IMPROVED DEFINITION VIDEO FRAME ENHANCEMENT

Richard R. Schultz and Robert L. Stevenson

Laboratory for Image and Signal Analysis
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556, USA

ABSTRACT

The human visual system seems to be capable of temporally integrating information in a video sequence in such a way that the perceived spatial resolution of a sequence appears much higher than the spatial resolution of an individual frame. This paper addresses how to utilize both the spatial and temporal information present in an image sequence to create a high-resolution video still. A novel observation model based on motion compensated subsampling is proposed for a video sequence. Since the reconstruction problem is ill-posed, Bayesian restoration with an edge-preserving prior image model is used to extract a high-resolution video frame from a low-resolution sequence. Estimates computed from an image sequence containing a camera pan show dramatic improvement over bilinear, cubic B-spline, and Bayesian single frame interpolations. Improved definition is also shown for a video sequence containing objects moving with independent trajectories.

1. INTRODUCTION

Image interpolation techniques have been researched quite extensively, with the zero-order hold, bilinear interpolation, cubic B-spline interpolation [1], and regularization methods [2], [3] providing progressively more accurate solutions. However, the quality of estimates generated by these methods is inherently limited by the number of constraints available within a single image. For this reason, multiframe methods have been proposed which use the additional data present within a sequence of temporally-correlated frames.

Multiframe image restoration was introduced by Tsai and Huang [4]. Their motivation came from generating a high-resolution frame from misregistered Landsat images. Provided that enough frames are available with different subpixel global displacements, their observation mapping becomes invertible. If this is not the case, a least squares approximation may be computed through a pseudoinverse of the constraint matrix. An extension of this algorithm for noisy data was provided by Kim *et al.* [5], resulting in a weighted least squares algorithm. Stark and Oskoui [6] formulated a projection onto convex sets (POCS) algorithm to compute an estimate from observations obtained by scanning or rotating an image with respect to the CCD image acquisition sensor array. Tekalp *et al.* [7] extended this POCS formulation to include sensor noise and later time-varying motion blur [8].

In this research, the ill-posed inverse problem of interpolation is placed into a Bayesian framework. The enhancement algorithm incorporates several ideas which improve the usability and quality of the estimated image frame. First, independent object motion in the video sequence will be assumed, rather than the simple cases of global displacement or rotation assumed in previous multiframe methods. Next, an edge-preserving image prior will be used to regularize the interpolation problem. This has the intent on improving upon least squares and POCS solutions, which typically contain smooth edges. Finally, the multiframe interpolation algorithm proposed in this paper reduces to the Bayesian method presented previously [3] when only a single frame is available.

The paper will be organized as follows. Section 2 proposes a novel observation model for a low-resolution video sequence. The video frame enhancement algorithm is formulated in a Bayesian framework in Section 3, including a discontinuity-preserving prior model for the data and a density for the modeling error. Simulation results are described in Section 4 for a synthetically-generated sequence containing camera panning and a real video sequence containing objects moving independently. Section 5 provides a brief summary of the research.

2. VIDEO OBSERVATION MODEL

The video frame enhancement problem is stated in this section, and an observation model is proposed for a video sequence which includes motion compensated subsampling.

2.1. Problem Statement

The objective is to estimate a high-resolution frame, by reconstructing the high-frequency components of the image lost through undersampling the data. Assume that each frame in a low-resolution image sequence contains $N_1 \times N_2$ square pixels. A lexicographical ordering of the l^{th} frame

Effort sponsored by Rome Laboratory, Air Force Materiel Command, USAF under grant number F30602-94-1-0017. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Rome Laboratory or the U.S. Government.

0-7803-2431-5/95 $4.00 © 1995 IEEE

results in the $N_1N_2 \times 1$ vector denoted as $\mathbf{y}^{(l)}$. Consider a low-resolution video subsequence

$$\left\{\mathbf{y}^{(l)}\right\} \quad \text{for } l = k - \frac{M-1}{2}, \ldots, k, \ldots, k + \frac{M-1}{2}, \quad (1)$$

where M represents an odd number of frames. A single high-resolution frame $\mathbf{z}^{(k)}$ coincident with the center frame $\mathbf{y}^{(k)}$ is to be estimated from the low-resolution subsequence. This unknown high-resolution data consists of $qN_1 \times qN_2$ square pixels, where q is an integer-valued interpolation factor in both the horizontal and vertical directions. Thus, $\mathbf{z}^{(k)}$ is a $q^2N_1N_2 \times 1$ lexicographically-ordered vector.

