MAP Based Resolution Enhancement of Video Sequences Using a Huber-Markov Random Field Image Prior Model

Hu He, Lisimachos P. Kondi

Dept. of Electrical Engineering, SUNY at Buffalo, Buffalo, NY 14260

ABSTRACT

In this paper, we propose two approaches for video sequence resolution enhancement using Maximum A Posteriori (MAP) estimation. Huber-Markov Random Fields (HMRF) are used as prior models. These models can better preserve image discontinuities (edges) when compared with Gaussian prior models. The two proposed approaches differ in the selection of image smoothness measure. The first approach employs a measure that is based on a discrete Laplacian kernel, while the second approach uses a finite difference approximation of second order derivatives at each pixel of the high-resolution image estimate. Experimental results are presented and conclusions are drawn.

1. INTRODUCTION

In many imaging systems, the resolution of the detector array of the camera is not sufficiently high for a particular application. Furthermore, the capturing process introduces additive noise and the point spread function of the lens and the effects of the finite size of the photo-detectors further degrade the acquired video frames. The goal of resolution enhancement is to estimate a high-resolution image from a sequence of low-resolution images while also compensating for the above-mentioned degradations.

In this paper, we propose two approaches for video sequence resolution enhancement using a joint Maximum A Posteriori (MAP) registration and high-resolution image estimation algorithm. The motion between the frames (registration) and the high-resolution image are jointly estimated. A Huber-Markov Random Field (HMRF) is used as prior model for the high-resolution image. The use of HMRF model better preserves image discontinuities when compared with a Gaussian prior model.

In this paper, we extend our previous work in [1] and [2] where a Gaussian prior image model was used. In [3], MAP estimation of a high-resolution image from a single low-resolution image using a HMRF prior model was presented. A finite difference approximation of second order derivatives was employed as a smoothness measure. The algorithm in [3] was extended in [4] to perform high-resolution image estimation from a sequence of low-resolution images. The relative motion between the frames (registration) was performed prior to the MAP estimation of the high-resolution image. In this paper, we propose a joint MAP registration (estimation of the relative motion between frames) and high-resolution image estimation algorithm using an HMRF prior model. We propose a smoothness measure based on a Laplacian kernel and compare its performance with the smoothness measure in [3] and [4].

The rest of the paper is organized as follows. In section 2, the MAP-based resolution enhancement algorithm is presented along with two different HMRF-based prior models. In section 3, experimental results are presented. Finally, in section 4, conclusions are drawn.

2. MAP-BASED RESOLUTION ENHANCEMENT

Several resolution enhancement techniques have been proposed in the literature. In this review, we concentrate on Bayesian methods. Maximum A Posteriori (MAP) estimation with an edge preserving Huber-Markov random field image prior is studied in [3], [4]. MAP based resolution enhancement with simultaneously estimation of registration parameters has been proposed [1], [2], [5]. In the following, we use the same model and notation as in [1].

2.1 Observation Model

We order all vectors lexicographically. We assume that p low-resolution frames are observed, each of size $N_1 \times N_2$. The desired high-resolution image $\mathbf{z}$ is of size $N = L_1 N_1 L_2 N_2$ and L_1 and L_2 represent the down-sampling factors in the horizontal and vertical directions, respectively. Thus, the observed low-resolution images are related to the high resolution image through blurring, motion shift and subsampling. Let the kth low-resolution frame be denoted as $\mathbf{y}_k = [y_{k,1}, y_{k,2}, \cdots y_{k,M}]^t$ for $k = 1,2,\cdots p$ and where $M = N_1 N_2$. The full set of p observed low-resolution images can be denoted as

$$\mathbf{y} = \left[\mathbf{y}_1^t, \mathbf{y}_2^t, \cdots, \mathbf{y}_p^t\right]^t = \left[y_1, y_2, \cdots, y_{pM}\right]^t \quad (1)$$

