MOTION ESTIMATION IN HIGH RESOLUTION IMAGE RECONSTRUCTION FROM COMPRESSED VIDEO SEQUENCES

*L.D. Alvarez[a], Rafael Molina[a] and Aggelos K. Katsaggelos[b] **

a) Dpto. de Ciencias de la Computación e I.A., Universidad de Granada, 18071 Granada, España
b) Dept. of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208-3118

ABSTRACT

In order to obtain a high resolution image from a compressed video sequence it is essential to correctly estimate the motion vectors in the sequence. Most of the approaches reported in the literature address this problem using standard motion estimation techniques. In this paper we tackle the correct estimation of the motion vectors by consistently estimating the optical flow across multiple images. Consistency is achieved by adding a regularization term to the classical Lucas-Kanade approach to motion estimation. The proposed algorithm is tested on real video sequences.

1. INTRODUCTION

Super resolution algorithms increase the resolution of an image without changing the resolution of the image sensor. This is accomplished by exploiting the underlying motion of a video sequence to provide multiple observations for each frame, and it mitigates the requirements for transporting and storing a high resolution sequence. Accurate motion estimation is essential in super resolution problems.

Most of the reported work on motion estimation for super resolution first estimate the displacement vectors between the images in the sequence either by first interpolating the low resolution observations and then finding the motion vectors or by first finding the low resolution motion vectors in the low resolution domain and then interpolating them. So, classical motion estimation techniques can be applied to the process of finding the high resolution motion vectors (see [1] for a review). Registration work developed specifically for the low to high resolution problem, in which in some cases the high resolution image and the registration parameters are estimated simultaneously, can be found in [2, 3, 4, 5].

The high resolution image reconstruction problem is further complicated when the available low resolution video is compressed [6, 7]. This is the case in many applications of interest. In this paper, starting from the low to high resolution method described in [8] we propose a new iterative method to consistently estimate the motion vectors from compressed low resolution video data.

The rest of the paper is organized as follows. The process to obtain a compressed low resolution video sequence from high resolution images is described in section 2. Prior information on the high resolution image we want to reconstruct is described in section 3. The process to estimate the motion between the images is described in section 4. Once the registration parameters have been obtained the application of the Bayesian paradigm to calculate the MAP high resolution image is described in section 5. Experimental results are described in section 6. Finally, section 7 concludes the paper.

2. OBTAINING LOW RESOLUTION COMPRESSED OBSERVATIONS FROM HIGH RESOLUTION IMAGES

The underlying high resolution (HR) video sequence is denoted by $\mathbf{f} = \{\mathbf{f}_1, \ldots, \mathbf{f}_k, \ldots, \mathbf{f}_L\}$, where the size of the high resolution images $\mathbf{f}_l, l = 1, \ldots, L$ is $PM \times PN$, ($P > 1$ being the magnification factor). Using matrix-vector notation, the $PM \times PN$ images can be transformed into a $PM \times PN$ column vector, obtained by lexicographically ordering the image by rows. The $(PM \times PN) \times 1$ vector that represents the lth image in the HR sequence will also be denoted by $\mathbf{f}_l$. The high resolution image we are trying to estimate will be denoted by $\mathbf{f}_k$.

Frames within the HR sequence are related through time. Here we assume that the camera captures the images in a fast succession and so we write

$$f_l(\mathbf{x}) = f_k(\mathbf{x} + \mathbf{d}_{l,k}(\mathbf{x})) + n_{l,k}(\mathbf{x}) \ , \qquad (1)$$

where $\mathbf{x} = (x, y)$ denotes pixel location, the vector containing the horizontal and vertical components of the displacement is denoted by $\mathbf{d}_{l,k}(\mathbf{x}) = (d^x_{l,k}(\mathbf{x}), d^y_{l,k}(\mathbf{x}))$, and $n_{l,k}(\mathbf{x})$ is the noise introduced by the motion compensation process. The above equation relates a gray level pixel value at location $\mathbf{x}$ at time l to the gray level pixel value of its corresponding pixel in the high resolution image we want to estimate $\mathbf{f}_k$.

*This work has been partially supported by the "Comisión Nacional de Ciencia y Tecnología" under contract TIC2003-00880.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

We can rewrite (1) in matrix-vector notation as

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

where $\mathbf{C}(\mathbf{d}_{l,k})$ is the $(PM \times PN) \times (PM \times PN)$ matrix that maps frame $\mathbf{f}_l$ to frame $\mathbf{f}_k$, and $\mathbf{n}_{l,k}$ is the registration noise.

