prevents excessive smoothing of image edges in the final result. The effect of Eq. (23) is such that DCT frequency components that have larger variance do not affect the gradient as much as DCT frequency components with lower variance. Thus, the proposed noise model automatically adjusts for different quantizers—if the quantization parameters decrease for certain regions of the frame, the model accounts for this by having lower quantization noise for those regions. Similarly, if the quantization parameters change from frame to frame, the quantization noise for each frame follows accordingly. This is a much more realistic model than assuming IID noise throughout each frame.

The convex optimization problem in Eq. (18) is solved using a gradient descent algorithm [17]. Denoting the estimate of $\mathbf{f}$ at iteration n as $\mathbf{f}^{(n)}$, and $g\left(\mathbf{f}\right)=\lambda\nabla\mu\left(\mathbf{f}\right)+\nabla v\left(\mathbf{f}\right)$, the gradient descent algorithm forms the new estimate as

$$\mathbf{f}^{(n+1)}=\mathbf{f}^{(n)}-\tau^{(n)}g\left(\mathbf{f}^{(n)}\right) \tag{24}$$

where $\tau^{(n)}$ is a step size that ideally reduces the objective function by as much as possible. By taking the derivative with respect to $\tau^{(n)}$ and manipulating terms, the step size for the n-th iteration is given by [10]

$$\tau^{(n)}=\frac{-\nabla\mu\left(\mathbf{f}^{(n)}\right)^t\mathbf{d}_x^{(n)}}{\left(\mathbf{d}_x^{(n)}\right)^t\nabla^2\mu\left(\mathbf{f}^{(n)}\right)\mathbf{d}_x^{(n)}} \tag{25}$$

This value for the step size $\tau^{(n)}$ gives the maximum decrease in a local region of the function for each iteration, ensuring global convergence in a reasonable number of iterations.

4. SIMULATIONS

In order to compare the performance of the proposed SR algorithm, two image sequences were used in the simulations. The first image is synthetically generated from a single digital image *mobile* sequence #5 frame to model a diagonal camera pan with known motion vectors. The second image sequence is captured by a camera with a slight panning. This sequence is composed of several objects moving independently. Both image sequences are are first sub-sampled by a factor of q (q = 2) and then compressed at 1Mbps with a MPEG-4 codec. The quantitative comparisons could be made using the improved signal-to-noise ratio (ISNR). The ISNR is a quantitative measure of how much the estimated frame $\hat{\mathbf{f}}^{(k)}$ has improved over the reference frame, given as

$$ISNR = 10\log\frac{\left\|\mathbf{f}^{(k)} - \mathbf{f}_0^{(k)}\right\|^2}{\left\|\mathbf{f}^{(k)} - \hat{\mathbf{f}}^{(k)}\right\|^2} \text{ (dB)} \quad (26)$$

In this expression, $\mathbf{f}_0^{(k)} = q^2\mathbf{W}^T\mathbf{y}^{(k)}$, $\mathbf{f}^{(k)}$ is the original HR image, and $\hat{\mathbf{f}}^{(k)}$ is the estimated HR image.

Estimating accurate subpixel-resolution motion vectors is a critically important component of modeling an image sequence for use in SR reconstruction algorithms [18]. The goal is to compute the motion fields as quickly and as accurately as possible. In full-search subpixel block matching, the search area must be increased in size by a factor of q over integer-resolution block matching, and the motion vectors are estimated using only the decoded, interpolated frames. This is a relatively accurate, albeit computationally expensive, method of estimating subpixel displacements. Since the MPEG encoder provides half-pel motion displacements in the compressed bit stream, a suboptimal search can be performed in a region centered at each MPEG motion vector, rather than by conducting an exhaustive search over all possible displacements. To significantly reduce the number of computations, the decoded half-pel MPEG motion vectors may be used as initial conditions, with a smaller search area centered at each up-sampled MPEG motion vector. For a frame interpolation factor of q, the half-pel MPEG motion fields must be up-sampled by a factor of $q/2$. Recall that a single MPEG vector represents the displacement for all 16×16 LR pixels contained within a macroblock. The up-sampled MPEG motion vector is used as the initial condition for all pixels within its macroblock, and an individual subpixel displacement is estimated for each block of $q\times q$ HR pixels.

Since inaccurate motion vectors can cause a great deal of damage to the SR enhanced frame, these displacements are detected and the corresponding pixels eliminated from the video observation model. The displaced frame difference (DFD) is computed between up-sampled frame $\mathbf{y}_l$ for $l \neq k$ and the compensated image constructed from the up-sampled I-picture and the estimated subpixel motion vectors. In regions where the DFD is large, the block matching displacements are generally inaccurate due to either a lack of edges within the data or pixel occlusions. To accommodate for these errors, pixels within $\mathbf{y}_l$ which are not also observable in the I-picture are ignored in the SR algorithm.

