EM-Based Simultaneous Registration, Restoration, and Interpolation of Super-resolved Images

Nathan A. Woods[1], Nikolas P. Galatsanos[1], and Aggelos K. Katsaggelos[2]

Illinois Institute of Technology[1]
Department of Electrical and Computer Engineering
3301 S. Dearborn St., Chicago, IL 60616, USA,

Northwestern University[2]
Department of Electrical and Computer Engineering
2145 Sheridan Rd., Evanston, IL 60208, USA

Nathan@BinaryMachinesInc.com, npg@ece.iit.edu, aggk@ece.nwu.edu

ABSTRACT

We present a maximum likelihood (ML) solution to the problem of obtaining high-resolution images from sequences of noisy, blurred, and low-resolution images. In our formulation, the registration parameters of the low-resolution images, the degrading blur, and noise variance are unknown. Our algorithm has the advantage that all unknown parameters are obtained simultaneously using all of the available data. An efficient implementation is presented in the frequency domain, based on the Expectation Maximization (EM) algorithm. Simulations demonstrate the effectiveness of the algorithm.

1. INTRODUCTION

The problem of obtaining an enhanced resolution image from a sequence of lower resolution images has been an active area of research over the last several years. Tsai and Huang [1] were the first to demonstrate that unique information in a sequence of translated and aliased images can be exploited to produce an enhanced resolution image. In their early work, the low-resolution images were neither blurred nor noisy, and they assumed that the shifts between the low-resolution still were known. Since then, several other approaches have been proposed in the literature, see for example [2] and [3].

In this paper, we attempt to solve the super-resolution problem where the blur, noise, and motion parameters are unknown. In our solution, registration of the low-resolution images, blind deconvolution, and interpolation are all performed jointly using all of the available data. The presented solution is optimal in the maximum-likelihood (ML) sense. The solution is obtained iteratively using the Expectation Maximization algorithm (EM) and is solved in the frequency domain.

The main advantages of the proposed formulation are: 1) The frequency domain implementation of the entire algorithm allows the efficient use, without any truncation, of the very large convolutional operators required in the registration step. 2) The EM algorithm apart from its guaranteed convergence properties captures naturally implicit constraints for ML estimation problems, such as the positive definite nature of covariance matrices. Furthermore, it provides fertile ground for Bayesian extensions with conjugate priors.

2. IMAGE MODEL

We first approximate the underlying, or actual, scene as a discrete, high-resolution image of size M_H x N_H. Let x be an $M_H N_H$ x 1 vector containing the discrete intensity values, of the underlying scene. We model the intensity values as samples from a random process with probability density

$$p(x) \propto \alpha^{(N_H M_H - 1)} \exp\left(-\frac{\alpha^2}{2} x^T Q^T Q x\right) \qquad (1)$$

with Q the Laplacian operator matrix and α^2 an unknown parameter. This image model is termed simultaneously autoregressive [7] and [8]. The observations, y_i, are related to the underlying image x by

$$y_i = B_i x + n_i \quad \text{for } i = 0,1,\ldots,P-1 \qquad (2)$$

where y_i, $i = 0, 1, \ldots, P-1$, is an $M_L N_L$ x 1 vector containing the discrete intensity values of the i^{th} observed image arranged lexicographically, B_i is a linear degradation operator of size $M_L N_L$ x $M_H N_H$, and n_i is an $M_L N_L$ x 1 noise vector. The noise is modeled as a Gaussian random process with zero mean,

$$n \sim N(0, \Lambda_n). \qquad (3)$$

In this problem white noise was used with $\Lambda_n = \sigma^2 I$ where σ^2 is the unknown noise variance. We also assume that the noise is uncorrelated with the underlying scene.

The linear degradation operator, B_i, spatially shifts, blurs, and decimates x. Therefore, we decompose B_i as follows,

$$B_i = D \cdot S(\delta_i) \cdot H \qquad (4)$$

where D is the decimation operator of size $M_L N_L$ x $M_H N_H$ common to all observed images, H is the blur operator of size $M_H N_H$ x $M_H N_H$ common to all images,

0-7803-7750-8/03/$17.00 ©2003 IEEE

and S is the shift operator of size $M_H N_H$ x $M_H N_H$ parameterized by the i^{th} shift vector $\boldsymbol{\delta}_i$. The shift vector $\boldsymbol{\delta}_i$ is a 1 x 2 vector containing the horizontal and vertical shifts, measured in pixels of $\mathbf{x}$, of the i^{th} image relative to the 0^{th} image.