The observed low resolution frames are related to the high-resolution image through the following model:

$$y_{k,m} = \sum_{r=1}^{N} w_{k,m,r}(\mathbf{s}_k) z_r + \eta_{k,m} \quad (2)$$

for $m = 1,2,\cdots M$ and $k = 1,2,\cdots p$. The weight $w_{k,m,r}(\mathbf{s}_k)$ represents the "contribution" of the *r*th high-resolution pixel to the *m*th low resolution observed pixel of the *k*th frame. The vector $\mathbf{s}_k = \left[s_{k,1}, s_{k,2}, \cdots s_{k,K}\right]^t$, is the *K* registration parameters for frame *k*, representing global translational shift, rotation, affine transformation parameters, or other motion parameters. This motion is measured in reference to a fixed high resolution grid. The term $\eta_{k,m}$ represents additive noise samples that will be assumed to be independent and identically distributed (i.i.d.) Gaussian noise samples with variance σ_η^2.

It will be convenient to represent the observation model in matrix notation.

$$\mathbf{y} = \mathbf{W_s z} + \mathbf{n} \quad (3)$$

where $\mathbf{W_s}$ contains the values $w_{k,m,r}(\mathbf{s}_k)$ and $\mathbf{n} = \left[\eta_1, \eta_2, \cdots \eta_{pM}\right]^t$. Note that since the elements of $\mathbf{n}$ are i.i.d. Gaussian samples, the multivariate pdf of $\mathbf{n}$ is given by

$$P_r(\mathbf{n}) = \frac{1}{(2\pi)^{\frac{pM}{2}} \sigma_\eta^{pM}} \exp\left\{-\frac{1}{2\sigma_\eta^2} \sum_{m=1}^{pM} \eta_m^2\right\}. \quad (4)$$

We can form a MAP estimate of the high-resolution image $\mathbf{z}$ and the registration parameters $\mathbf{s}$ simultaneously, given the observed low resolution images $\mathbf{y}$. The estimates can be computed as

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z,s}} P_r(\mathbf{z}, \mathbf{s} | \mathbf{y}) \quad (5)$$

Using Bayes rule, assuming that all possible motion vectors are equally probable and after some algebra, the joint estimates can be expressed as:

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\min_{\mathbf{z,s}} \left\{-\log\left[P_r(\mathbf{y}|\mathbf{z},\mathbf{s})\right] - \log\left[P_r(\mathbf{z})\right]\right\}. \quad (6)$$

2.2 Cost Function Using a Gaussian Image Prior

The prior image can be chosen to be a Gaussian random field with density of the form:

$$\Pr(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\mathbf{z}^t \mathbf{C}^{-1} \mathbf{z}\right\} \quad (7)$$

where $\mathbf{C}$ is the $N \times N$ covariance matrix of $\mathbf{z}$. More information can be found in [1], [2]. The assumption of a Gaussian prior model tends to produce a high-resolution estimate with smooth edges. Huber-Markov models address this problem, as explained in the next section.

2.3 Approach I: Cost Function with HMRF and Laplacian Smoothness Kernel

By making the assumption of a Gaussian prior model, edges are statistically unlikely to appear in the MAP estimate. Effectively, high-frequency components are suppressed by the image model, since it assumes that smooth edges will be more highly probable than sharp discontinuities. A more realistic assumption is that the image data are piece-wise smooth; i.e., the image consists of smooth regions which are separated by discontinuities. A general form of the HMRF density is

$$\Pr(\mathbf{z}) = A \exp\left\{-\frac{1}{\lambda} \sum_{c \in C} \rho\left(\mathbf{d}_c^t \mathbf{z}\right)\right\} \quad (8)$$

where *A* is a constant, λ is the temperate or tuning parameter of the density, *c* is a local group of points called clique and *C* denotes the set of all cliques in the image. $\mathbf{d}_c$ is a coefficient vector for clique *c* and $\rho(\bullet)$ is the Huber function:

$$\rho(x) = \begin{cases} x^2, & |x| \le T \\ T^2 + 2T(|x| - T), & |x| > T \end{cases} \quad (9)$$

where *T* is the threshold of Huber function.