The HR sequence, through filtering and downsampling, produces an unobserved uncompressed low resolution (LR) sequence. This unobserved LR discrete sequence will be denoted by $\mathbf{g} = \{\mathbf{g}_1, \ldots, \mathbf{g}_L\}$. The size of the LR images $\mathbf{g}_l, l = 1, \ldots, L$ is $M \times N$. Matrix-vector notation will also be used for this sequence. Each LR image $\mathbf{g}_l$, $l = 1 \ldots, L$ is related to the corresponding HR image $\mathbf{f}_l$ by

$$\mathbf{g}_l = \mathbf{AHf}_l + \nu_l \qquad l = 1, 2, 3, ... \ , \tag{3}$$

where $\mathbf{H}$ of size $(PM \times PN) \times (PM \times PN)$ describes the filtering of the HR image, $\mathbf{A}$ is the down-sampling matrix with size $MN \times (PM \times PN)$ and ν_l denotes the acquisition noise. We assume here for simplicity that all the blurring matrices $\mathbf{H}$ are the same, although they can be time dependent. Matrices $\mathbf{A}$ and $\mathbf{H}$ are assumed to be known.

Using (2) and (3) we obtain the following equation that provides us the acquisition system for a LR image from the high resolution image $\mathbf{f}_k$ that we want to estimate

$$\mathbf{g}_l = \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k + \mathbf{e}_{l,k}, \tag{4}$$

where $\mathbf{e}_{l,k}$ is the combination of the registration and acquisition noise.

The LR frames are now compressed with a hybrid motion compensation video compression system resulting in $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_L\}$. The size of the LR compressed images is $M \times N$. The compression system also provides the motion vectors $\mathbf{v}_{l,m}$ that at position (i, j) predict the pixel $\mathbf{y}_l(i, j)$ from some previously coded $\mathbf{y}_m$. These motion vectors that predicts $\mathbf{y}_l$ from $\mathbf{y}_m$ are represented by the $(2 \times M \times N) \times 1$ vector that is formed by stacking the transmitted horizontal and vertical offsets.

During compression frames are divided into blocks that are encoded with one of two available methods, intracoding or intercoding (see [9] for details). The relationship between the acquired low resolution frame and its compressed observation becomes

$$\begin{aligned} \mathbf{y}_l \ &= \ T^{-1}Q\left[T\left(\mathbf{g}_l - MC_l(\mathbf{y}_l^P, \mathbf{v}_l)\right)\right] + MC_l(\mathbf{y}_l^P, \mathbf{v}_l) \\ &\quad l = 1, \ldots, L, \end{aligned} \tag{5}$$

where $Q[.]$ represents the quantization procedure, T and T^{-1} are the forward and inverse-transform operations, respectively, and $MC_l(\mathbf{y}_l^P, \mathbf{v}_l)$ is the motion compensated prediction of $\mathbf{g}_l$ formed by motion compensating the appropriate previously decoded frame/frames depending on whether the current frame at l is an I, P or B frame. Note that, to be precise, we should make clear that MC_l depends on $\mathbf{v}_l$ and only a subset of $\mathbf{y}_1, \ldots, \mathbf{y}_L$. However, we will keep the above notation for simplicity and generality.

Using (4) and (5) we have for the distribution of the high resolution image we want to estimate [6],

$$\begin{aligned} &P(\mathbf{y}_l|\mathbf{f}_k, \mathbf{d}_{l,k}) \propto \\ &\exp\left[-\frac{1}{2}(\mathbf{y}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)^T\mathbf{K}_Q^{-1}(\mathbf{y}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)\right], \end{aligned} \tag{6}$$

where $\mathbf{K}_Q$ is the covariance matrix that describes the noise.

Following [7] we model the displaced frame difference within the encoder using the distribution

$$\begin{aligned} &P(\mathbf{v}_{l,k}|\mathbf{f}_k, \mathbf{d}_{l.k}, \mathbf{y}_l) \propto \\ &\quad \exp[-\frac{1}{2}(MC_l(\mathbf{y}_l^P, \mathbf{v}_l) - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)^T\mathbf{K}_{MV}^{-1} \\ &\qquad (MC_l(\mathbf{y}_l^P, \mathbf{v}_l) - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)\big] \ , \end{aligned} \tag{7}$$

where $\mathbf{K}_{MV}$ is the covariance matrix for the prediction error between the frame $\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k$ and its motion compensated estimate $MC_l(\mathbf{y}_l^P, \mathbf{v}_l)$.

