Bayesian Super Resolution Estimation for EBCOT Compressed Video

Zoran Ivanovski[1], Ljupcho Panovski[1], Cvetko Mitrovski[2]

Abstract - **In this paper, a new approach to video enhancement based on super-resolution is presented. The proposed method is used to enhance video sequences compressed with the MotionJPEG2000 standard. It is based on a spatial-domain Bayesian maximum a posteriori probability estimator. Simulations and real video experiments show improvement in the peak signal-to-noise ratio as well as improvement in visual quality.**

Keywords - **super-resolution, enhancement, EBCOT, compressed video, MAP estimator.**

I. INTRODUCTION

In the past twenty years, the area of super-resolution (SR) has received considerable attention. Since its beginning, dating back to the work of Huang and Tsai in 1984, numerous algorithms have been proposed with different levels of success. Some of the algorithms approach the SR problem in the frequency domain [1], [2]. Some exploit mapping of low-resolution images onto a HR image plane followed by interpolation and restoration [3]. Another approach to SR, which incorporates prior knowledge about the solution into the reconstruction process, is based on Projection onto Convex Sets (POCS). This is done by restricting the solution to be a member of a closed convex set of vectors that satisfy a particular property [4]. Reconstruction constraints have also been used in a statistical estimation framework. The constraints can be easily embedded in a Bayesian estimator incorporating a prior knowledge about the HR image. Both the Maximum Likelihood (ML) and Maximum a Posteriori Probability (MAP) estimators have been used. Tom and Katsaggelos [5] proposed an ML estimator that estimates the subpixel shifts, the noise variance of each image, and the HR image simultaneously. A MAP estimator that uses an edge preserving Huber-Markov random field for an image prior was proposed by Shultz and Stevenson [6]. Hardie et al. used a MAP estimator for simultaneous estimation of the image registration parameters and the HR image.

Recently, as the use of compressed video grew higher, the problem of enhancing compressed video came into focus.

[1] Zoran Ivanovski and Ljupcho Panovski are with the Faculty of Electrical Engineering, Karpos II bb, 1000 Skopje, Macedonia, E-mails: mars@cerera.etf.ukim.edu.mk; panovski@ukim.edu.mk

[2] Cvetko Mitrovski is with the Faculty of Technical Sciences, Ivo Ribar Lola bb, 7000 Bitola, Macedonia, E-mail: cvetko.mitrovski@ uklo.edu.mk

In the last few years several techniques for enhancing compressed video have been proposed. Early techniques usually estimate subpixel motion vectors (using the information from the transmitted ones) before applying the SR process [8], [9], [10]. Later on, techniques that simultaneously estimate motion field and high resolution frame were proposed [11]. Patti and Altunbasak demonstrated the importance of the proper handling of the quantization information and proposed a solution that explicitly exploits the compression process by incorporating the quantization information into a POCS-based algorithm that operates in the compression domain [12], [13]. In order to establish a stochastic framework that can utilize the quantization information, Gunturk et al. in [14] model the additive noise and than transform it to the compressed domain. A Gaussian model for conditional probability is used. In [15] a method to experimentally estimate the conditional probability is proposed. Bayesian reconstruction methods have also been proposed for the minimization of the artifacts introduced by the compression process [10], [16].

Although the framework developed in most of the aforementioned works is general and can be used with all video coding standards where the transform utilized is linear, most of them deal with DCT-based standards. On the other side, new compression techniques are mostly wavelet-based. Usually, the wavelet transform is applied to the image, without dividing it into blocks (tiles), although the possibility of tiling is still given in JPEG2000. The avoidance of tiles eliminates the blocking effect as one of the artifacts of the compression process, and consequently the necessity to optimize the SR technique for its elimination. High compression ratio in JPEG2000 is achieved by optimizing the compression through minimization of the MSE for the desired bit-rate (Tier 2 coding). The minimization (bit-stream generation) phase makes the modeling of the conditional probability density function very difficult. In this paper we demonstrate the possibility of the use of a Gaussian model for the conditional PDF in combination with a prior PDF model based on Markov Random Fields.

II. IMAGING MODEL

The observation model for the low-resolution images in the video sequence assumes that low-resolution images are generated from the high-resolution (HR), ideal, and undegraded images, $\mathbf{z}_k$. These HR images are representing scene values sampled at or above the Nyquist rate. The pixels in the low-resolution image are defined as a weighted sum of appropriate HR pixels with additive noise, with weighting being used to model the blurring that is due to the finite detector size and point spread function (PSF) of the detector and the optics.