In this paper, we propose to measure the image smoothness at a given pixel using a Laplacian kernel, i.e., $\mathbf{d}_i^t \mathbf{z} = \sum_{j=1}^{N} d_{i,j} z_j$ with $d_{i,j}$ a Laplacian smoothness kernel defined as

$$d_{i,j} = \begin{cases} 1 & for\ i = j \\ -1/4 & for\ j : z_j\ is\ a\ cardinal\ neighbor\ of\ z_j \end{cases} \quad (10)$$

The corresponding cost function is:

$$L(\mathbf{z}, \mathbf{s}) = \frac{1}{2\sigma_\eta^2} \sum_{m=1}^{pM} \left(y_m - \sum_{r=1}^{N} w_{m,r}(\mathbf{s}) z_r\right)^2 + \frac{1}{2\lambda} \sum_{N_T} \left(\sum_{j=1}^{N} d_{i,j} z_j\right)^2 + \frac{1}{2\lambda} \sum_{N_T^C} \left\{T^2 + 2T\left(\left|\mathbf{d}_i^t \mathbf{z}\right| - T\right)\right\} \quad (11)$$

where N_T is the set that $\left\{i \mid i \in (1,2,\cdots,N), \left|d_i^t z\right| \le T\right\}$ and N_T^C is the complement set of N_T.

2.4 Approach II: Cost Function with HMRF Image Prior and Finite Difference Approximation to Second Order Derivatives as Smoothness Measure

We next present the smoothness measure in [3], [4]. The Huber-Markov Random field model and Huber function are same as those in equation (8), (9). The difference is that image smoothness is measured by finite difference approximation to second order derivatives, i.e., $\mathbf{d}_i^t\mathbf{z}$ is defined differently.

The quantity $\mathbf{d}_i^t\mathbf{z} = \sum_{q=1}^{4} \rho(\mathbf{d}_{i,q}^t\mathbf{z})$ is the spatial activity measure that can be computed at each pixel (x, y) in the high resolution image, given by the following second order finite differences:

$$\begin{aligned}
\mathbf{d}_{x,y,1}^t\mathbf{z} &= z_{x,y-1} - 2z_{x,y} + z_{x,y+1} \\
\mathbf{d}_{x,y,2}^t\mathbf{z} &= \left(z_{x+1,y-1} - 2z_{x,y} + z_{x-1,y+1}\right)/2 \\
\mathbf{d}_{x,y,3}^t\mathbf{z} &= z_{x-1,y} - 2z_{x,y} + z_{x+1,y} \\
\mathbf{d}_{x,y,4}^t\mathbf{z} &= \left(z_{x-1,y-1} - 2z_{x,y} + z_{x+1,y+1}\right)/2
\end{aligned} \tag{12}$$

The smoothness measure in this approach is contributed from four directions of the current high-resolution pixel: horizontal, vertical and two diagonal directions, while in approach I, only an overall smooth quantity at the high-resolution pixel is used and it is directionless.

The corresponding cost function of approach II is:

$$L(\mathbf{z},\mathbf{s}) = \frac{1}{2\sigma_\eta^2}\sum_{m=1}^{pM}\left(y_m - \sum_{r=1}^{N} w_{m,r}(\mathbf{s})z_r\right)^2 + \frac{1}{2\lambda}\sum_{i=1}^{N}\sum_{q=1}^{4}\rho\left(\mathbf{d}_i^t\mathbf{z}\right) \tag{13}$$

