Multi-channel Reconstruction of Video Sequences from Low-Resolution and Compressed Observations

Luis D. Alvarez[1], Rafael Molina[1], and Aggelos K. Katsaggelos[2]

[1] Departamento de Ciencias de la Computación e Inteligencia Artificial,
University of Granada, 18071 Granada, Spain
{ldac, rms}@decsai.ugr.es
http://decsai.ugr.es/{~ldac, ~rms}/

[2] Department of Electrical and Computer Engineering,
Northwestern University, Evanston, IL 60208, USA
{aggk}@ece.nwu.edu
http://www.ece.northwestern.edu/~aggk/

Abstract. A framework for recovering high-resolution video sequences from sub-sampled and compressed observations is presented. Compression schemes that describe a video sequence through a combination of motion vectors and transform coefficients, e.g. the MPEG and ITU family of standards, are the focus of this paper. A multichannel Bayesian approach is used to incorporate both the motion vectors and transform coefficients in it. Results show a discernable improvement in resolution in the whole sequence, as compared to standard interpolation methods.

1 Introduction

High-frequency information is often discarded during the acquisition and processing of an image. This data reduction begins at the image sensor, where the original scene is spatially sampled during acquisition, and continues through subsequent sampling, filtering or quantization procedures. Recovering the high-frequency information is possible though, as multiple low-resolution observations may provide additional information about the high-frequency data. This information is introduced through sub-pixel displacements in the sampling grid, which allows for the recovery of resolution.

Although work has been devoted to the problem of reconstruction of one high resolution image from a sequence of low resolution ones (see for instance [1-5] and [6] for a review), not much work has been reported on the problem of increasing the resolution of a whole image sequence simultaneously (see however [7–9]).

In this paper we present a new method to obtain a whole high resolution sequence from a set of low resolution observations. The method will use the relationship between the high resolution images in the sequences and also the process to obtain the low resolution compressed ones from their corresponding high resolution images.

The rest of this paper is organized as follows: In section 2, we formulate the problem within the Bayesian framework, define the acquisition system to be considered and the prior information we are going to use on the high resolution image

A. Sanfeliu and J. Ruiz-Shulcloper (Eds.): CIARP 2003, LNCS 2905, pp. 46–53, 2003.

sequence. In section 3, we introduce an iterative algorithm for estimating the high-resolution sequence. In section 4, we present results from the proposed procedure. Conclusions are presented in section 5.

2 System Model

When images from a single camera are captured at closely spaced time instances, then it is reasonable to assume that the content of the frames is similar. That is, we can say that

$$f_l(x,y) = f_k\left(x + d_{l,k}^x(x,y),\, y + d_{l,k}^y(x,y)\right) + n_{l,k}(x,y) \ , \tag{1}$$

where $f_l(x,y)$ and $f_k(x,y)$ are the gray level values at spatial location (x,y) in the high-resolution images at times l and k, respectively, $d_{l,k}^x(x,y)$ and $d_{l,k}^y(x,y)$ comprise the displacement that relates the pixel at time k to the pixel at time l, and $n_{l,k}(x,y)$ is an additive noise process that accounts for any image locations that are poorly described by the displacement model.

The expression in (1) can be rewritten in a matrix-vector form as

$$\mathbf{f}_l = \mathrm{C}(\mathbf{d}_{l,k})\mathbf{f}_k + \mathbf{n}_{l,k} \ , \tag{2}$$

where $\mathbf{f}_l$ and $\mathbf{f}_k$ are formed by lexicographically ordering each image into an one-dimensional vector, $C(\mathbf{d}_{l,k})$ is the two-dimensional matrix that describes the displacement across the entire frame, $\mathbf{d}_{l,k}$ is the column vector defined by lexicographically ordering the values $\left(d_{l,k}^x(x,y), d_{l,k}^y(x,y)\right)$ and $\mathbf{n}_{l,k}$ is the noise process. When the images are PMxPN arrays, then $\mathbf{f}_l$, $\mathbf{f}_k$, $\mathbf{d}_{l,k}$ and $\mathbf{n}_{l,k}$ are column vectors with length $PMPN$ and $C(\mathbf{d}_{l,k})$ has dimension $PMPN$x$PMPN$.