Consider low-resolution frames of size $N_1 \times N_2$ pixels. The values of the pixels in the k-th low-resolution frame of the sequence can be expressed as

$$y_{k,m} = \sum_{r=1}^{q^2 N_1 N_2} w_{k,m,r} z_{k,r} + \eta_{k,m} \quad m = 1,\dots, N_1 N_2 \tag{1}$$

where $\eta_{k,m}$ is additive noise representing the error in estimating $w_{k,m}$. The noise, $\eta_{k,m}$, is assumed to be an independent and identically distributed (i.i.d.) Gaussian random variable. In Eq. (1), $z_{k,r}$ is the r-th element of the lexicographic representation of the k-th undegraded HR image. Different model weights for every pixel in the low-resolution frames correspond to different amounts of motion at each pixel during the acquisition of each frame. The subpixel motion of every low-resolution pixel relative to the HR grid is essential for the estimation of the HR image. We can express Eq. (1) in matrix notation as

$$\mathbf{y}_k = \mathbf{w}_k \mathbf{z}_k + \mathbf{n}_k \tag{2}$$

The relation between the HR images in a video sequence can be expressed as:

$$\mathbf{z}_k = \mathbf{C}(\mathbf{d}_{k,k-i})\mathbf{z}_{k-1} \tag{3}$$

where $\mathbf{C}(\mathbf{d}_{k,k-i})$ represents the motion compensation matrix from frame k to frame $k-i$ and $\mathbf{d}_{k,k-i}$ are motion vectors. Hence, the relation between low the resolution images and the different HR images is:

$$\mathbf{y}_{k-i} = \mathbf{AHC}(\mathbf{d}_{k,k-i})\mathbf{z}_k + \mathbf{n}_{k-i} \tag{4}$$

where $\mathbf{A}$ is the downsampling matrix, and $\mathbf{H}$ is the filtering matrix.

The low resolution images are than compressed before transmission and decompressed at the receiver, yielding

$$\mathbf{g}_{k-i} = \mathbf{T}_{DWT}^{-1} Q^* \left[Q\left[\mathbf{T}_{DWT} \left(\mathbf{AHC}(\mathbf{d}_{k,k-i})\mathbf{z}_k + \mathbf{n}_{k-i} \right) \right] \right] \tag{5}$$

where $\mathbf{T}_{DWT}$ and Q are the transform matrix and the quantization operator, respectively.

II. BAYESIAN MAP ESTIMATOR

In order to estimate the high resolution image $\mathbf{z}_k$ at time k, and the matrix of motion vectors $\mathbf{d}$ that describes the motion between $\mathbf{z}_k$ and M_F+M_B neighboring frames, we form a Bayesian MAP estimator given the low-resolution decompressed frames $\mathbf{g}$ and the appropriate prior. The estimate can be computed by maximizing the a posteriori probability $\Pr(\mathbf{z}_k, \mathbf{d} | \mathbf{g})$, which, according to Bayes theorem, gives:

$$\hat{\mathbf{z}}_k, \hat{\mathbf{d}} = \arg\max\{\Pr(\mathbf{g} \mid \mathbf{z}_k, \mathbf{d})\Pr(\mathbf{z}_k)\Pr(\mathbf{d})\} \tag{6}$$

or by maximizing the log-likelihood function

$$\begin{aligned}\hat{\mathbf{z}}_k, \hat{\mathbf{d}} = \arg\ \max\{&\log\ [\Pr(\mathbf{g} \mid \mathbf{z}_k, \mathbf{d})] \\ &+ \log[\Pr(\mathbf{z}_k)] + \log[\Pr(\mathbf{d})]\}\end{aligned} \tag{7}$$

The probability $\Pr(\mathbf{d})$ is ignored since it is not a function of $\mathbf{z}_k$ or $\mathbf{d}$. It is assumed that $\mathbf{z}_k$ and $\mathbf{d}$ are statistically independent.

The first term in Eq. (7) is a conditional probability that models the errors in estimating the motion vectors, errors introduced during the conversion of the high resolution image to the low resolution observation and the noise introduced during the compression. Assuming that the error between frames is zero-mean i.i.d. Gaussian, we can write:

$$\Pr(\mathbf{g} \mid \mathbf{z}_k, \mathbf{d}) \propto \exp\left\{ -\frac{1}{2}\beta \sum_{i=-MB}^{MF} \left\| AC(\mathbf{d}_{k,k-i})\mathbf{z}_k - \mathbf{g}_{k-i} \right\|^2 \right\} \tag{8}$$

where $\beta = \dfrac{1}{\sigma_n^2}$ and σ_n^2 is the noise variance.

The second term in Eq. (7) is the prior HR image model. We assumed a local conditional PDF based on Markov Random Fields which penalizes the difference between the pixel intensity and the averaged intensity value of its four neighboring pixels:

$$\Pr(\mathbf{z}_k) \propto \exp\left\{ -\frac{1}{\lambda} \left\| \mathbf{z}_k - \mathbf{a}\mathbf{z}_k \right\|^2 \right\} \tag{9}$$

where $\mathbf{a}$ represents the four neighbors averaging matrix and λ is a "tuning" parameter.

