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 the estimation of high resolution frames from a low resolution compressed video sequence. The algorithm exploits the existing correlation between the high and low resolution frames and the information provided by the encoder to obtain a high resolution frame. The performance of the algorithm is demostrated experimentally.

1. INTRODUCTION

High-resolution images are useful and often important tools 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. Another disadvantage of this approach is that the signal to noise ratio decreases as the sensor size decreases. 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 in 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. Fortunately, the encoder provides some information with the video sequence such as the estimated motion vectors, quantization information, macroblock type, that 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, 5, 6] and the references therein), not much work has been reported on the problem when the low resolution video sequence has been compressed. In [4] the problem is approached by using the quantization information. In this work we propose an iterative algorithm for the estimation of high resolution frames from a low resolution compressed video sequence.

The paper is divided as follows. In section 2 the problem is formulated and notation is established. Section 3 discusses the proposed estimation model and algorithm for the resolution enhancement of the compressed low resolution frames. In section 4 some issues regarding the motion estimation process are addressed. In section 5 results with the proposed algorithm are presented and, finally, section 6 concludes the paper.

2. PROBLEM FORMULATION

Let $\mathbf{f}_k^t = \{(\mathbf{f}_k^Y)^t \ (\mathbf{f}_k^U)^t \ (\mathbf{f}_k^V)^t\}$ be the kth high resolution color frame, whose color components will be denoted by $\mathbf{f}_k^c$, $c \in \{Y, U, V\}$, and let $\mathbf{x}_k$ be the corresponding non-compressed low resolution frame. If the original size of the low resolution frame $\mathbf{x}_k^Y$ is $M_1 \times N_1$ and we assume a 4:1:1 subsampling, we note that for the luminance band, Y, a $M^Y \times N^Y$ image is obtained with $M^Y = M_1$ and $N^Y = N_1$ and, for the chrominance bands, the images are $M^c \times N^c$ where $M^c = M_1/2$ and $N^c = N_1/2$ with $c \in \{U, V\}$. Then, each high resolution frame is $PM^c \times PN^c$, $c \in \{Y, U, V\}$, with P an integer determining the degree of subsampling applied to obtain the low resolution image.

The relationship between the bands of the high and

*The work of Javier Mateos is partially supported by the Ministerio de Educación y Cultura.

†This paper was partially supported by the "Comisión Nacional de Ciencia y Tecnología" under contract TIC97-0989.

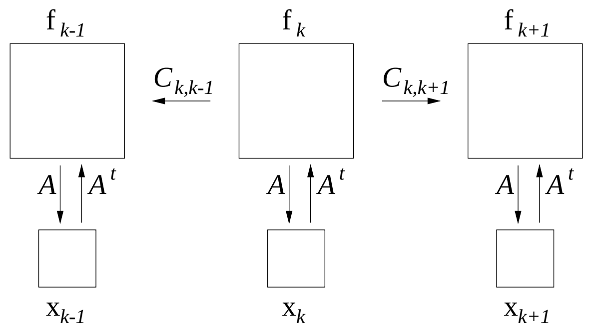


Figure 1: Relationship between high and low resolution frames.

low resolution frames is described by

$$\mathbf{x}_k^c = A^c \mathbf{f}_k^c,$$

where, for $c \in \{Y, U, V\}$, A^c, is an $(M^c \times N^c) \times (PM^c \times PN^c)$ matrix denoting the integration and subsampling operations. With the use of $A^t = \{(A^Y)^t \; (A^U)^t \; (A^V)^t\}$, the relation between the high and low resolution frames is

$$\mathbf{x}_k = A\mathbf{f}_k. \tag{1}$$

This relation is represented by vertical arrows in figure 1. Notice that while A is a downsampling by averaging operator, A^t is an upsampling and zero-order hold operator.