Note that from the observation model of the low resolution compressed images and motion vectors we have described, we can write the joint observational model of the low resolution compressed images and low resolution motion vectors given the high resolution image and motion vectors as

$$P(\mathbf{y}, \mathbf{v}|\mathbf{f}_k, \mathbf{d}) = \prod_l P(\mathbf{y}_{l,k}|\mathbf{f}_k, \mathbf{d}_{l,k})P(\mathbf{v}_{l,k}|\mathbf{f}_k, \mathbf{d}_{l,k}, \mathbf{y}_l). \tag{8}$$

In this paper we assume that the high resolution motion vectors are estimated prior to the estimation of $\mathbf{f}_k$ by one of the methods to be described in section 4 and so we have

$$P(\mathbf{y}, \mathbf{v}|\mathbf{f}_k) = \prod_l P(\mathbf{y}_l|\mathbf{f}_k)P(\mathbf{v}_{l,k}|\mathbf{f}_k, \mathbf{y}_l). \tag{9}$$

3. REGULARIZATION IN HR

The distribution we use for $\mathbf{f}_k$, $P(\mathbf{f}_k)$ reflects the facts that we expect the images to be smooth within homogeneous regions and also that the LR image obtained from the HR image should be free of blocking artifacts. So, we have [6]

$$P(\mathbf{f}_k) \propto \exp\left[-\left(\frac{\lambda_3}{2}\|\mathbf{Q}_3\mathbf{f}_k\|^2 + \frac{\lambda_4}{2}\|\mathbf{Q}_4\mathbf{AHf}_k\|^2\right)\right] \ , \tag{10}$$

where $\mathbf{Q}_3$ represents a linear high-pass operation that penalizes non-smooth estimates, $\mathbf{Q}_4$ represents a linear high-pass operator that penalizes estimates with block boundaries and λ_3 and λ_4 control the weight of the norms.

4. MOTION ESTIMATION

In this section we describe the motion estimation method we use to estimate the high resolution motion vectors in (1).

Since the high resolution sequence, $\mathbf{f}$, is not available we use an initial estimate of $\mathbf{f}$ to find the motion vectors. Most results reported in the literature to obtain an initial estimate of $\mathbf{f}$, upsample the low resolution observations by using tipically bilinear interpolation. However, we note that this process does not take into account the blurring in the sequence $\mathbf{y}$. In this paper we use the method described in [8], according to which dowsampling and blurring in just one frame is removed, with the addition of smoothness to its high resolution version.

Let us denote by $\overline{\mathbf{f}} = \{\overline{\mathbf{f}}_1, \ldots, \overline{\mathbf{f}}_k, \ldots, \overline{\mathbf{f}}_L\}$ the initial estimate of the high resolution sequence. We now proceed to estimate the high resolution motion vector from this sequence.

For each $\mathbf{x}_0 = (x_0, y_0)$ the goal of the Lucas-Kanade algorithm [10] when applied to our problem becomes finding $\mathbf{d}_{l,k}(\mathbf{x}_0) = (d^x_{l,k}(\mathbf{x}_0), d^y_{l,k}(\mathbf{x}_0))$ that minimizes

$$\sum_{\mathbf{x}\in\mathcal{N}_{\mathbf{x}_0}} \left(\overline{f}_l(\mathbf{x}) - \overline{f}_k(\mathbf{x} + \mathbf{d}_{l,k}(\mathbf{x}_0))\right)^2 \qquad (11)$$

where $\mathcal{N}_{\mathbf{x}_0}$ denotes a set of neighbouring pixels of $\mathbf{x}_0$.