We restrict our model to integer vertical and horizontal decimation factors, d_y and d_x, so that

$$\frac{M_H}{M_L} = d_y \quad \text{and} \quad \frac{N_H}{N_L} = d_x \quad \{d_x, d_y : d_x \in \mathbb{Z}, d_y \in \mathbb{Z}\} \tag{5}$$

The shift operator, $S(\boldsymbol{\delta}_i)$, is the Shannon 2-D interpolation operator which is shift invariant and is given by

$$h_{shift}(m,n,\boldsymbol{\delta}) = \frac{\sin(\pi(m-\delta_y))}{\pi(m-\delta_y)} \frac{\sin(\pi(n-\delta_x))}{\pi(n-\delta_x)} \tag{6}$$

for $m = 0,1,\ldots,M_H - 1$, and $n = 0,1,\ldots N_H - 1$. To reduce the number of unknown parameters, we model the blur operator H as linear convolution with the following 2-D impulse response, parameterized by β^2:

$$h_{blur}(m,n) = \frac{1}{2\pi\beta^2} \exp\left(-\frac{1}{2\beta^2}\left(n^2 + m^2\right)\right) \tag{7}$$

Let us represent the entire sequence of observed images as a single $PM_L N_L$ x 1 vector,

$$\mathbf{y} = \begin{bmatrix} y_0^T & y_1^T & \cdots & y_{P-1}^T \end{bmatrix}^T \tag{8}$$

so that the model becomes

$$\mathbf{y} = \mathbf{DSHx} + \mathbf{n} \tag{9}$$

where $\mathbf{D}$ is the $PM_L N_L$ x $M_H N_H$ block diagonal decimation matrix, $\mathbf{S}$ is a $PM_H N_H$ x $M_H N_H$ matrix containing the shift operators stacked on top of one another, and $\mathbf{H}$ is an $M_H N_H$ x $M_H N_H$ blur matrix equal to H.

Equation (9) is the spatial-domain version of the model. We produce an estimate of $\mathbf{x}$ from the observation vector $\mathbf{y}$, given only the decimation factors d_x and d_y, and the matrix Q. All other model parameters are considered unknown, including the covariance parameter α, the noise variance, σ^2, the P shift vectors $\boldsymbol{\delta}_i$, and the blur matrix $\mathbf{H}$.

3. E-M MAXIMUM LIKELIHOOD SUPER RESOLUTION ALGORITHM

We use the maximum-likelihood (ML) criteria to find the best estimate of the parameters of the model and ultimately of $\mathbf{x}$. Direct maximization of the likelihood function is difficult, however, due to the high non-linearity of the likelihood function with respect to those parameters [4]. Instead, an iterative approach called the Expectation-Maximization (EM) algorithm [5] is employed to find the ML estimate of the model parameters and $\mathbf{x}$.

In the EM approach, instead of maximizing the likelihood of the observations with respect to the model parameters, one maximizes the expectation of the "complete data" conditioned on the "incomplete data". Let us denote θ as the set of unknown parameters and $f_z(z;\theta)$ as the PDF of the complete data. In mathematical form, the EM is given by the alternate computation of

$$Q\left(\theta;\theta^{(p)}\right) = E\left[\log f_z(z;\theta) \mid \mathbf{y};\theta^{(p)}\right] \tag{10}$$

and

$$\theta^{(p+1)} = \arg\left\{\max_\theta Q(\theta;\theta^{(p)})\right\} \tag{11}$$

where $\theta^{(p)}$ is the estimate of θ at the p-th iteration.