Clearly, there exists a relation between every two high resolution frames $\mathbf{f}_k$ and $\mathbf{f}_{k-i}$ that can be expressed as

$$\mathbf{f}_{k-i}^c = C_{k,k-i}^c \mathbf{f}_k^c,$$

or

$$\mathbf{f}_{k-i} = C_{k,k-i}\mathbf{f}_k, \tag{2}$$

where the matrix $C_{k,k-i}^c$, $c \in \{Y, U, V\}$, represents the motion compensation operator for band c of frame k to frame $k-i$ and $C_{k,k-i}$ represents the motion compensation operator from the color frame k to color frame $k-i$. This relationship, represented by the horizontal lines in figure 1, can be also expresed pixelwise by $\mathbf{f}_{k-i}^c(r) = \mathbf{f}_k^c(r + d_{k,k-i}^c(r))$, where $d_{k,k-i}^c(r)$ is the motion vector for band c from frame k to frame $k-i$, for pixel index r.

According to Eqs. (1) and (2), we observe that the low resolution sequence acquired by the camera is related to the high resolution sequence by

$$\mathbf{x}_{k-i} = A\mathbf{f}_{k-i} = AC_{k,k-i}\mathbf{f}_k. \tag{3}$$

The low resolution frames $\mathbf{x}_k$, $k = 1, \ldots, L$, obtained according to Eq. (3) are now compressed using any of the video compression standards, such as MPEG4 [2], obtaining the observed compressed sequence of low resolution frames $\mathbf{g}_k$.

3. RECOVERY ALGORITHM

Our objective is to find an estimate of frame $\mathbf{f}_k$, denoted by $\hat{\mathbf{f}}_k$, given $M1 + M2 + 1$ observed compressed frames $\mathbf{g}_i$, $i = k - M1, \ldots, k + M2$. The spatial resolution enhancement algorithm we propose uses a regularized reconstruction approach which takes into account that the observed frames were compressed and, hence, they may exhibit blocking and mosquito artifacts.

Based on the degradation model in Eq. (3), and in order to reduce the artifacts introduced by the compression process while enhancing the spatial resolution, the high resolution frame is selected as

$$\begin{aligned} \hat{\mathbf{f}}_k \;=\; & \arg\min_{\mathbf{f}} \Bigg\{ \sum_{i=-M2}^{M1} \| AC_{k,k-i}\mathbf{f}_k - \mathbf{g}_{k-i} \|^2 \\ & + \lambda_1 \| Q_1 \mathbf{f}_k \|^2 + \lambda_2 \| Q_2 \mathbf{f}_k \|^2 \\ & + \sum_{i=-M2}^{M1} \lambda_{k,k-i} \| C_{k,k-i}\mathbf{f}_k - \mathbf{f}_{k-i} \|^2 \Bigg\}, \end{aligned} \tag{4}$$

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, the term $\| C_{k,k-i}\mathbf{f}_k - \mathbf{f}_{k-i} \|^2$ enforces temporal smoothness between the high resolution frame being considered and one of its neighbors, and λ_1, λ_2 and $\lambda_{k,k-i}$ are the regularization parameters that control the within-block, between-block and temporal smoothness, respectively. These parameters can be adjusted according to the macroblocks characteristics to obtain optimal results [7]. Note that while the first term in Eq. (4) imposes fidelity to the received data from the encoder, the temporal smoothness constrain imposes continuity between adjacent high resolution frames. This high resolution frames may have been obtained by previous executions of the algorithm or by any type of interpolation from the corresponding low resolution frame.

The minimization in Eq. (4) can be carried out by an iterative gradient descendent algorithm described by

$$\begin{aligned} \mathbf{f}_k^{l+1} \;=\; & \mathbf{f}_k^l + \beta \Bigg[\sum_{i=-M2}^{M1} C_{k,k-i}^t A^t \left(AC_{k,k-i}\mathbf{f}_k^l - \mathbf{g}_{k-i} \right) \\ & + \lambda_1 Q_1^t Q_1 \mathbf{f}_k^l + \lambda_2 Q_2^t Q_2 \mathbf{f}_k^l \\ & + \sum_{i=-M2}^{M1} \lambda_{k,k-i} C_{k,k-i}^t (C_{k,k-i}\mathbf{f}_k^l - \mathbf{f}_{k-i}) \Bigg], \end{aligned} \tag{5}$$

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 β is the relaxation parameter that controls the convergence and the rate of convergence of the algorithm.