2.5 Joint MAP Registration Algorithm

The above cost functions can be minimized using the coordinate-descent method. This iterative method starts with an initial estimate of **z** obtained using interpolation from a low resolution frame. Then, for a fixed **z**, the cost function is minimized with respect to **s**. Thus, the motion of each frame is estimated. Then, for fixed **z**, a new estimate for **z** is obtained. This procedure continues until convergence is reached, i.e., **z** and **s** updated in a cyclic fashion. In order to update the estimate **z**, we first estimate

$$\hat{\mathbf{s}}_k^n = \arg\min_{\mathbf{s}_k}\left\{\sum_{m=1}^{M}\left(y_m - \sum_{r=1}^{N} w_{m,r}(\mathbf{s})z_r\right)^2\right\} \tag{14}$$

Also, the gradient can be obtained from

$$g_k(\mathbf{z},\mathbf{s}) = \frac{\partial L(\mathbf{z},\mathbf{s})}{\partial z_k} \tag{15}$$

The step size ε^n can be found by solving

$$\frac{\partial L(\hat{\mathbf{z}}^{n+1},\hat{\mathbf{s}}^n)}{\partial \varepsilon^n} = 0 \tag{16}$$

Then **z** can be updated recursively as

$$\hat{z}_k^{n+1} = \hat{z}_k^n - \varepsilon^n g_k(\hat{\mathbf{z}}^n,\hat{\mathbf{s}}^n) \tag{17}$$

until convergence is reached. n is the iteration number starting from 0.

3. EXPERIMENTAL RESULTS

We use real data of a truck video sequence provided by the Naval Research Laboratory (Washington, DC) to test the reconstruction results. 20 frames of low-resolution image with size 128x128 pixels are used. The up-sample ratio was L_1= L_2=4. The point spread function is pre-determined before the resolution enhancement. For our case, we used a Gaussian blur with variance 1.7. The variance of noise is estimated at a "smooth" area of the low-resolution. The first frame is selected as reference frame and bilinear interpolation of the first frame is chosen as the first estimate of high-resolution image **z**. To estimate the motion, the current **z** is compared with the low-resolution frames as shown in equation (14). Three Point Search (TSS) with search window ± 7 and Sum of Squared Errors (SSE) criterion are applied to decide the motion vector of current macro block. When reconstructing the high resolution image, these motion vectors are used to compensate the motion. The threshold of Huber function is set at T=1 for both approaches. The coordinate-descent method is carried out for 20 iterations or until convergence is reached $\left(\left\|\hat{\mathbf{z}}^{n+1} - \hat{\mathbf{z}}^n\right\|/\left\|\hat{\mathbf{z}}^n\right\| < 10^{-6}\right)$.

The first frame of the low video sequence is shown in Fig.1. In Fig. 2, bilinear interpolation of the first frame is shown as a comparison to the reconstructed high-resolution images using Joint MAP algorithm with HMRF image prior shown in Fig 3 and 4 respectively. In Fig.5, the difference between the two approaches is displayed.

4. CONCLUSION

Both approaches give a good reconstruction result and preserve more information at the edges than that using Gaussian prior model in [2]. The smoothness measure in approach II is contributed from four directions of the current high-resolution pixel while in approach I, only an overall smoothness quantity is used and it is directionless. This gives approach I (which proposed in this paper) less computation cost compared to approach II (from [3], [4]) with almost the same visual result. We should notice that the result is highly dependent on the choice of threshold T and there is no explicit rule of selection for T. Also, the high-frequency term of the cost function is weighted by the inverse of the tuning parameter, whose selection is empirical for most cases. As λ increases, the MAP estimate approaches the Maximum Likelihood (ML) estimate and the choice of T has less influence on the superresolution result. More research needs to be done on this topic.

5. REFERENCES

[1] L. P. Kondi, D. Scribner, J. Schuler, "A Comparison of digital image resolution enhancement techniques", SPIE AeroSense, Orlando, FL, April 2002.

[2] H. He, L. P. Kondi, "Resolution enhancement of video sequences with simultaneous estimation of the regularization parameter," SPIE Electronic Imaging, Santa Clara, CA, January 2003.