2.2. Center Frame Subsampling Model

Subsampling for the center frame is accomplished by averaging a square block of high-resolution pixels,

$$y_{i,j}^{(k)} = \frac{1}{q^2}\left(\sum_{r=qi-q+1}^{qi}\sum_{s=qj-q+1}^{qj} z_{r,s}^{(k)}\right), \quad (2)$$

for $i = 1, \ldots, N_1$ and $j = 1, \ldots, N_2$. This models the spatial integration of light intensity over a square surface region performed by CCD image acquisition sensors [3]. The center frame observation model is given as

$$\mathbf{y}^{(k)} = \mathbf{A}^{(k,k)}\mathbf{z}^{(k)}, \quad (3)$$

where $\mathbf{A}^{(k,k)} \in \mathbb{R}^{N_1N_2 \times q^2N_1N_2}$ is the subsampling matrix. Each row of $\mathbf{A}^{(k,k)}$ maps a square block of $q \times q$ high-resolution samples into a single low-resolution pixel.

2.3. Motion Compensated Subsampling Model

The idea is to extract knowledge about the high-resolution frame from the low-resolution frames. An exact model is given as $\mathbf{y}^{(l)} = \mathbf{A}^{(l,k)}\mathbf{z}^{(k)} + \mathbf{u}^{(l,k)}$, $\forall\, l \neq k$. The motion compensated subsampling matrix models the subsampling of the high-resolution frame and accounts for object motion occurring between frames $\mathbf{y}^{(l)}$ and $\mathbf{y}^{(k)}$. For pixels in $\mathbf{z}^{(k)}$ which are not observable in $\mathbf{y}^{(l)}$, $\mathbf{A}^{(l,k)}$ contains a column of zeros. Object motion will also cause pixels to be present in $\mathbf{y}^{(l)}$ which are not in $\mathbf{z}^{(k)}$. The vector $\mathbf{u}^{(l,k)}$ accommodates for these pixels with nonzero elements. Since $\mathbf{u}^{(l,k)}$ is unknown, it is obviously difficult to utilize these nonzero rows. Rows of $\mathbf{A}^{(l,k)}$ containing useful information are those for which elements of $\mathbf{y}^{(l)}$ are observed entirely from motion compensated elements of $\mathbf{z}^{(k)}$. Write these useful rows as the reduced set of equations $\mathbf{y}'^{(l)} = \mathbf{A}'^{(l,k)}\mathbf{z}^{(k)}$. In practice, the motion compensated subsampling matrix must be estimated initially from the low-resolution frames; *i.e.*, an estimate $\hat{\mathbf{A}}'^{(l,k)}$ must be computed from $\mathbf{y}^{(l)}$ and $\mathbf{y}^{(k)}$. In the construction of $\hat{\mathbf{A}}'^{(l,k)}$, estimates of the subpixel motion vectors between frames $\mathbf{y}^{(l)}$ and $\mathbf{y}^{(k)}$ are required, as well as estimates of pixel locations which are not observable within both frames simultaneously [9]. The relationship between $\mathbf{y}^{(l)}$ and $\mathbf{z}^{(k)}$ for $l \neq k$ is defined as

$$\mathbf{y}'^{(l)} = \hat{\mathbf{A}}'^{(l,k)}\mathbf{z}^{(k)} + \mathbf{n}^{(l,k)}, \quad (4)$$

where $\mathbf{n}^{(l,k)}$ is an additive noise term representing the error in estimating $\hat{\mathbf{A}}'^{(l,k)}$. This noise is assumed to be independent and identically distributed (i.i.d.) Gaussian.

3. VIDEO FRAME ENHANCEMENT

The problem of estimating the high-resolution frame $\hat{\mathbf{z}}^{(k)}$ given the low-resolution subsequence $\left\{\mathbf{y}^{(l)}\right\}$ is ill-posed in the sense of Hadamard, since a number of solutions could satisfy the model constraints. A well-posed problem will be formulated using Bayesian maximum *a posteriori* (MAP) estimation, resulting in a constrained optimization problem with a unique minimum.