Choosing $\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ as the complete data permits the simultaneous identification of the blur and image parameters and the restoration of the image. Using this $\mathbf{z}$ as the complete data, the expected value of the negative of the conditional log likelihood, dropping the constant term, is

$$\begin{aligned} Q'(\theta;\theta^{(p)}) = {} & \log|\Lambda_x| + \log|\Lambda_n| \\ & + \operatorname{tr}\left[\left(\Lambda_x^{-1} + \mathbf{B}^H \Lambda_n^{-1}\mathbf{B}\right)\Lambda_{x|y}^{(p)}\right] \\ & + \mu_{x|y}^{H\,(p)}\left(\Lambda_x^{-1} + \mathbf{B}^H\Lambda_n^{-1}\mathbf{B}\right)\mu_{x|y}^{(p)} \\ & - 2\operatorname{Re}\left\{y^H \Lambda_n^{-1}\mathbf{B}\mu_{x|y}^{(p)}\right\} + y^H\Lambda_n^{-1}y \end{aligned} \tag{12}$$

where $\mathbf{B} = \mathbf{DSH}$ and

$$\mu_{x|y} = \Lambda_x \mathbf{B}^H \left(\mathbf{B}\Lambda_x\mathbf{B}^H + \Lambda_n\right)^{-1} \mathbf{y} \tag{13}$$

$$\Lambda_{x|y} = \Lambda_x - \Lambda_x\mathbf{B}^H\left(\mathbf{B}\Lambda_x\mathbf{B}^H + \Lambda_n\right)^{-1}\mathbf{B}\Lambda_x \tag{14}$$

Since the distribution of the complete data conditioned on the observations is Gaussian, equations (13), and (14) completely specify this distribution and therefore constitute the E-step of the EM algorithm in the spatial domain. However, computation of the E-step in the spatial domain is impractical due to the matrix inversion required in the updates of the conditional mean in (13) and conditional covariance in (14). As will be shown, by transforming the problem to the frequency domain the matrix inversion becomes manageable.

4. EFFICIENT COMPUTATIONS OF EM ITERATIONS IN FREQUENCY DOMAIN

In [6], the matrix $\mathbf{B}\Lambda_x\mathbf{B}^H + \Lambda_n$ assumed a doubly circulant form, as B represented 2-D linear convolution. It is well known that then it can be diagonalized by the similarity transform $A\left(\mathbf{B}\Lambda_x\mathbf{B}^H + \Lambda_n\right)A^{-1}$, where A is the

2-D DFT matrix. Thus, the matrix inversion of $\mathbf{B}\Lambda_x\mathbf{B}^H + \Lambda_n$ in the spatial domain was reduced to scalar division in the frequency domain.

Due to the decimation in our formulation (which was absent in [6]), the matrix $\mathbf{B}\Lambda_x\mathbf{B}^H + \Lambda_n$ is not a doubly circulant form. Nonetheless, let us define an interlaced observation vector, $\mathbf{Y}'$, in the frequency domain as follows:

$$\mathbf{Y}' = \left[Y''(0) \quad Y''(1) \quad \cdots \quad Y''(M_L N_L - 1) \right]^T \quad (15)$$

where $Y''(n) = \left[Y_0(n) \quad Y_1(n) \quad \cdots \quad Y_{P-1}(n) \right]^T$ and Y_i is the 2-D DFT of the i^{th} low-resolution image. It can be shown using similar techniques as in [6] that in the frequency domain the conditional mean and conditional covariance assume the forms

$$M_{x|y} = d_x d_y \mathbf{U}\mathbf{V}^{-1}\mathbf{Y}' \qquad (16)$$

$$\mathbf{S}_{x|y} = \mathbf{S}_x - \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^H \qquad (17)$$

where $\mathbf{V}$ is *a block diagonal matrix* composed of P x P submatrices, and $\mathbf{S}_x$ is a diagonal, frequency-domain version of the covariance matrix associated with $\mathbf{x}$ and is given by

$$\left[\mathbf{S}_x \right]_{m,m} = \frac{1}{\alpha^2} \left| \tilde{Q}(m) \right|^{-2} \qquad (18)$$

where $\tilde{Q}(m)$ are the eigenvalues of the regularization matrix Q.