(a) (b) (c)

Figure 2: High resolution estimation of frame 9. (a) By bilinear interpolation; With the proposed algorithm estimating the motion vectors (b) from the original frames, and (c) from the initial frames.

4. ON MOTION ESTIMATION

It is known that one of the most crucial steps in obtaining high resolution frames from low resolution ones is the estimation of the motion vectors. In this section we propose a method to estimate the motion vectors, using the information provided by the encoder.

Note that the motion vectors provided by the encoder cannot be used directly for the problem at hand since: (i) they do not provide enough accuracy; (ii) they are referred to the low resolution frames and; (iii) not all needed motion fields are available, that is, we may have $d_{k-i,k}(r)$ for a given frame pair but not $d_{k,k-i}(r)$. Therefore, we have to estimate a new set of motion vectors for the high resolution frames. Since the encoder provides us with the motion vectors used for the compression of the low resolution frames, we can use this information to improve the estimate of the high resolution motion vectors. The motion vectors provided by the encoder can be incorporated into the estimation procedure by combining them with the motion vectors estimated from the high resolution image, according to

$$d_{k,k-i}(r) = \alpha_{k,k-i} d^{enc}_{k,k-i}(r) + (1-\alpha_{k,k-i}) d^{hr}_{k,k-i}(r), \quad (6)$$

where $d^{enc}_{k,k-i}(r)$ are the upsampled motion vectors provided by the encoder, $d^{hr}_{k,k-i}(r)$ are the motion vectors estimated from an estimate of the high resolution frames, and $0 \leq \alpha_{k,k-i} \leq 1$ is the parameter that controls the relative contribution of the two elements. Note that if $\alpha_{k,k-i} = 0$ no use of the motion information provided by the encoder is made, either because we have no confidence on $d^{enc}_{k,k-i}(r)$ or, simply, it is not available.

Although as verified experimentally this approach improves the estimated motion vectors by forcing them to be close to the ones provided by the encoder, in some situations it fails. This is the case when a motion vector provided by the encoder belongs to a macroblock containing parts or two differently moving objects. In this case it is desirable not to use the information provided by the encoder. To do so, the convex combination in Eq. (6) is used only if the difference between $d^{enc}_{k,k-i}(r)$ and $d^{hr}_{k,k-i}(r)$ is less than a small quantity, δ, that is,

$$d_{k,k-i} = \begin{cases} \alpha_{k,k-i} d^{enc}_{k,k-i} + (1-\alpha_{k,k-i}) d^{hr}_{k,k-i} & d \leq \delta \\ d^{hr}_{k,k-i} & d > \delta \end{cases}, \quad (7)$$

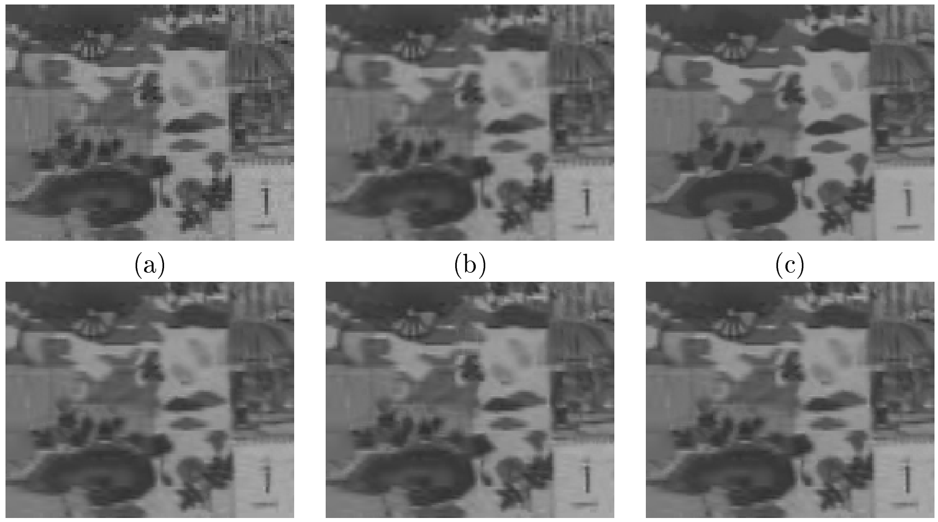
where $d = |d^{enc}_{k,k-i}(r) - d^{hr}_{k,k-i}(r)|$.

