Identification and Restoration of Noisy Blurred Images Using the Expectation-Maximization Algorithm

REGINALD L. LAGENDIJK, STUDENT MEMBER, IEEE, JAN BIEMOND, SENIOR MEMBER, IEEE, AND DICK E. BOEKEE

Abstract—In image restoration, it is nearly always assumed that the point-spread function of the degrading system, as well as the variance of the observation noise and a model of the original image, are known *a priori*. Since these parameters are unknown for practical images of interest, they have to be estimated from the noisy blurred images themselves. This paper presents a maximum likelihood approach to the blur identification problem, and proposes to employ the expectation-maximization algorithm to optimize the nonlinear likelihood function in an efficient way. In order to improve the performance of the identification algorithm, low-order parametric image and blur models are incorporated into the identification method. The resulting iterative technique simultaneously identifies and restores noisy blurred images.

I. INTRODUCTION

PROBABLY the best approach to the restoration of noisy blurred images would be to prevent the degradations from occurring at all. Unfortunately, in many cases these degradations cannot be avoided, and the data must be accepted as it is (e.g., blurred pictures of unique events). In other situations, the degradations can be prevented only at very high cost due to the required quality of the imaging system (e.g., in astronomical imaging), or by physical requirements which are unacceptable or unrealizable (e.g., in medical imaging). For these reasons, a variety of image restoration algorithms has been developed over the last decades, including Fourier domain [1], [2], recursive [3], [4], and iterative [5], [6] filters. However, in the use of these restoration methods, it is nearly always assumed that all the information that is required to restore an image is known *a priori*. Since this is uncommon for practical images of interest, one of the most challenging questions in image restoration nowadays is how to obtain this information.

If a blurred image is modeled as the output of a two-dimensional (2-D) linear system with finite impulse response, and the (unknown) original image as its input, the blur identification problem can be specified as the estimation of the unknown point-spread function (PSF) of this system from the observed noisy blurred image. The ear-

liest work on blur identification concentrated on PSF's which can parametrically be modeled, and which have a regular pattern of zeros on the unit bicircle, such as linear motion blur [1]. Both spectral and cepstral techniques were used to determine the distance between the zero crossings of the transfer function of the blur [7], [8]. By virtue of the parametric modeling, this distance immediately determines the PSF. Shortcomings of this method are that PSF's which do not satisfy the above mentioned conditions cannot be identified in this way, and that the presence of noise in the recorded data is not directly taken into account.

In more recent work on blur identification, the original image is first modeled as a 2-D autoregressive (AR) process [9]. Next, the identification problem is formulated as a maximum likelihood (ML) problem, which involves estimating the PSF, the observation noise variance, and the model of the original image. Tekalp *et al.* [9] showed that in a recursive estimation procedure, this ML problem can be interpreted as a 2-D autoregressive moving average (ARMA) model identification problem, where the AR coefficients determine the image model coefficients, and where the MA part determines the PSF of the blurring system. They were able to employ 2-D recursive estimation procedures by assuming that the PSF can be decomposed into 4 quarter-plane convolutional factors [9], [10]. Along the same lines, Wagner [11] proposed to use a dyadic factoring of symmetric noncausal PSF's in order to identify the horizontal and vertical blur components independently. Biemond *et al.* [12] showed that the 2-D ARMA identification can be done in parallel, where each of the parallel channels requires the identification of a 1-D complex ARMA process. In [13] and [14], Lagendijk *et al.* used gradient-based procedures to optimize the likelihood function: in [13] the gradients are computed analytically, and in [14] a 2-D Kalman restoration filter is incorporated in order to evaluate the gradients numerically.

In this paper, we present a class of blur identification algorithms which is based on the expectation-maximization (EM) algorithm to compute ML parameter estimates [15]. Unlike other blur identification methods [9]–[13] which require that the amount of noise in the observed

blurred image is negligible, the presence of noise in the observed data is directly taken into account. As a result, the proposed methods are particularly well suited for identifying blurred images which contain a considerable amount of noise. The resulting algorithms constitute iterative procedures which simultaneously identify and restore a noisy blurred image (see also [14], [16], [17]). In Section II we will start with the basic image and blur model development, followed by the formulation of the blur identification problem as an ML problem. The basic identification scheme will be derived in Section III. To improve the performance of this identification algorithm, low-order parametric image and blur models will be incorporated in Section IV. Finally, various identification and restoration results using the proposed techniques will be given in Section V.