Let us define the matrix $\mathbf{H}_D$ to be the (diagonal) frequency domain version of the blur matrix $\mathbf{H}$. Define the matrix $\mathbf{S}_D$ to be the frequency domain version of the shift matrix $\mathbf{S}$. $\mathbf{S}_D$ consists of P diagonal matrices stacked vertically. For notational convenience, let us define

$$u = (jM_L + k)N_H + mN_L + x \qquad (19)$$

and

$$v = (lM_L + k)N_H + mN_L + x \qquad (20)$$

After significant algebra, the frequency domain equivalent of (12) is given by

$$Q''\left(\theta; \theta^{(p)}\right) = pM_L N_L \log \sigma^2 +$$

$$\sum_{m=0}^{M_H N_H - 1} \left[\log\left(\left[\mathbf{S}_x \right]_{m,m} \right) + \frac{\left[\mathbf{S}_{x|y} \right]_{m,m} + \frac{1}{M_H N_H} \left| M_{x|y}(m) \right|^2}{\left[\mathbf{S}_x \right]_{m,m}} \right] +$$

$$\frac{1}{d_x d_y \sigma^2} \sum_{l=0}^{d_y-1} \sum_{j=0}^{d_y-1} \sum_{k=0}^{M_L-1} \sum_{n=0}^{d_x-1} \sum_{m=0}^{d_x-1} \sum_{x=0}^{N_L-1} \left\{ \left[\frac{1}{M_H N_H} M_{x|y}(u) M^*_{x|y}(v) + \right. \right.$$

$$\left. \mathbf{H}_D(u)\mathbf{H}^*_D(v) \sum_{i=0}^{P-1} \mathbf{S}_D(\delta_i, u) \mathbf{S}^*_D(\delta_i, v) \right\} -$$

$$\frac{2}{M_H N_H \sigma^2} \mathrm{Re} \left\{ \sum_{j=0}^{d_y-1} \sum_{k=0}^{M_L-1} \sum_{m=0}^{d_x-1} \sum_{x=0}^{N_L-1} M_{x|y}(u) \mathbf{H}_D(u) \sum_{i=0}^{P-1} \mathbf{S}_D(\delta_i, u) \cdot \right.$$

$$\left. Y^*_i(kN_L + x) \right\} + \frac{1}{M_L N_L \sigma^2} \sum_{m=0}^{pM_L N_L} \left| \mathbf{Y}(m) \right|^2$$

(21)

In the M-step, we must minimize the likelihood function given by (21) with respect to the unknown parameters, α^2, β^2, $\boldsymbol{\delta}$, and σ^2 this yields

$$\alpha^{2(p+1)} = \frac{M_H N_H}{\sum_{m=0}^{M_H N_H - 1} \left| \tilde{Q}(m) \right|^2 \left(\left[\mathbf{S}^{(p)}_{x|y} \right]_{m,m} + \frac{1}{M_H N_H \left| M^{(p)}_{x|y}(m) \right|^2} \right)} \qquad (22)$$

$$\sigma^{2(p+1)} = \frac{1}{pM_H N_H} \sum_{l=0}^{d_y-1} \sum_{j=0}^{d_y-1} \sum_{k=0}^{M_L-1} \sum_{n=0}^{d_x-1} \sum_{m=0}^{d_x-1} \sum_{x=0}^{N_L-1} \left\{ \left[\frac{1}{M_H N_H} M^{(p)}_{x|y}(u) \cdot \right. \right.$$

$$\left. M^{*\,(p)}_{x|y}(v) + \left[\mathbf{S}^{(p)}_{x|y} \right]_{u,v} \right] \mathbf{H}_D\left(u; \beta^{(p)}\right) \mathbf{H}^*_D\left(v; \beta^{(p)}\right) \cdot$$

$$\left. \sum_{i=0}^{P-1} \mathbf{S}_D\left(u; \delta^{(p)}_i\right) \mathbf{S}^*_D\left(v; \delta^{(p)}_i\right) \right\} - \frac{2 d_x d_y}{P (M_H N_H)^2 \sigma^2} \cdot$$

$$\mathrm{Re} \left\{ \sum_{j=0}^{d_y-1} \sum_{k=0}^{M_L-1} \sum_{m=0}^{d_x-1} \sum_{x=0}^{N_L-1} M^{(p)}_{x|y}(u) \mathbf{H}_D\left(u; \beta^{(p)}\right) \cdot \right.$$

$$\left. \sum_{i=0}^{P-1} \mathbf{S}_D\left(u; \delta^{(p)}_i\right) Y^*_i(kN_L + x) \right\}$$

$$+ \frac{1}{P} \left(\frac{d_x d_y}{M_H N_H} \right)^2 \sum_{m=0}^{pM_L N_L} \left| \mathbf{Y}(m) \right|^2 \qquad (23)$$

For β and $\boldsymbol{\delta}$, we cannot find closed for expressions and we must resort to numerical methods to minimize (21).