The MAP estimate is located at the maximum of the posterior probability $\Pr\left(\mathbf{z}^{(k)} \middle| \left\{\mathbf{y}^{(l)}\right\}\right)$. Equivalently, the estimate is computed as

$$\hat{\mathbf{z}}^{(k)} = \arg\max_{\mathbf{z}^{(k)}} \left\{\log \Pr\left(\mathbf{z}^{(k)}\right) + \log \Pr\left(\left\{\mathbf{y}^{(l)}\right\} \middle| \mathbf{z}^{(k)}\right)\right\} \quad (5)$$

through use of the log-likelihood function. Both the prior image model and the conditional density will be defined.

Bayesian estimation distinguishes between possible solutions through a prior image model. Commonly, an assumption of global smoothness is made for the data, which is incorporated into the estimation problem by a Gaussian prior. The Huber-Markov random field (HMRF) model [3] is a Gibbs prior which models piece-wise smooth data, given as

$$\Pr\left(\mathbf{z}^{(k)}\right) = \frac{1}{Z}\exp\left\{-\frac{1}{2\beta}\sum_{c \in \mathcal{C}} \rho_\alpha\left(\mathbf{d}_c^t\mathbf{z}^{(k)}\right)\right\}. \quad (6)$$

In this expression, Z is a normalizing constant known as the partition function, β is the "temperature" parameter, and c is a local group of pixels contained within the set of all image cliques $\mathcal{C}$. The quantity $\mathbf{d}_c^t\mathbf{z}^{(k)}$ is a spatial activity measure, with a small value in smooth image regions and a large value at edges. Four spatial activity measures are computed at each pixel in the high-resolution image, implemented as second-order finite differences [3]. The likelihood of edges is controlled by the Huber edge penalty function [3],

$$\rho_\alpha(x) = \begin{cases} x^2, & |x| \leq \alpha, \\ 2\alpha|x| - \alpha^2, & |x| > \alpha, \end{cases} \quad (7)$$

where α is a threshold parameter controlling the size of discontinuities modeled by the prior.

The conditional density models the error in estimating $\hat{\mathbf{A}}^{(l,k)}$. Error is independent between frames, so that the complete density may be written as

$$\Pr\left(\left\{\mathbf{y}^{(l)}\right\} \middle| \mathbf{z}^{(k)}\right) = \prod_{l=k-\frac{M-1}{2}}^{k+\frac{M-1}{2}} \Pr\left(\mathbf{y}^{(l)} \middle| \mathbf{z}^{(k)}\right). \quad (8)$$

Since $\mathbf{A}^{(k,k)}$ is known exactly,

$$\Pr\left(\mathbf{y}^{(k)} \middle| \mathbf{z}^{(k)}\right) = \begin{cases} 1, & \text{for } \mathbf{y}^{(k)} = \mathbf{A}^{(k,k)}\mathbf{z}^{(k)}, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

All other conditional densities are given by the zero-mean i.i.d. Gaussian probability density

$$\Pr\left(\mathbf{y}^{(l)} \middle| \mathbf{z}^{(k)}\right) = \quad (10)$$

$$\frac{1}{(2\pi)^{\frac{N_1N_2}{2}}\sigma^{(l,k)N_1N_2}} \exp\left\{-\frac{1}{2\sigma^{(l,k)2}}\left\|\mathbf{y}'^{(l)} - \hat{\mathbf{A}}'^{(l,k)}\mathbf{z}^{(k)}\right\|^2\right\}$$

$\forall\, l \neq k$. Although the error variance $\sigma^{(l,k)^2}$ for each frame is unknown, it is assumed to be proportional to the frame index difference $|l-k|$.

The MAP estimate of the high-resolution data becomes

$$\hat{\mathbf{z}}^{(k)} = \arg \min_{\mathbf{z}^{(k)} \in \mathcal{Z}} \left\{ \sum_{c \in C} \rho_\alpha \left(\mathbf{d}_c^t \mathbf{z}^{(k)} \right) + \sum_{\substack{l = k - \frac{M-1}{2} \\ l \neq k}}^{k + \frac{M-1}{2}} \lambda^{(l,k)} \left\| \mathbf{y}^{\prime(l)} - \hat{\mathbf{A}}^{\prime(l,k)} \mathbf{z}^{(k)} \right\|^2 \right\}, \quad (11)$$

in which the solution is constrained to the set

$$\mathcal{Z} = \left\{ \mathbf{z}^{(k)} \; : \; \mathbf{y}^{(k)} = \mathbf{A}^{(k,k)} \mathbf{z}^{(k)} \right\}. \quad (12)$$