The estimation process works iteratively. Given a current estimate $\mathbf{d} = (d_1, d_2)$ of $\mathbf{d}_{l,k}(\mathbf{x}_0)$, $\Delta\mathbf{d} = (\Delta d_1, \Delta d_2)$ is found by minimizing

$$\sum_{\mathbf{x}\in\mathcal{N}_{\mathbf{x}_0}} \left(\overline{f}_l(\mathbf{x}) - \overline{f}_k(\mathbf{x} + \mathbf{d} + \Delta\mathbf{d})\right)^2 \qquad (12)$$

where

$$\begin{aligned}\overline{f}_k(\mathbf{x} + \mathbf{d} + \Delta\mathbf{d}) &\approx \overline{f}_k(\mathbf{x} + \mathbf{d}) \\ &+(\overline{f}_k)_x(\mathbf{x} + \mathbf{d})\Delta d_1 + (\overline{f}_k)_y(\mathbf{x} + \mathbf{d})\Delta d_2\end{aligned}$$

The idea of using consistent high resolution motion estimation has not been seen much coverage in the literature (see however [11] and [3]). Here we consider the following model to enforce consistency in the high resolution motion vectors.

In order to constrain the trajectory of the motion we first find $\mathbf{d}_{l,l+1}$, $l = 1, \ldots, k-1$, and $\mathbf{d}_{l+1,l}$, $l = k, \ldots, L-1$, by using the Lucas-Kanade [10] algorithm.

For $l = k-2, \ldots, 1$ we then estimate $\mathbf{d}_{l,k}$, once $\mathbf{d}_{l+1,k}$ has already been calculated, by minimizing

$$\begin{aligned}L(\mathbf{d}_{l,k}(\mathbf{x}_0)) = \lambda \parallel \mathbf{d}_{l,k}(\mathbf{x}_0) - \mathbf{d}_{l+1,k}(\mathbf{x}_0 + \mathbf{d}_{l,l+1}(\mathbf{x}_0)) \parallel^2 \\ + \mu \sum_{\mathbf{x}\in\mathcal{N}_{\mathbf{x}_0}} \left(\overline{f}_l(\mathbf{x}) - \overline{f}_k(\mathbf{x} + \mathbf{d}_{l,k}(\mathbf{x}_0))\right)^2 \quad (13)\end{aligned}$$

where λ and μ are weighting parameters.

For $l = k+2, \ldots, L$ we estimate $\mathbf{d}_{l,k}$, once $\mathbf{d}_{l-1,k}$ has already been calculated, by minimizing

$$\begin{aligned}L(\mathbf{d}_{l,k}(\mathbf{x}_0)) = \lambda \parallel \mathbf{d}_{l,k}(\mathbf{x}_0) - \mathbf{d}_{l-1,k}(\mathbf{x}_0 + \mathbf{d}_{l,l-1}(\mathbf{x}_0)) \parallel^2 \\ + \mu \sum_{\mathbf{x}\in\mathcal{N}_{\mathbf{x}_0}} \left(\overline{f}_l(\mathbf{x}) - \overline{f}_k(\mathbf{x} + \mathbf{d}_{l,k}(\mathbf{x}_0))\right)^2 \quad (14)\end{aligned}$$

where λ and μ are weighting parameters.

5. ESTIMATING HIGH RESOLUTION IMAGES

Having described in the previous sections the high resolution image prior, the acquisition model and the estimation of the high resolution motion vectors, we turn now our attention to computing the high resolution frame.

We aim at finding the maximum of the posterior distribution of the high resolution image given the observations.

For compressed sequences our goal becomes finding $\hat{\mathbf{f}}_k$ that satisfies

$$\hat{\mathbf{f}}_k = \arg\max_{\mathbf{f}_k} P(\mathbf{f}_k)P(\mathbf{y}, \mathbf{v}|\mathbf{f}_k) \ , \qquad (15)$$

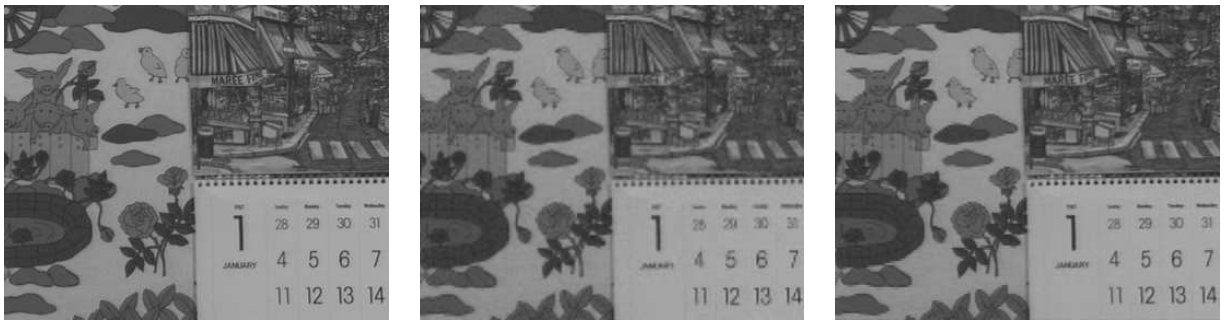
where the distribution of high resolution intensities is given in section 3 and the acquisition model is described in section 2.

