SIMULTANEOUS MOTION ESTIMATION AND RESOLUTION ENHANCEMENT OF COMPRESSED LOW RESOLUTION VIDEO

Javier Mateos[(a)], *Aggelos K. Katsaggelos*[(b)] *and Rafael Molina*[(a)]

(a) Depto. de Ciencias de la Computación e I. A., Universidad de Granada, 18071 Granada, Spain.
{jmd, rms}@decsai.ugr.es
(b) Dept. of Electrical and Computer Eng., Northwestern University. Evanston, IL.
aggk@ece.nwu.edu

ABSTRACT

In this work we propose an iterative algorithm for simultaneously estimating the motion field and high resolution frames from a compressed low resolution video sequence. The algorithm exploits the existing correlation between high and low resolution frames and information provided by the encoder, such as coding modes and motion vectors (when available), to obtain a higher resolution frame. The performance of the algorithm is demonstrated experimentally.

1. INTRODUCTION

High-resolution images are useful and often critical for many applications. Remote sensing applications, medical imaging, surveillance or frame freeze in video are some of the applications where high resolution images are crucial. An approach to obtain high resolution images is to increase the number of sensor in the camera that captures the images. Although this approach is feasible for some applications, obtaining a dense detector array may be very costly or simply unavailable. An alternative approach is to estimate a high resolution image from a sequence of low resolution aliased images. This is possible if there exists subpixel motion between the acquired frames.

In many applications the available low resolution video has been compressed. This is for example the case with most digital video cameras, in which the acquired data are compressed using one of the video compression standards, in order to reduce the storage requirements. This compression can introduce artifacts in the low resolution video sequence, such as blocking and mosquito artifacts, that should also be removed by the resolution enhancement algorithm. Information provided by the encoder, such as, motion vectors, quantization information, and macroblock type, could be used by the resolution enhancement algorithm.

Although a number of algorithms for the resolution enhancement of video have appeared in the literature (see for example [1, 3, 8, 9, 10, 12] and references therein), not much work has been reported on the problem when the low resolution video sequence has been compressed. In [7] the problem is approached by incorporating the quantization information in the high resolution image estimation process. In [5] the motion vectors from the encoder are also used but the estimation of the motion field and the high resolution frame is performed independently. In this paper we propose an iterative algorithm for simultaneously estimating the high resolution frames from a compressed low resolution video sequence and the motion relating the high and low resolution frames using the motion vectors provided by the encoder when available.

The paper is divided as follows. In section 2 the problem is formulated within the Bayesian paradigm and the needed degradation and prior models are presented. Section 3 discusses the algorithm for the simultaneous resolution enhancement of the compressed low resolution frames and estimation of the motion vectors. In section 4 results with the proposed algorithm are presented and, finally, section 5 concludes the paper.

2. NOTATION AND MODEL

2.1. Notation

Let $\mathbf{f}_k$ be the k-th frame of a high resolution uncompressed video sequence, each frame of size $PM \times PN$ pixels. The relation between each pair of frames in this sequence, $\mathbf{f}_k$ and $\mathbf{f}_{k-i}$, can be expressed pixelwise as

$$\mathbf{f}_{k-i}(r) = \mathbf{f}_k(r + \mathbf{d}_{k,k-i}(r)), \tag{1}$$

where $\mathbf{d}_{k,k-i}(r)$ is the motion vector for the pixel location r from frame k to frame $k-i$. Let us represent by $\mathbf{d}_{k,k-i}$ the vector containing the motion information relating the high resolution frame k to the frame $k-i$, $\mathbf{d}_{k,k-i} = (\mathbf{d}_{k,k-i}(1), \mathbf{d}_{k,k-i}(2), \ldots, \mathbf{d}_{k,k-i}(PM \times PN))^t$. The relation in Eq. (1), represented by the horizontal lines in figure 1, can be expressed in matrix notation as

$$\mathbf{f}_{k-i} = C(\mathbf{d}_{k,k-i})\mathbf{f}_k, \tag{2}$$

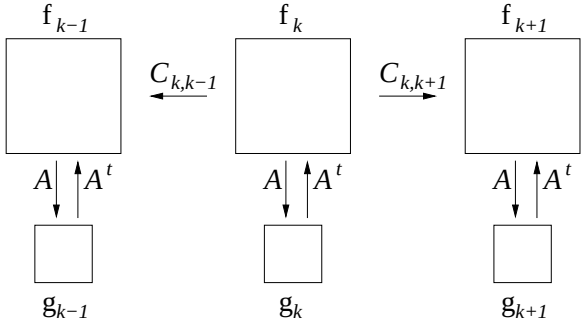


Fig. 1. Relationship between high and low resolution frames.

where the matrix $C(\mathbf{d}_{k,k-i})$ represents the motion compensation operator from frame k to frame $k-i$ obtained through the motion vectors in $\mathbf{d}_{k,k-i}$.