The conversion of a high-resolution frame to its low-resolution and compressed observation is expressed as

$$\mathbf{y}_k = T_{DCT}^{-1} Q\left[T_{DCT}\left(\mathbf{AHf}_k - \sum_{\forall i} C(\mathbf{v}_{k,i})\mathbf{y}_i\right)\right] + \sum_{\forall i} C(\mathbf{v}_{k,i})\mathbf{y}_i \ , \tag{3}$$

where $\mathbf{y}_k$ is a vector that contains the compressed low-resolution images with dimension $MNx1$, $\mathbf{f}_k$ is the high-resolution data, $\mathbf{v}_{k,i}$ is the motion vector transmitted by the encoder that signals the prediction of frame k from previously compressed frame i, $C(\mathbf{v}_{k,i})$ represents the prediction process with a matrix (for images said to be "intra-coded", the prediction from all frames is zero), $\mathbf{A}$ is an MNx$PMPN$ matrix that sub-samples the high-resolution image, $\mathbf{H}$ is an $PMPN$x$PMPN$ matrix that filters the high-resolution image, T_{DCT} and $T^{1}{}_{DCT}$ are the forward and inverse DCT calculations, and Q represents the quantization procedure.

Let $\mathbf{F}$ be the vector $\left(\mathbf{f}_1^T, \mathbf{f}_2^T, \ldots, \mathbf{f}_k^T, \ldots, \mathbf{f}_L^T\right)^T$ that contains all the high-resolution frames and let $\mathbf{Y}$ be the vector that contains all the low-resolution frames $\left(\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_k^T, \ldots, \mathbf{y}_L^T\right)^T$. We propose to follow a maximum *a posteriori* (MAP) estimation approach in recovering the high resolution information from the low resolution compressed observations. Towards this task, we will use the following approximation for the conditional distribution of the observed low resolution images given the high resolution sequence

$$p(\mathbf{Y}\,|\,\mathbf{F}) = p(\mathbf{y}_1,\ldots,\mathbf{y}_L\,|\,\mathbf{f}_1,\ldots,\mathbf{f}_L) = \prod_i p(\mathbf{y}_i\,|\,\mathbf{f}_i)\,, \tag{4}$$

where

$$p(\mathbf{y}_i\,|\,\mathbf{f}_i) \propto \exp\left\{-\lambda_4 \left\|\mathbf{y}_i - \mathbf{AHf}_i\right\|^2\right\}. \tag{5}$$

This conditional distribution enforces similarity between the compressed low resolution image and its high resolution image (through a process of blurring and downsampling, represented by **H** and **A** respectively). With λ_4 we control this resemblance.

In this paper we assume that the high resolution motion vectors $\mathbf{d}_{l,k}$ have been previously estimated (see [6] for different approaches to perform this task).

In the literature about motion estimation there are methods based on optical flow (see [10] and [11]), block matching [12], and feature matching. Simoncelli in [13] uses the optical flow equation but also adds an uncertainty model to solve the extended aperture problem and a Gaussian pyramid to deal with big displacements. Another interesting method was proposed by Irani and Peleg (see [14]) using an object based approach. The motion parameters and the location of the objects (it is supposed that there are several moving objects in the image sequence) are computed sequentially taking into account only one object at a time by using segmentation. A Gaussian pyramid from coarse to finer resolution is also used to avoid problems with the displacements.

In our implementation the motion field has been computed, for all the compressed low resolution frames, mapping the previous frame into the current one, and then interpolating the resulting low resolution motion field to obtain the high resolution motion field. Better motion field estimation procedures, which probably would provide better reconstruction results, are currently under study.

From equation (2) ,assuming smoothness within the high resolution images and trying to remove the blocking artifacts in the low resolution uncompressed images, we use the following prior model to describe the relationship between the high resolution images:

$$p(\mathbf{f}_1,\ldots,\mathbf{f}_L) \propto \exp\left\{-\lambda_1 \sum_{i=2}^{L} \left\|\mathbf{f}_{i-1} - C(\mathbf{d}_{i-1,i})\mathbf{f}_i\right\|^2 - \lambda_2 \sum_{i=1}^{L} \left\|\mathbf{Q}_1\mathbf{f}_i\right\|^2 - \lambda_3 \sum_{i=1}^{L} \left\|\mathbf{Q}_2\mathbf{AHf}_i\right\|^2\right\}. \tag{6}$$

In the first term of the above prior distribution we are including the quality in the prediction (if the prediction of our frame $\mathbf{f}_i$ from the previous one is a good prediction, this term will be small). The second and third terms represent smoothness constraints, where $\mathbf{Q}_1$ represents a linear high-pass operation, $\mathbf{Q}_2$ represents a linear high-pass operation across block boundaries, and λ_2 and λ_3 control the influence of the two norms. By increasing the value of λ_2, the density describes a smoother image frame, while increasing the value of λ_3 results in a frame with smooth block boundaries.