The solution of (15) can be found using gradient descent techniques.

6. EXPERIMENTS

To examine the performance of the proposed methodology to estimate the high resolution motion vectors and then the high resolution image, we have used frames from the Mobile sequence. The size of each original image in the sequence is 704x576 pixels. The images are decimated by a factor of two (in each dimension), then they are cropped (central part) to a size of 176x144 pixels and compressed with an MPEG-4 encoder operating at 1024Kbps. We will reconstruct frame 11 using three previous and three future frames.

The original image is shown in figure 1a. The compressed observation after bi-linear interpolation is shown in figure 1b, and the image obtained adding consistency to the optical flow estimate, according to (13) and (14), is shown in figure 1c.



(a) (b) (c)

Fig. 1. a) Original 352×288 high-resolution image, b) bi-linear interpolation of the compressed low resolution frame 11, and c) estimated high resolution image using the consistent motion estimation method in (13) and (14).

The performance of the proposed algorithm was evaluated by measuring the peak signal-to-noise ratio (PSNR) defined as PSNR$= 10 \times \log_{10}[352 \times 288 \times 255^2 / \parallel \mathbf{f} - \hat{\mathbf{f}} \parallel^2]$, where $\mathbf{f}$ and $\hat{\mathbf{f}}$ are the original and estimated high resolution images, respectively.

The PSNR of the bi-linear interpolation was 24.5316dB. We obtained a PSNR of 30.24dB when using to estimate the motion vectors non-overlapping block-matching with squared error as comparison criterion, a window size of 3×3 and a search area of 2 pixels in each direction. The chosen parameters for the regularized Lucas-Kanade algorithm in (13) and (14) were $\lambda = 0.1$ and $\mu = 0.9$. The PSNR of the reconstruction was 32.92dB.

7. CONCLUSIONS

In this paper we have presented a new method to consistently estimate the motion vectors in order to reconstruct a high resolution image from a sequence of compressed low resolution observations. Consistency of the estimated optical flow across multiple images has been imposed by adding a regularization term to the classical Lucas-Kanade approach to motion estimation. The proposed method has been tested on real sequences.

8. REFERENCES

[1] M. G. Kang and S. Chaudhuri, Eds., *Super-resolution image reconstruction*, IEEE Signal Processing Magazine, vol. 20, no. 3, 2003.

[2] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," in *Proceedings of the IEEE International Conference on Image Processing*, 1995, vol. 2, pp. 539–542.

[3] W. Zhao and H. Sawhney, "Is super-resolution with optical flow feasible?," in *In Proc. European Conf. Computer Vision, LNCS 2350*, 2002, pp. 599–613.

[4] N.A. Woods, N.P. Galatsanos, and A.K. Katsaggelos, "EM-based simultaneous registration, restoration, and interpolation of super resolved images," in *Proceedings of the IEEE International Conference on Image Processing*, 2003, vol. II, pp. 303–306.

[5] C. M. Bishop, A. Blake, and B. Marthi, "Super-resolution enhancement of video," in *C. M. Bishop and B. Frey (Eds.), Proceedings Artificial Intelligence and Statistics*, 2003.

[6] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Processing Magazine*, vol. 20, pp. 37–48, 2003.

[7] C.A. Segall, R. Molina, A.K. Katsaggelos, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Transactions on Image Processing, to appear*, 2004.

[8] R. Molina, M. Vega, J. Abad, and A.K. Katsaggelos, "Parameter estimation in bayesian high-resolution image reconstruction with multisensors," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1655–1667, December 2003.

[9] A.M. Tekalp, *Digital Video Processing*, Prentice Hall, Signal Processing Series, 1995.

[10] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, 2004.

[11] J.C. Brailean and A.K. Katsaggelos, "A recursive nonstationary map displacement vector field estimation algorithm," *IEEE Transactions on Image Processing*, vol. 4, pp. 416–429, 1995.