In order to obtain a low resolution compressed video sequence, the high resolution sequence is downsampled by a factor of P in each direction and then compressed using any of the video compression standards, such as MPEG4 [2], obtaining a sequence of compressed low resolution frames, $\mathbf{g}_k$, of size $M \times N$. The observed low resolution sequence is related to the original high resolution sequence by

$$\mathbf{g}_k(r) = (A\mathbf{f}_k)(r) + \mathbf{n}_k(r), \tag{3}$$

where $\mathbf{n}_k(r)$ represents the noise which is assumed to be additive Gaussian with zero mean and variance σ_n^2, and A is a $(M \times N) \times (PM \times PN)$ matrix denoting the integration and subsampling operations. This relation is represented by the vertical arrows in figure 1. Notice that while A is a downsampling by averaging operator, A^t is an upsampling by zero-order hold operator.

From Eqs. (2) and (3) it is clear that each low resolution frame is related to the high resolution frame $\mathbf{f}_k$ by

$$\mathbf{g}_{k-i} = AC(\mathbf{d}_{k,k-i})\mathbf{f}_k + \mathbf{n}_{k-i}, \tag{4}$$

that is, we can see the low resolution video sequence as different subsampled realizations of a single high resolution frame via different motion compensation operations.

2.2. Bayesian formulation

Our objective is, given a sequence of $M1+M2+1$ observed compressed low resolution frames $\mathbf{g}_{k-i}$ to find an estimate of the frame $\mathbf{f}_k$ and the motion vectors $\mathbf{d}_{k,k-i}$.

Let $\mathbf{d} = (\mathbf{d}_{k,k-M1}^t, \ldots, \mathbf{d}_{k,k+M2}^t)^t$ and $\mathbf{g} = (\mathbf{g}_{k-M1}^t, \ldots, \mathbf{g}_{k+M2}^t)^t$, the Bayesian paradigm dictates that inference about the true $\mathbf{f}_k$ and $\mathbf{d}$ should be based on $p(\mathbf{f}_k, \mathbf{d}|\mathbf{g})$, given by

$$\begin{aligned} p(\mathbf{f}_k, \mathbf{d}|\mathbf{g}) &= \frac{p(\mathbf{g}|\mathbf{f}_k, \mathbf{d})p(\mathbf{f}_k, \mathbf{d})}{p(\mathbf{g})} \\ &\propto p(\mathbf{g}|\mathbf{f}_k, \mathbf{d})p(\mathbf{f}_k)p(\mathbf{d}), \end{aligned} \tag{5}$$

where we are assuming that $\mathbf{f}_k$ and $\mathbf{d}$ are statistically independent.

Maximization of Eq. (5) with respect to $\mathbf{f}_k$ and $\mathbf{d}$ yields

$$\hat{\mathbf{f}}_k, \hat{\mathbf{d}} = \arg\max_{\mathbf{f}_k, \mathbf{d}} \{p(\mathbf{g}|\mathbf{f}_k, \mathbf{d})p(\mathbf{f}_k)p(\mathbf{d})\}, \tag{6}$$

the maximum *a posteriori* (MAP) estimator.

Clearly,in utilizing Eq. (6) we must specify the conditional density $p(\mathbf{g}|\mathbf{f}_k, \mathbf{d})$, that is, the degradation model for the problem at hand, and the prior densities of the image, $p(\mathbf{f}_k)$, and the motion, $p(\mathbf{d})$, in which we incorporate the information we have about the form of the high resolution image and the motion field, respectively.

2.3. Noise, image, and motion models

Given the observation model in Eq. (4), the conditional density in Eq. (6) is given by

$$p(\mathbf{g}|\mathbf{f}_k, \mathbf{d}) \propto \exp\left\{-\frac{1}{2}\beta \sum_{i=-M2}^{M1} \parallel AC(\mathbf{d}_{k,k-i})\mathbf{f}_k - \mathbf{g}_{k-i} \parallel^2\right\},$$

where $\beta = 1/\sigma_n^2$.

Our prior knowledge about the object luminosity distribution of the original image includes the fact that it had no blocking artifacts and it was smooth within the blocks. This knowledge results in a model of the *a priori* distribution of $\mathbf{f}_k$ given by

$$p(\mathbf{f}_k) \propto \exp\left\{-\frac{1}{2}\left(\lambda_1 \parallel Q_1\mathbf{f}_k \parallel^2 + \lambda_2 \parallel Q_2\mathbf{f}_k \parallel^2\right)\right\},$$

where Q_1 and Q_2 are high pass operators that capture the within-block and between-block smoothness of the estimated frame $\mathbf{f}_k$, respectively, and λ_1 and λ_2 are the regularization parameters that control the within-block and between-block smoothness, respectively [11].