Each frame has an associated parameter, $\lambda^{(l,k)} = \beta / \sigma^{(l,k)^2}$, representing the confidence in $\hat{\mathbf{A}}^{\prime(l,k)}$. Since the objective function is convex, a unique solution to the optimization problem exists. The gradient projection algorithm [9] is used to compute $\hat{\mathbf{z}}^{(k)}$, with a zero-order hold of the center frame, $\mathbf{z}_0^{(k)} = q^2 \mathbf{A}^{(k,k)^t} \mathbf{y}^{(k)}$, used as the initial condition.

4. SIMULATIONS

The *Airport* test sequence consists of seven high-resolution frames. Diagonal panning of an airport scene was simulated by extracting subimages from a digitized image, shifting each successive frame seven vertical and seven horizontal pixels. Each low-resolution frame $\mathbf{y}^{(l)}$ was generated by averaging 4×4 pixel blocks within each high-resolution frame $\mathbf{z}^{(l)}$ and then subsampling by a factor of $q = 4$. The center low-resolution frame $\mathbf{y}^{(k)}$ was expanded using various single and multiframe techniques. For the video frame enhancement algorithm, subpixel motion vectors were estimated using a modified version of hierarchical block matching [10]. Table I provides a quantitative comparison of the estimates by showing the improved signal-to-noise ratio,

$$\Delta_{SNR} = 10 \log_{10} \|\mathbf{z}^{(k)} - \mathbf{z}_0^{(k)}\|^2 / \|\mathbf{z}^{(k)} - \hat{\mathbf{z}}^{(k)}\|^2 \; dB. \quad (13)$$

In the Bayesian methods, linear estimates ($\alpha = \infty$) and nonlinear estimates ($\alpha = 1$) were both computed to show the improvement gained by the preservation of edges. In this particular example it is known that only panning occurs, so that the exact displacement vectors can be recovered by averaging the estimated motion vector fields. This results in a significant resolution improvement. Figure 1 depicts several enhanced *Airport* estimates.

The *Mobile Calendar* test sequence consists of seven frames, composed of objects possessing fine detail. Within the sequence, a wall calendar moves with subpixel translational motion, and a toy train engine moving with translational motion pushes a ball undergoing rotational motion. Each high-resolution frame was subsampled by a factor of $q = 4$ in the same manner as described for the *Airport* sequence. Table II shows quantitative results for various *Mobile Calendar* frame interpolations, and Figure 2 shows details of the estimated frames in a region of the wall calendar. Again, the video frame enhancement algorithm with the Huber-Markov image model provides the best result, although the resolution improvement is not as dramatic as in the previous sequence.

5. CONCLUSION

An observation model was proposed for low-resolution video frames, which models the subsampling of the unknown high-resolution data and accounts for general object motion occurring between frames. Simulation results from the video frame enhancement algorithm were reported for sequences containing global and general motion. In the case of camera panning, definition was significantly improved. More modest improvements were visible for the sequence containing objects moving with independent trajectories. In future research, a robust regularization technique will be applied to the ill-posed inverse problem of motion estimation.

6. REFERENCES

[1] H. H. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. ASSP*, vol. 26, no. 6, pp. 508–517, 1978.

[2] N. B. Karayiannis and A. N. Venetsanopoulos, "Image interpolation based on variational principles," *Signal Processing*, vol. 25, no. 3, pp. 259–288, 1991.

[3] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Processing*, vol. 3, no. 3, pp. 233–242, 1994.

[4] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing* (R. Y. Tsai and T. S. Huang, eds.), vol. 1, pp. 317–339, JAI Press Inc., 1984.

[5] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. ASSP*, vol. 38, no. 6, pp. 1013–1027, 1990.

[6] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. Am. A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[7] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in *Proc. ICASSP*, (San Francisco, CA), pp. III–169 to III–172, 1992.

[8] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur," in *Proc. ICIP*, (Austin, TX), pp. I–343 to I–347, 1994.

[9] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences." Submitted to *IEEE Trans. Image Processing*.

[10] M. Bierling and R. Thoma, "Motion compensating field interpolation using a hierarchically structured displacement estimator," *Signal Processing*, vol. 11, no. 4, pp. 387–404, 1986.