After having defined the prior and degradation models, several points are worth mentioning. First, note that the degradation process (see equations (4) and (5)) relate each low resolution observation to its corresponding high resolution one and so no prediction of high resolution images is included in it. Due to this fact, this model is different from the one currently used in most high resolution methods (see Segall et al. [6] for a review). Note also that the prior model is responsible for relating the high resolution images and so a change in a high resolution image will enforce (through the prior model) changes in the other high resolution images. Finally, note that, although it is also possible to include prior models over the high-resolution motion vectors, in this work we assume that they have been estimated previously to the reconstruction process, see however [15–16] for the simultaneous estimation of high-resolution motion and images. Work on prior motion models which are consistent over time will be reported elsewhere.

3 Problem Formulation and Proposed Algorithm

The maximum *a posteriori* (MAP) estimate provides the necessary framework for recovering high-resolution information from a sequence of compressed observations. Following the Bayesian paradigm the MAP high resolution sequence reconstruction satisfies

$$\hat{\mathbf{F}} = \arg\max_{\mathbf{F}} \{p(\mathbf{F})p(\mathbf{Y}\,|\,\mathbf{F})\}\,. \tag{7}$$

Applying logarithms to equation (7) we find that the high resolution image sequence estimate $\hat{\mathbf{F}}$ satisfies

$$\hat{\mathbf{F}} = \arg\min_{\mathbf{F}} \left\{ \lambda_1 \sum_{i=2}^{L} \left\| \mathbf{f}_{i-1} - C(\mathbf{d}_{i-1,i})\mathbf{f}_i \right\|^2 + \lambda_2 \sum_{i=1}^{L} \left\| \mathbf{Q}_1 \mathbf{f}_i \right\|^2 + \lambda_3 \sum_{i=1}^{L} \left\| \mathbf{Q}_2 \mathbf{A}\mathbf{H}\mathbf{f}_i \right\|^2 + \right. \\ \left. + \lambda_4 \sum_{i=1}^{L} \left\| \mathbf{y}_i - \mathbf{A}\mathbf{H}\mathbf{f}_i \right\|^2 \right\} \tag{8}$$

In order to find the MAP we propose the following iterative procedure. Let $\mathbf{F}^0$ be an initial estimate of the high resolution sequence. Then given the sequence $\left((\mathbf{f}_1^n)^T, (\mathbf{f}_2^n)^T, \ldots, (\mathbf{f}_k^n)^T, \ldots, (\mathbf{f}_L^n)^T\right)^T$ we obtain, for *l=1,…,L* the high resolution image $\mathbf{f}_l^{n+1}$ at step *n+1* by using the following equation

$$\mathbf{f}_l^{n+1} = \mathbf{f}_l^n - \alpha_{f_l} \left\{ -\lambda_1 C^T(\mathbf{d}_{l-1,l}) \left[\mathbf{f}_{l-1}^{n+1} - C(\mathbf{d}_{l-1,l})\mathbf{f}_l^n\right] + \lambda_1 \left[\mathbf{f}_l^n - C(\mathbf{d}_{l,l+1})\mathbf{f}_{l+1}^n\right] + \right. \\ \left. + \lambda_2 \mathbf{Q}_1^T \mathbf{Q}_1 \mathbf{f}_l^n + \lambda_3 \mathbf{H}^T \mathbf{A}^T \mathbf{Q}_2^T \mathbf{Q}_2 \mathbf{A}\mathbf{H}\mathbf{f}_l^n - \lambda_4 \mathbf{H}^T \mathbf{A}^T (\mathbf{y}_l - \mathbf{A}\mathbf{H}\mathbf{f}_l^n) \right\} \tag{9}$$

The relaxation parameter α_{fl} determines convergence as well as the rate of convergence of the iteration. It is important to note that for the first and last frames in the sequence, $\mathbf{f}_1$ and $\mathbf{f}_L$ respectively, the frames $\mathbf{f}_0$ and $\mathbf{f}_{L+1}$ do not exist and so the above equation has to be adapted by removing the presence of $\mathbf{f}_0$ and $\mathbf{f}_{L+1}$ respectively.