Taking into account that the encoder provides us with information about the motion vectors used during the coding process, we can include this information in the prior distribution of $\mathbf{d}$ by

$$p(\mathbf{d}) \propto \exp\left\{-\frac{1}{2}\gamma \parallel \mathbf{d} - \mathbf{d}^{enc} \parallel^2\right\},$$

where $\mathbf{d}^{enc}$ is the upsampled version of the motion vectors provided by the encoder and γ, the inverse of their variance,

determines the confidence we have on the data provided by the encoder. For $\gamma = \infty$ the use of the vectors provided by the encoder is enforced while $\gamma = 0$ implies that we either do not have confidence on the received vectors or they are not available. Note that in this distribution we can incorporate other constraints, such as spatio-temporal smoothness of the motion vector field.

3. IMAGE AND MOTION FIELD ESTIMATION

By substituting the models described in section 2.3 into Equation (6) and taking the negative of the logarithm we obtain

$$\hat{\mathbf{f}}_k, \hat{\mathbf{d}} = \arg\min_{\mathbf{f}_k, \mathbf{d}} L(\mathbf{f}_k, \mathbf{d}) \tag{7}$$

where

$$\begin{aligned} L(\mathbf{f}_k, \mathbf{d}) = \beta \sum_{i=-M2}^{M1} \parallel AC(\mathbf{d}_{k,k-i})\mathbf{f}_k - \mathbf{g}_{k-i} \parallel^2 \\ +\lambda_1 \parallel Q_1\mathbf{f}_k \parallel^2 +\lambda_2 \parallel Q_2\mathbf{f}_k \parallel^2 \\ +\gamma \parallel \mathbf{d} - \mathbf{d}^{enc} \parallel^2 . \end{aligned} \tag{8}$$

The optimization of Eq. (7) is carried out by a cyclic coordinate-descent optimization procedure [4]. With it in minimizing the global cost function, the cost function is minimized separately with respect to $\mathbf{d}$ and with respect to $\mathbf{f}_k$ in a cyclic fashion, that is, at each iteration of the algorithm, $\mathbf{d}$ is obtained for $\mathbf{f}_k$ fixed and then $\mathbf{f}_k$ is obtained for $\mathbf{d}$ fixed (the result at a previous iteration).

So, for fixed $\mathbf{f}_k$, the motion estimate is computed as

$$\begin{aligned} \hat{\mathbf{d}} \quad = \quad \arg\min_{\mathbf{d}} \Bigg\{ \beta \sum_{i=-M2}^{M1} \parallel AC(\mathbf{d}_{k,k-i})\mathbf{f}_k - \mathbf{g}_{k-i} \parallel^2 \\ +\gamma \parallel \mathbf{d} - \mathbf{d}^{enc} \parallel^2 \Big\} , \end{aligned} \tag{9}$$

and, for fixed $\mathbf{d}$, the image $\hat{\mathbf{f}}_k$ is computed using a gradient descent algorithm as

$$\begin{aligned} \mathbf{f}_k^{l+1} = \mathbf{f}_k^l + \epsilon \Bigg[\lambda_1 Q_1^t Q_1 \mathbf{f}_k^l + \lambda_2 Q_2^t Q_2 \mathbf{f}_k^l \\ +\beta \sum_{i=-M2}^{M1} C^t(\mathbf{d}_{k,k-i}) A^t \left(AC(\mathbf{d}_{k,k-i})\mathbf{f}_k^l - \mathbf{g}_{k-i} \right) \Bigg], \end{aligned} \tag{10}$$

where $\mathbf{f}_k^l$ and $\mathbf{f}_k^{l+1}$ are the enhanced frames in the lth and $(l+1)$st iterations, respectively, and ϵ is the relaxation parameter that controls the convergence and the rate of convergence of the algorithm.

Any motion vector search method can be applied to minimize Eq. (9) including efficient techniques for traditional block matching. Note that in Eq. (9) the high-resolution frame is compared, after motion compensation and subsampling, to a low resolution frame instead of a (bilinearly interpolated) high resolution frame.

4. EXPERIMENTAL RESULTS

In order to test the proposed algorithm, the color *Mobile* sequence was used. Each frame, of size 720×576 pixels, was subsampled to obtain the 180×144 pixels low resolution frame, that is, the downsampling factor P was equal to four. The first 40 frames of the sequence were compressed at 128Kbps using the baseline mode MPEG4 (TM5) video coding standard [2].

For this experiments $M1 = M2 = 1$ in Eq. (8) were used to obtain each high resolution frame. The distribution parameters were set to $\beta = 100$, $\lambda_1 = 3$, $\lambda_2 = 5$ and $\gamma = 233$. Methods for the automatic selection of these parameters, based on [6] are currently under investigation To obtain an initial estimate of the high resolution frames, bilinear interpolation was used on the low resolution frames, without using motion information.