The procedure used to compute the SR enhanced estimates is as follows:

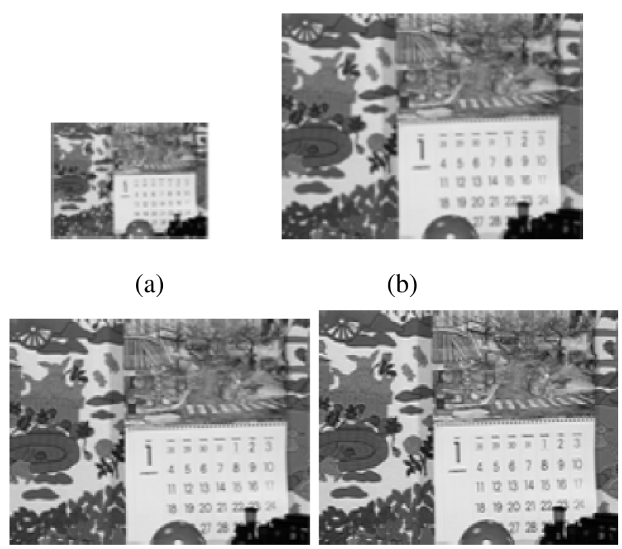
1. Identify the location of the I-picture to be enhanced in the MPEG sequence, and save the neighboring P- and B-pictures and decoded motion fields;
2. Up-sample both LR frames $\mathbf{y}_k$ and $\mathbf{y}_l$ by a factor of $q = 2$, using bilinear interpolation ,and perform reduced-search subpixel block matching on the up-sampled video frames using the half-pel motion fields as initial conditions;
3. Detect inaccurate motion vectors and eliminate the corresponding pixels from the video observation model by examining the DFD. Then, down-sample the displacement vectors by averaging $q \times q$ vector blocks.
4. Estimate the HR frame using decoded frames and subpixel motion fields.

In all experiments, the Huber threshold parameter was set to $\alpha = 1.0$, $\lambda = 0.00075$, $\beta = 10$, and the total frame number $L = 4$. Fig. 1 and 2 provide a visual comparison of reconstruction results using the SR techniques.

Fig.1 (a) is the LR image and Fig.1 (b) is the up-sampled image synthetic sequence. Fig.1 (c) represents the SR result (ISNR = 5.24dB) using all estimated motion vectors(MVs), while Fig.1 (d) denotes that the inaccurate motion vector estimates have been detected and eliminated prior to the application of the SR algorithm (ISNR=6.13dB.)

Fig. 2 (a) is the LR image and Fig. 2 (b) is the up-sampled image for actual sequence. Fig. 2(c) is of ISNR = 3.14dB while Fig. 2 (d) with ISNR = 3.63dB.

From the synthetic and actual image sequences, we can see that the coding artifacts are attenuated by proposed SR algorithm with visual quality improvement throughout the frame by retaining detailed information. We also note that in most cases, the ISNR is improved after inaccurate motion vectors have been detected and eliminated.



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

Figure 1: Results from Synthetic Sequence (a)LR image (b)Up-Sampled Image(c) SR with all MVs (d) SR with Accurate MVs

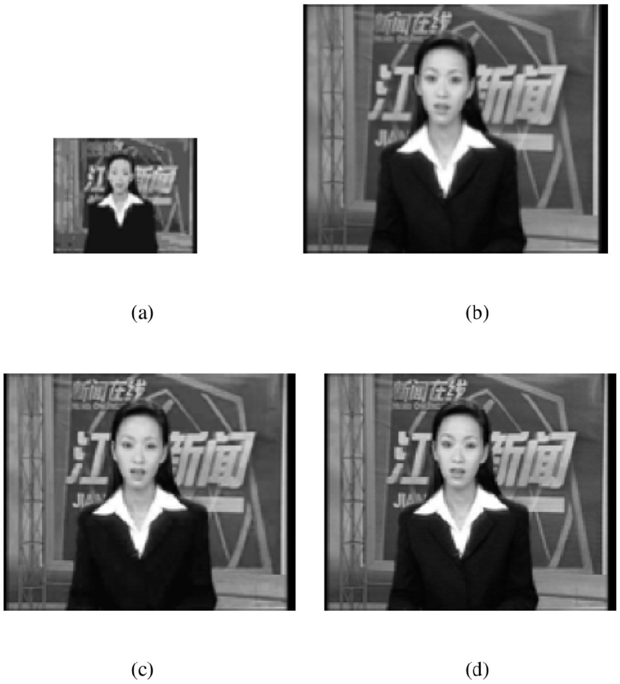


Figure 2: Results from Actual Sequence (a) LR Image (b) Up-Sampled Image(c) SR with all MVs (d) SR with Accurate MVs