5. 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 a CIF format frame. These frames represent the original high resolution frames $\mathbf{f}^k$. They were further subsampled by two in each direction, according to Eq. (1), to obtain the QCIF low resolution frames. The first 40 frames of the sequence were compressed at 128kbp using the baseline mode MPEG4 video coding standard [2].

For this experiments $M1 = M2 = 1$ in Eq. (5) was used. The smoothness parameters were set to $\lambda_1 = 0.1$, $\lambda_2 = 0.25$ and $\lambda_{k,k-i} = 0.3$, $i \in [-1,1]$, and $\beta = 1.0$. To obtain the initial high resolution frames estimation, bilinear interpolation was used on the low resolution frames, without using motion information.

Figure 2a shows the Y band of the bilinearly interpolated frame 9 and the high resolution image obtained with the proposed algorithm using the motion vectors estimated from the original high resolution images (Fig. 2b) and using the vectors estimated from the initial bilinearly interpolated high resolution images (Fig. 2c). The PSNR of the Y-band for each image is 22.31dB, 27.85dB and 24.62dB, respectively. By comparing these figures a good improvement is ob-



(a) (b) (c)

(d) (e) (f)

Figure 3: A section of the estimated frame 9 (a) by zero-order hold. PSNR = 14.33dB (b) By bilinear interpolation, PSNR= 22.31dB. With the proposed algorithm computing the motion vectors (c) from the original images, PSNR= 27.85dB, (d) from the initial bilinearly interpolated images, PSNR= 24.62dB, (e) using Eq. (6), PSNR= 24.53dB, and (f) using Eq. (7), PSNR= 24.56dB.

served, both in visual quality and $PSNR$.

In order to test the dependency of the different approaches on the various motion estimates we ran experiments using $\alpha_{k,k-i} = 0.5$, $i \in [-1, 1]$, and $\delta = 2$ in Eq. (7), when data from the decoder was available, $\alpha_{k,k-i} = 0$ otherwise. Figure 3 shows the results on a small section of the Y band of frame 9. It is clear that the best results are obtained when the motion vectors are estimated from the original images. Also, it is clear that the motion estimation scheme proposed in Eq. (6) fails if the motion vectors from the encoder are not reliable (see central part of fig. 3e). On the other hand, using the scheme proposed in Eq. (7) this problem is solved and better visual quality is obtained. Results obtained by using only the initial images to estimate the motion vectors are slightly better in terms of PSNR than the obtained by the scheme proposed in Eq. (7); however, the latter one gives better visual quality than the former, as it can be observed in Figure 3.

6. CONCLUSIONS

In this paper, a new method for enhancing the resolution of a compressed color video sequence is presented. The algorithm takes into account the special characteristics of the compressed video sequence reducing the artifacts introduced by the compression process. Also we propose a method for estimating the motion vectors by combining the information provided by the encoder with ones the estimated from the initial high resolution frames.

Although the proposed method treats each band independently, the possibility of combining the between-band information is currently being investigated. Work is also in progress on the use of additional information provided from the coder.

REFERENCES

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

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

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

[4] 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.

[5] R. R. Schultz and R. L. Stevenson, "Extraction of high resolution frames from video sequences", IEEE Trans. on IP, vol 5, n. 6, pp. 996–1011, 1996.

[6] 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.

[7] 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.