4 Experimental Results

The performance of the algorithm is illustrated by processing frames from the *Mobile* sequence. Each original image is *704x576* pixels and it is decimated by a factor of two in each dimension, cropped to a size of *176x144* pixels and compressed with an *MPEG-4* encoder operating at *1024Kbps*. Three frames from the compressed bit-stream are then sequentially provided to the proposed algorithm, $\mathbf{Q}_1$ is a *3x3* discrete Laplacian, $\mathbf{Q}_2$ is a difference operation across the horizontal and vertical block boundaries, and the model parameters were experimentally chosen to be $\lambda_1=100$, $\lambda_2=0.01$, $\lambda_3=0.002$, $\lambda_4=1$, and $\alpha_{fl}=0.125$. The algorithm is terminated when

$$\left\|\hat{\mathbf{F}}_{k+1}-\hat{\mathbf{F}}_k\right\|^2/\left\|\hat{\mathbf{F}}_k\right\|^2 < 1x10^{-7} \quad \textbf{(10)}$$

The performance of the algorithm is defined in terms of the improvement in signal-to-noise ratio, defined by

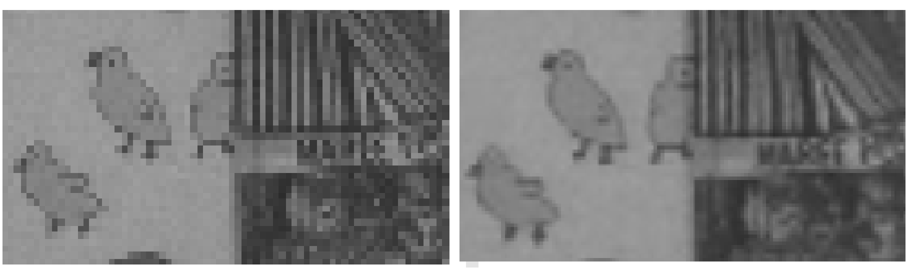
$$ISNR = 10\log_{10}\left(\left\|\mathbf{F}-\overline{\mathbf{Y}}\right\|^2/\left\|\mathbf{F}-\hat{\mathbf{F}}\right\|^2\right), \quad \textbf{(11)}$$

where $\overline{\mathbf{Y}}$ is the zero-order hold of image $\mathbf{Y}$. A representative algorithm result is presented in the Figs. 1-4. The original image is shown in Fig. 1, the compressed observation after bi-linear interpolation in Fig. 2, and the image provided by the proposed algorithm is depicted in Fig. 3. Fig. 4 zooms a part of the image obtained by bi-linear interpolation (Fig. 4a) and our proposed method (Fig. 4b). The sign "Maree" unreadable in the image shown in Fig. 4a whereas is almost readable in Fig. 4b. The smoothness constraint also performs well, as can be observed in the left area of the images. The corresponding ISNR values for Fig. 2 and 3 are 30.4123dB and 31.1606dB, respectively. These figures, as well as the visual inspection, demonstrate the improvement obtained by the proposed algorithm. Figure 5 plots the value of the stopping criterium of the algorithm (see equation (10)) as a function of the number of iteration k, demonstrating the convergence of the algorithm.

Fig. 1. Cropped part of the original image from the sequence before decimation and compression. The proposed method's aim is to estimate all these images at the same time.

Fig. 2. Decoded observations after bi-linear interpolation. The compression artifacts are easily noticeable.

Fig. 3. Image obtained by the proposed method. The comparison should be established between this figure and Fig. 2.



a) **b)**

Fig. 4. a) Decoded image after bi-linear interpolation. **b)** The improvement achieved by the method. The sign "Maree" unreadable in the image a) whereas is almost readable in b). The smoothness constraint works well, as can be observed in the left area of the images.