5. NUMERICAL EXPERIMENTS

We implemented the frequency-domain version of the algorithm as described in section 4. We chose the 256x256, 8-bit grayscale aerial image shown in Fig. 1(a) as the ideal, high-resolution image for our experiments. First, we synthesized blurred, shifted, and noisy low-resolution images from the high-resolution image. A 2-D Gaussian PSF was used to blur the high-resolution image. Spatial shifts were chosen randomly. White, Gaussian noise was added to the blurred and shifted high-resolution image. Finally, the blurred, shifted, noisy image was down-sampled to produce the low-resolution images used in simulation. An example of a shifted, blurred, and noisy image is shown in Fig 1(b). A 256 x 256 image was restored from four 128 x 128 images in our simulations. Convergence was declared when the negative of the likelihood function decreased by less than 0.01%.

Fig. 1(d) shows the result of restoration from four low-resolution images with a noise variance of 2.0 and a blur variance of 0.75. These parameters were chosen to be representative of a low-noise, moderately blurred case. Fig. 2 shows one of the images used to produce the image in Fig. 4. For visual comparison, Fig 1(c) shows a 256 x 256 image produced from Fig. 1(b) by simple bi-cubic interpolation. The visual improvement produced by our restoration method is quite noticeable. The PSNR of the image in Fig. 3, produced using bi-cubic interpolation, is 25.8. The PSNR of the image in Fig. 4, produced using our method, is 29.8, an improvement of 4 dB.

REFERENCES

[1] T.S. Huang and R. Y. Tsai, "Multiple frame image reconstruction and registration," in *Advances in Computer Vision and Image Processing*, T.S. Huang, Ed. Greenwich, CT: JAI, 1984, ch. 7.

[2] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing,* vol 3., pp. 233-242, May 1994..

[3] R. C. Hardie, K. J. Barnard, and E.E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Transactions on Image Processing*, vol. 6, no.12, pp. 1621-1633, Dec. 1997.

[4] A. K. Katsaggelos, ed., *Digital Image Restoration*, New York: Springer-Verlag, 1991.

[5] A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM algorithm," *J. Roy. Stat. Soc.*, vol B39, pp 1-37, 1977.

[6] A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos, "A General Framework for Frequency Domain Multi-Channel Signal Processing," *IEEE Trans. Image Processing*, vol. 2, no. 3, July 1993.

[7] B. D. Ripley, *Spatial Statistics*, Wiley, New York, 1981.

[8] R. Molina, B. D. Ripley, "Using spatial models as priors in astronomical images analysis", *J. Appl. Stat.*, vol.16, pp.193-206, 1989.

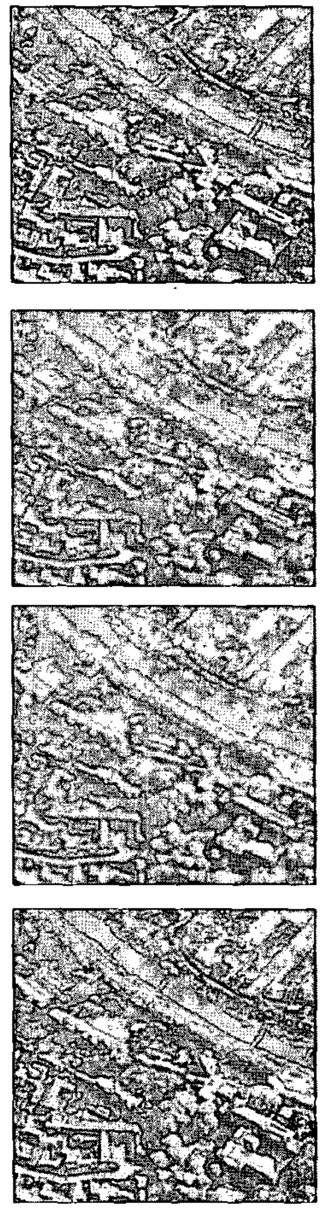

Figure 1: From top to bottom (a) Original Image (b) Shifted, Blurred, and Decimated Noisy Image (c) Bi-cubic Interpolation of Fig. 1(b) (d) Restored Super-Resolved Image.