[3] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," IEEE Trans. Image Processing, vol. 3, pp. 233–242, May 1994.

[4] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," IEEE Trans. Image Processing, Vol. 5, pp. 996-1011, June 1996.

[5] R. C. Hardie, K. J. Barnard "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," IEEE Trans. Image Processing, vol. 6, pp. 621–1633, December 1997.

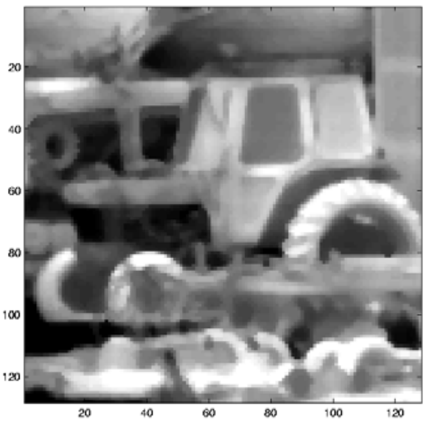

Fig. 1 Original image (first frame of truck sequence)

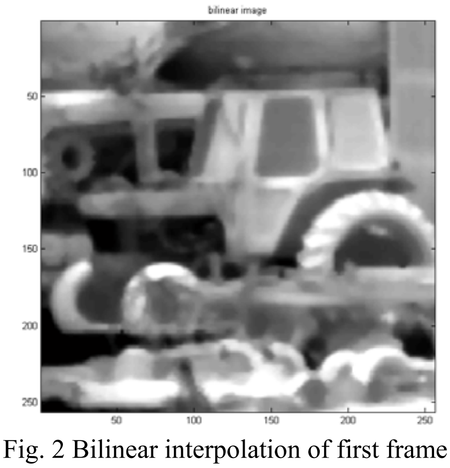

Fig. 2 Bilinear interpolation of first frame

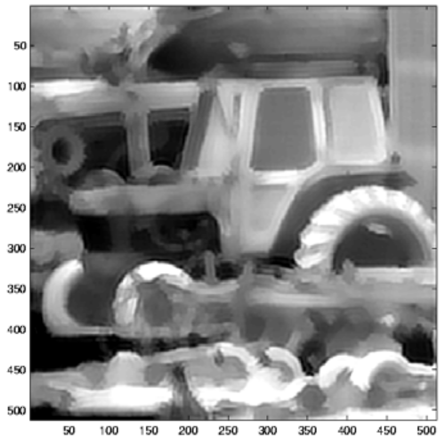

Fig.3 Reconstructed high-resolution image using joint MAP with HMRF, approach I

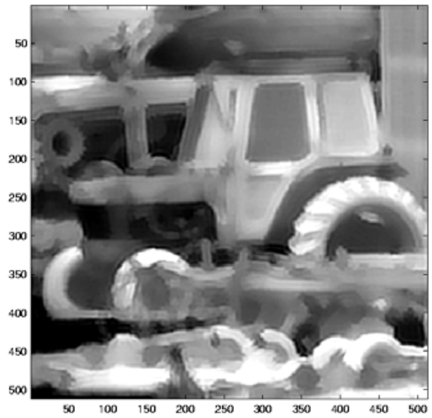

Fig. 4 Reconstructed high-resolution image using joint MAP with HMRF, approach II

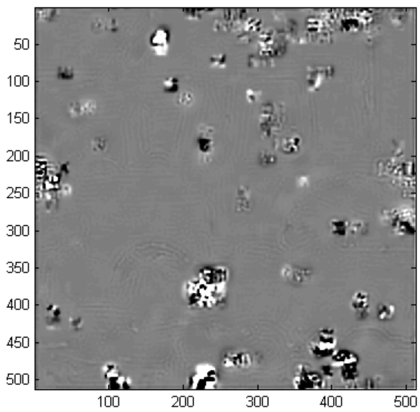

Fig. 5 Difference between the reconstructed high-resolution Image using approach I and II (peak value=2.5)