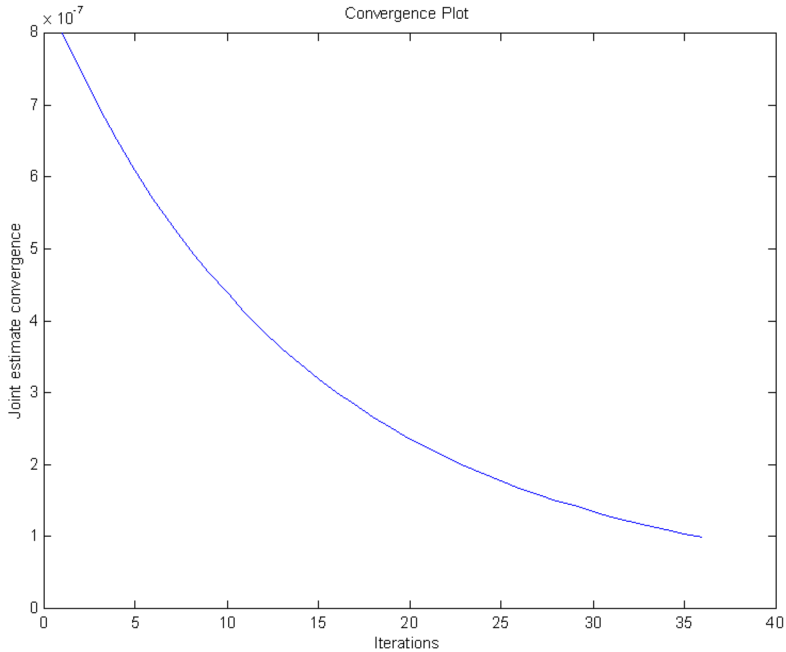


Fig. 5. Convergence plot of the iterative procedure. The implemented method guarantees convergence.

5 Conclusions

In this paper we have proposed a new iterative procedure to estimate a high resolution video sequence from low resolution observations. The method uses fidelity to the low resolution data and smoothness constraints whithin and between the high resolution images to estimate the sequence. Incorporating temporal coherence of the high resolution motion vectors as well as the development of a parallel implementation of the algorithm are currently under study. The proposed method has been experimentally validated.

References

1. Chen, D., Schultz, R.R.: Extraction of High-Resolution Still from MPEG Sequences. Proceedings of the IEEE ICIP, Chicago, IL, (1998) 465–469
2. Altunbasak, Y., Patti, A.J., Mersereau, R.M.: Super-resolution still and video reconstruction from mpeg-coded video. IEEE Transactions on Circuits and Systems for Video Technology, vol. 12 (2002) 217–226

3. Segall, C.A., Molina, R., Katsaggelos, A.K., Mateos, J.: Bayesian High-Resolution Reconstruction of Low-Resolution and Compressed Video. Proceedings of the IEEE ICIP, vol 2, Thessaloniki, Greece (2001) 21–25
4. Segall, C. A., Katsaggelos, A. K., Molina, R., Mateos, J.: Super-Resolution from Compressed Video. In: Chaudhuri, S.: Super-Resolution Imaging, Kluwer Academic Publishers (2001) 211–242
5. Gunturk, B.K., Altunbasak, Y., Mersereau, R.M.: Multiframe Resolution-Enhancement Methods for Compressed Video. IEEE Signal Processing Letters, vol. 9 (2002) 170–174
6. Segall, C. A., Molina, R., Katsaggelos, A. K.: High Resolution Images from a Sequence of Low Resolution and Compressed Observations: A Review. IEEE Signal Processing Magazine, vol. 20, no. 3 (2003) 37–48
7. Choi, M. C., Yang, Y., Galatsanos, N. P.: Multichannel Regularized Recovery of Compressed Video Sequences. IEEE Trans. on Circuits and Systems II, vol. 48 (2001) 376–387
8. Elad, M., Feuer, A.: Super-Resolution Reconstruction of Image Sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 21 (1999) 817–834
9. Elad, M., Feuer, A.: Super-Resolution Restoration of an Image Sequence: Adaptive Filtering Approach. IEEE Transactions on Image Processing, vol 8 (1999) 387–395
10. Horn, B. K. P., Schunk, B. G.: Determining Optical Flow. Artificial Intelligence, vol. 17 (1981) 185–203
11. Lucas, B. D., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. Proceedings of Imaging Understanding Workshop (1981) 121–130
12. Trucco, E., Verri, A.: Introductory Techniques for 3-D Computer Vision. Prentice Hall (1998)
13. Simoncelli, E. P.: Bayesian Multi-Scale Differential Optical Flow: In: Jähne, B., Haussecker, H., Geissler, P.: Handbook of Computer Vision and Applications, Academic Press, vol. 2 (1999) 97–402
14. Irani, M., Rousso, B., Peleg, S.: Computing Occluding and Transparent Motions. Int. J. Computer Vision, vol. 12, no. 1 (1994) 5–16
15. Hardie, R. C., Barnard, K. J., Armstrong, E. E.: Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images. IEEE Transactions on Image Processing, vol. 6, no. 12 (1997) 1621–1633
16. Hardie, R. C., Barnard, K. J., Bognar, J. G., Armstrong, E. E., Watson, E. A.: High-Resolution Image Reconstruction from a Sequence of Rotated and Translated Frames and its Application to an Infrared Imaging System. Optical Engineering, vol. 73, no. 1 (1998) 247–260