CONCLUSIONS

In this paper, we utilize the Bayesian framework to incorporate information from the bit-stream as well as to model synthetic coding artifacts, and it relies on a gradient descent optimization for realization. We model the process of video compression and exploit the quantization step size and motion vector information available in the bit-stream. The method also uses the source statistics and additional reconstruction constraints, such as those that might aid in blocking artifact reduction and edge enhancement. A synthetic sequence and an actual image sequence are simulated. Experimental results demonstrate that the proposed algorithm has an improvement in terms of both objective and subjective quality. However, there are still some open issues. The nature of the distributions of the DCT coefficients is of high importance for determining DCT domain quantization noise. Rather than assuming uniform DCT-domain quantization noise, prior knowledge of the DCT coefficients can be incorporated in the model, which can yield better results. Again, more thorough analysis needs to be done.

ACKNOWLEDGMENTS

This work was supported by Nature Science Foundation of Jiangsu Province, P. R. China. (No.BK2004151).

REFERENCES

[1] Tsai R and Huang T. Multi-frame Image Restoration and Registration. *Advances in Computer Vision and Image Processing*, 1984, (1): 317-339.

[2] Stark H and Oskoui P. High-resolution Image Recovery from Image-plane Arrays, using Convex Projections. *Journal of the Opt. Soc. of America (A)*, 1989, 6 (11): 1715-1726.

[3] Patti AJ, Sezan MI and Tekalp AM. Super-resolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time. *IEEE Trans. Image Processing*, 1997, 6(8): 1064–1076.

[4] Schultz RR and Stevenson RL. Extraction of High Resolution Frames from Video Sequences. *IEEE Trans. Image Processing*, 1996, 5 (6): 996-1001.

[5] Hardie RC, Barnard KJ, Armstrong EE. Joint MAP Registration and High-resolution Image Estimation using A Sequence of under-sampled Images. *IEEE Trans. Image Processing*, 1997, 38(6): 1621-1633.

[6] Segall CA, Molina R, Katsaggelos AK and Mateos J. Reconstruction of High-resolution Image Frames from a Sequence of Low-resolution and Compressed Observations. *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Orlando, Florida, 2002, vol. 2: 1701–1704.

[7] Gunturk BK, Altunbasak Y and Mersereau RM. Multi-frame Resolution-enhancement Methods for Compressed Video. *IEEE Signal Processing Letters*, 2002, 9(6): 170–174.

[8] Patti AJ, Altunbasak Y. Super-resolution Image Estimation for Transform Coded Video with Application to MPEG. in Proc. *IEEE Int. Conf. Image Processing*, Kobe, Japan, 1999.

[9] Chen D and Schultz RR. Extraction of High-resolution Video Stills from MPEG Image Sequences. *IEEE Int. Conf. Image Processing*, Chicago, IL, 1998, vol. 2: 465–469.

[10] Gunturk BK, Altunbasak Y and Mersereau RM. Super-Resolution Reconstruction of Compressed Video using Transform-domain Statistics. *IEEE Trans. Image Processing*, 2004, 13(1): 33-43.

[11] Robertson MA and Stevenson RL. DCT Quantization Noise in Compressed Images. *IEEE Trans. Circuits and Systems for Video Technology*. 2005, 15(1): 27-38.

[12] Segall CA, Katsaggelos AK, Molina R and Mateos J. Bayesian Resolution Enhancement of Compressed Video. *IEEE Trans. Image Processing*, 2004, 13(7): 898–911.

[13] Robertson MA and Stevenson RL. Temporal Resolution Enhancement in Compressed Video Sequences. *EURASIP Journal on Applied Signal Processing*, 2001, 4: 230-238.

[14] Jain. AK. *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.

[15] Schultz RR, Stevenson RL. A Bayesian Approach to Image Expansion for Improved Definition. *IEEE Trans. Image Processing*, 1994, 3(2): 233-242.

[16] O'Rourke TP and Stevenson RL. Improved Image Decompression for Reduced Transform Coding artifacts. *IEEE Trans. on Circuits and System for Video Technology*. 1995, 5(6): 490–499.

[17] Jacoby SLS, Kowalik JS and Pizzo JT. *Iterative Methods for Nonlinear Optimization Problems*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1972.

[18] Schultz RR, Li Meng and Stevenson RL. Subpixel Motion Estimation for Super-resolution Image Sequence Enhancement. *Journal of Visual Communications and Image Representation*. 1998, 9(1) 1:38–50.