Figure 2 shows the Y band of the zero-order hold frame 32, the bilinearly interpolated and the high resolution image obtained with the proposed algorithm. The PSNR of the Y-band for each image is 20.44dB, 21.98dB and 25.64dB, respectively. By comparing these figures a good improvement is observed, both in visual quality and $PSNR$.

5. CONCLUSIONS AND EXTENSIONS

A simultaneous motion estimation and high resolution video reconstruction algorithm from compressed low resolution sequences is presented. The algorithm in its current form uses only part of the information provided by the encoder, that is, the location of the block boundaries and the motion vectors. It is also possible, however, to incorporate additional information, such as a temporal smoothness constraint in the image prior model in order to take advantage of the correlation between the high resolution frames is under study. This temporal smoothness constraint will allow to use the information on the high resolution frames reconstructed by the algorithm at the previous iteration step. In addition, more complex models that take into account both spatial and temporal constraints can be incorporated into the motion prior distribution. The proposed method allows for great flexibility in incorporating knowledge from the encoder in a simple way.

The incorporation of the information about the macroblock type in the image and motion prior models, in order to make them adaptive, and the preservation of the fidelity to the quantization intervals is currently under study.

6. REFERENCES

[1] R. C. Hardie, K. J. Barnard, and E. E. Amstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE*

Trans. on Image Processing, vol. 6, no. 12, pp. 1621–1633, 1997.

[2] International Organization for Standarisation, *ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio*, 1999.

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

[4] D. G. Luenberger, *Linear and Nonlinear Programming*, Reading, MA:Addison-Wesley, 1984.

[5] J. Mateos, A. K. Katsaggelos, and R. Molina, "Resolution Enhancement of Compressed Low Resolution Video" in *Proceedings of the ICASSP 2000*, 2000.

[6] J. Mateos, A. K. Katsaggelos, and R. Molina, "A Bayesian Approach for the Estimation and Transmission of Regularization Parameters for Reduced Blocking Artifacts", *IEEE Trans. on Image Processing*, vol. 9, n. 7, July 2000.

[7] A. J. Patti, and Y. Altunbasak, "Super-Resolution Image Estimation for Transform Coded Video with Application to MPEG", in *Proceedings of the ICIP99*, n. 27A03.7, 1999.

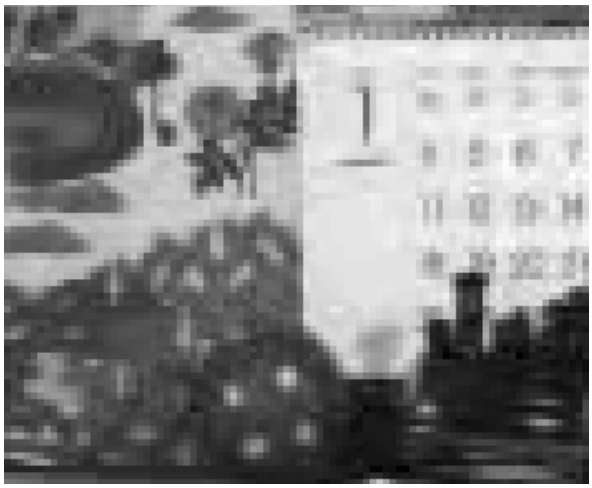
[8] R. R. Schultz, and R. L. Stevenson, "Extraction of High Resolution Frames from Video Sequences", *IEEE Trans. on Image Processing*, vol 5, n. 6, pp. 996–1011, 1996.

[9] B. C. Tom, *Reconstruction of a High Resolution Image from Multiple Degraded Mis-Registered Low Resolution Images*", PhD. thesis, Dept. of ECE, Northwestern University, Dec. 1995.

[10] B. B. Tom, and A. K. Katsaggelos, "Resolution Enhancement of Monochrome and Color Video Using Motion Compensation", to appear *IEEE Trans. on Image Processing*, 2000.

[11] C.-J. Tsai, P. Karunaratne, N. P. Galatsanos, and A. K. Katsaggelos, "A Compressed Video Enhancement Algorithm", in *Proceedings of the ICIP99*, n. 27AP5.1, 1999.

[12] T. R. Tuinstra, and R. C. Hardie, "High-Resolution Image Reconstruction from Digital Video by Exploitation of Nonglobal Motion", *Optical Eng.* vol. 38, n. 5, pp. 806–814, 1999.



(a)

(b)

(c)

Fig. 2. A section of the estimated frame 32 (a) by zero-order hold, PSNR = 20.44dB. (b) By bilinear interpolation, PSNR = 21.98dB. (c) With the proposed algorithm, PSNR = 25.64dB.