TABLE I
COMPARISON OF INTERPOLATION METHODS ON THE SYNTHETIC *Airport* SEQUENCE

Interpolation Technique	Δ_{SNR} (dB)
Bilinear, $M = 1$	0.57
Cubic B-Spline, $M = 1$	1.25
MAP Estimation, $M = 1$, $\alpha = \infty$	1.43
MAP Estimation, $M = 1$, $\alpha = 1$	1.51
Video Frame Enhancement with Motion Estimates, $M = 7$, $\alpha = \infty$, $\lambda^{(l,k)} = \frac{10}{\|l-k\|}$	3.27
Video Frame Enhancement with Motion Estimates, $M = 7$, $\alpha = 1$, $\lambda^{(l,k)} = \frac{10}{\|l-k\|}$	5.16
Video Frame Enhancement with Panning, $M = 7$, $\alpha = \infty$, $\lambda^{(l,k)} = \frac{1000}{\|l-k\|}$	6.67
Video Frame Enhancement with Panning, $M = 7$, $\alpha = 1$, $\lambda^{(l,k)} = \frac{1000}{\|l-k\|}$	6.96

TABLE II
COMPARISON OF INTERPOLATION METHODS ON THE *Mobile Calendar* SEQUENCE

Interpolation Technique	Δ_{SNR} (dB)
Bilinear, $M = 1$	0.24
Cubic B-Spline, $M = 1$	0.72
MAP Estimation, $M = 1$, $\alpha = \infty$	0.82
MAP Estimation, $M = 1$, $\alpha = 1$	1.05
Video Frame Enhancement with Motion Estimates, $M = 7$, $\alpha = \infty$, $\lambda^{(l,k)} = \frac{10}{\|l-k\|}$	1.37
Video Frame Enhancement with Motion Estimates, $M = 7$, $\alpha = 1$, $\lambda^{(l,k)} = \frac{10}{\|l-k\|}$	2.11

Figure 1: Synthetic *Airport* sequence. *Left-to-Right*: High-resolution frame $\mathbf{z}^{(k)}$; One of the low-resolution frames $\mathbf{y}^{(k)}$; MAP estimator from [3], $M = 1$, $\alpha = 1$; Video frame enhancement with panning, $M = 7$, $\alpha = 1$, $\lambda^{(l,k)} = \frac{1000}{\|l-k\|}$.

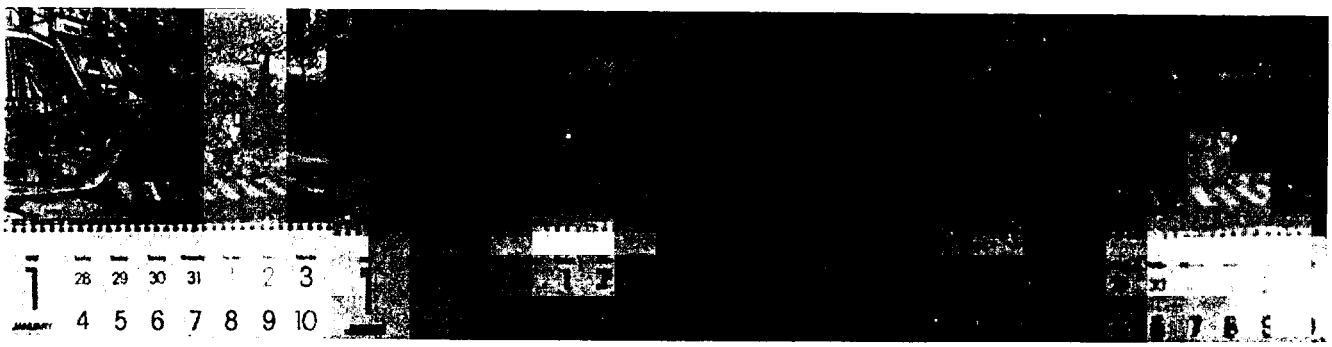

Figure 2: Details of the *Mobile Calendar* sequence. *Left-to-Right*: High-resolution frame $\mathbf{z}^{(k)}$; One of the low-resolution frames $\mathbf{y}^{(k)}$; MAP estimator from [3], $M = 1$, $\alpha = 1$; Video frame enhancement with motion estimates, $M = 7$, $\alpha = 1$, $\lambda^{(l,k)} = \frac{10}{\|l-k\|}$.