By substituting Eqs. (8) and (9) into Eq. (7) we obtain:

$$\hat{\mathbf{z}}_k, \hat{\mathbf{d}} = \arg\min\left\{ \beta \sum_{i=-MB}^{MF} \left\| AC(\mathbf{d}_{k,k-i})\mathbf{z}_k - \mathbf{g}_{k-i} \right\|^2 + \frac{1}{\lambda} \left\| \mathbf{z}_k - \mathbf{a}\mathbf{z}_k \right\|^2 \right\} \tag{10}$$

The minimization of Eq. (10) is accomplished through a cyclic coordinate-descent minimization procedure. At each iteration n, the motion parameters estimates are updated through a search procedure to minimize Eq. (10) with respect to $\mathbf{d}$, given that $\mathbf{z}_k = \hat{\mathbf{z}}_k^n$.

$$\hat{\mathbf{d}}^n = \arg\min\left\{ \beta \sum_{i=-MB}^{MF} \left\| AC(\mathbf{d}_{k,k-i})\mathbf{z}_k^n - \mathbf{g}_{k-i} \right\|^2 \right\} \tag{11}$$

Adaptive block-matching algorithm is used as a search procedure.

Once the estimate for the motion parameter is found, a steepest descent technique is employed to minimize Eq. (10) with respect to $\mathbf{z}$ and to estimate $\hat{\mathbf{z}}_k^{n+1}$.

IV. EXPERIMENTAL RESULTS

Two sets of experiments were conducted to test the performance of the algorithm. The first set of experiments was performed on a simulated sequence of images. A single HR image was used to produce a sequence of low-resolution images, with given subpixel motion. The generated sequence of low-resolution images was compressed and decompressed using the EBCOT compression algorithm (JPEG2000). In the second set of experiments, a real video sequence was used.

In the first set of experiments, the "cameraman" image was used to produce a sequence of three 144 × 176 low-resolution images, the second image being shifted -0.3 pixels in both directions, and the third being shifted 0.5 pixels in both directions. All three images were then compressed with the EBCOT algorithm to a compression ratio of 40:1 and decompressed. The decompressed images were used as the input for the SR estimation procedure. Despite the fact that only three images were used, a PSNR gain of 0.85 dB was achieved. The visual quality of the image also improved. Fig. 1(a) shows the original image, Fig. 1(b) shows the compressed image, and Fig. 1(c) shows the image obtained after the SR procedure was applied. The results shown were achieved with $\lambda = 1$ and with a fixed-size block matching algorithm in a motion estimation procedure with 0.1 pixel accuracy. The procedure was stopped after 2 iterations, since no further significant visual improvement was noticed.

Fig. 1(a) The original "cameraman" image

Fig. 1(b) The reconstruction of the EBCOT compressed "cameraman" image

Fig. 1(c) The SR enhanced reconstruction of the EBCOT compressed "cameraman" image

In the second set of experiments, the “coastguard” video sequence was used, consisting of 300 frames with resolution 144×176, recorded at 30 fps. Fig. 2 shows the PSNR gain for all frames following the applicatoin of the SR procedure with $\lambda = 1$ and motion estimation with 0.1 pixel accuracy. Again, the procedure was stopped after 2 iterations, since no further significant visual improvement was noticed. The region with large negative values in Fig. 2 demonstrates that the SR procedure could be destructive for the quality if inappropriate parameters are applied. Namely, this region corresponds to a particular part of the "coastguard" video sequence, which contains very large motion shifts, much larger than the range of the motion estimation search window. However, outside this region, the average PSNR gain is at the level of 0.5 dB.

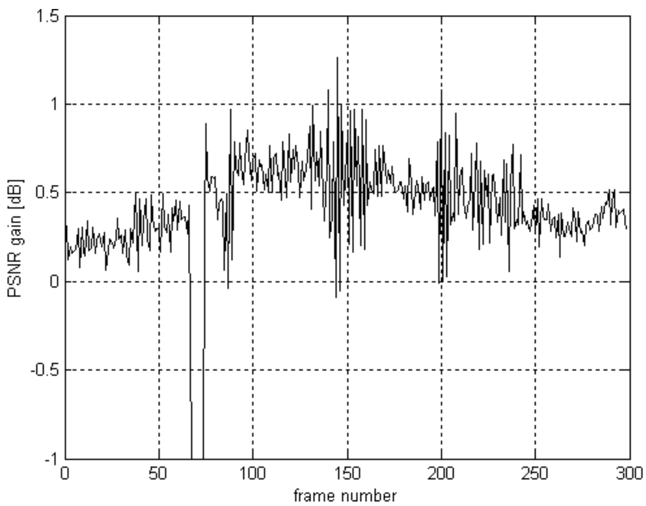

Fig 2. PSNR gain for the "coastguard" sequence

Figs. 3 (a), (b) and (c) show frame #145 of the original, of the compressed, and of the SR processed sequence, respectively.

Fig 3(a) Frame #145 of the original "coastguard" sequence

Fig. 3(b) Reconstruction of frame #145 of the EBCOT compressed "costguard" sequence

Fig. 3(c) SR enhanced reconstruction of frame #145 of the EBCOT compressed "coastguard" sequence

V. CONCLUSIONS

An algorithm for video enhancement of video sequences compressed with MotionJPEG2000 is proposed, which is based on a spatial-domain Bayesian maximum a posteriori probability estimator. Simulations and real video experiments show improvements in the peak signal-to-noise ratio of up to 1 dB both in the case of simulated video sequences and in real video. At the same time, along with the improvements in the PSNR, improvements in the visual quality are achieved as well.

REFERENCES

[1] R. Y. Tsai and T. S. Huang, "Multiframe Image Restoration and Registration", *Advances in Computer Vision and Image Processing*, vol. 1, pp. 317-339, JAI Press Inc., 1984.

[2] S. P. Kim, N. K. Bose and H. M. Valenzuela, "Recursive Reconstruction of High-Resolution Image from Noisy Undersampled Multiframes", IEEE Trans. Acust., Speech, Signal Processing, vol. 38, No. 6, pp. 1013-1027, 1990.

[3] N. Nguyen and P. Milanfar, "An Efficient Wavelet-Based Algorithm for Image Superresolution", Proc. Int. Conf. Image Processing, vol. 2, pp. 351-354, 2000.

[4] H. Stark and P. Oskoui, "High Resolution Image Recovery from Image-Plane Arrays, Using Convex Projection", J. Opt. Soc. Am. A, vol. 6, pp. 1715-1726, 1989.

[5] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a High-Resolution Image by Simultaneous Registration, Restoration, and Interpolation of Low-Resolution Images", Proc. 1995 IEEE Int. Conf. Image Processing, vol. 2, pp. 539-542, 1995.

[6] R. R. Shultz and R. L. Stevenson, "Extraction of High-Resolution Frames from Video Sequences", IEEE Transaction on Image Processing, vol. 5, no. 6, pp. 996-1011, 1996.

[7] 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 Transaction on Image Processing, vol. 6, no. 12, pp. 1621-1633, 1997.

[8] Y. Altunbasak and A. J. Patti, “A Maximum A Posteriori Estimator for Higher Resolution Video Reconstruction from MPEG Video”, Proc. of the IEEE ICIP, Vancouver, BC, Canada, Sept. 10-13, 2000.

[9] A. K. Katsaggelos and N. P. Galatsanos, eds. *Signal Recovery Techniques for Image and Video Compression and Transmission*, Kluwer Academic Publishers, 1998.

[10] J. Mateos, A. K. Katsaggelos and R. Molina, “Resolution Enhancement of Compressed Low Resolution Video”, Proc. of ICASSP 2000, vol. 4, pp. 1919-1922, 2000.

[11] J. Mateos, A. K. Katsaggelos and R. Molina, “Simultaneous Motion Estimation and Resolution Enhancement of Compressed Low Resolution Video”, Proc. of ICASSP 2000, 2000.

[12] Y. Altunbasak, A. J. Patti and R. M. Mersereau, “Super-Resolution Still and Video Reconstruction from MPEG-Coded Video”, IEEE Trans. Circuits Syst. VId. Technol., vol. 12, pp. 217-226, Apr. 2002.

[13] Y. Altunbasak and A. J. Patti, “Artifacts Reduction for Set Theoretic Super Resolution Image Reconstruction with Edge Adaptive Constraints and High-Order Interpolants”, IEEE Trans. Image Processing, vol. 10, pp. 179-186, Jan. 2001.

[14] B. K. Gunturk, Y. Altunbasak and R. M. Mersereau, “Bayesian Resolution Enhancement Framework for Transform-Coded Video”, Proc. IEEE ICIP, vol. 2, pp. 41-44, 2001.

[15] B. K. Gunturk, Y. Altunbasak and R. M. Mersereau, “Multiframe Resolution-Enhancement Methods for Compressed Video”, IEEE Signal Processing Letters, vol. 9, no. 6, pp. 170-174, June 2002.

[16] C. A. Segall, R. Molina, A. K. Katsaggelos and J. Mateos, “Bayesian High-Resolution Reconstruction of Low-Resolution Compressed Video”, Proc. ICIP, vol. 2, pp. 25-28, 2001.