II. PROBLEM FORMULATION

A. Image and Blur Model Development

In image restoration, it is appropriate to model a discrete original image $f(i, j)$, with size $N \times N$, by a 2-D autoregressive process of low order [18]

$$
\begin{aligned}
f(i, j) &= a(i, j) * f(i, j) + v(i, j) \\
&= \sum_{k, l \in S_a} a(k, l) f(i - k, j - l) + v(i, j).
\end{aligned}
$$

$$(1)$$

Here $a(k, l)$ are the MSE image model coefficients which minimize $E(v(i, j)^2)$, and S_a is the quarter-plane image model support (see Fig. 1). The modeling error $v(i, j)$ is a zero-mean homogeneous Gaussian distributed noise process with covariance $Q_v(i, j) = \sigma_v^2 \delta(i, j)$, which is independent of $f(i, j)$. A more compact notation of (1) can be arrived at by lexicographically ordering the image data [19], yielding

$$
f = Af + v.
\tag{2}
$$

Here A is the image model matrix, which has the size $N^2 \times N^2$. If a circular convolution is assumed in (1), then A has a block-circulant structure [19]. The probability density function (pdf) of f follows from the characteristics of v and the model (2), and is given by:

$$
\begin{aligned}
p(f; A, Q_v) &= \sqrt{\frac{\det|I - A|^2}{2\pi^{N^2} \det|Q_v|}} \\
&\quad \cdot \exp\left\{ -\frac{1}{2} f'(I - A)' Q_v^{-1}(I - A) f \right\}
\end{aligned}
$$

$$(3)$$

where $Q_v = \sigma_v^2 I$ is the $N^2 \times N^2$ (diagonal) covariance matrix of the modeling error with $\sigma_v^2 > 0$, and where $(I - A)$ is assumed to be nonsingular.

The observed image $g(i, j)$ is modeled as the output of a 2-D linear space-invariant system, which is characterized by its point-spread function $d(m, n)$. The observa-
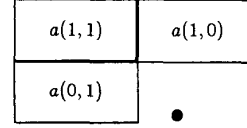

Fig. 1. Quarter-plane image model support S_a.

tion noise $w(i, j)$ is assumed to be an additive zero-mean homogeneous Gaussian distributed process with covariance $Q_w(i, j) = \sigma_w^2 \delta(i, j)$. Then, the noisy blurred image is given by [1]:

$$
\begin{aligned}
g(i, j) &= d(i, j) * f(i, j) + w(i, j) \\
&= \sum_{m, n \in S_d} d(m, n) f(i - m, j - n) + w(i, j).
\end{aligned}
$$

$$(4)$$

Here S_d is the support of the PSF $d(m, n)$, which can have any shape but is typically noncausal. Again, (4) can be written in a more compact matrix-vector notation by lexicographically ordering the data:

$$
g = Df + w.
\tag{5}
$$

The blur matrix D, with size $N^2 \times N^2$, has a block-circulant structure if a circular convolution is assumed in (4). The conditional probability density of g, given the original image, follows from the characteristics of the observation noise and the blurring system:

$$
\begin{aligned}
p(g/f; D, Q_w) &= \frac{1}{\sqrt{2\pi^{N^2} \det|Q_w|}} \\
&\quad \cdot \exp\left\{ -\frac{1}{2}(g - Df)' Q_w^{-1}(g - Df) \right\}
\end{aligned}
$$

$$(6)$$

where $Q_w = \sigma_w^2 I$ is the $N^2 \times N^2$ (diagonal) covariance matrix of the observation noise, and where we assume that $\sigma_w^2 > 0$.

B. Maximum Likelihood Blur Identification

The purpose of image restoration is to obtain an estimate $\hat{f}$, which is "as close as possible" to the original image f. In order to derive restoration filters, the PSF of the blurring system, the variance of the observation noise, and the model of the original image need to be available. Blur identification focuses on estimating the unknown parameters $\theta = \{d(m, n), a(k, l), \sigma_w^2, \sigma_v^2\}$ from the observed noisy blurred image $g(i, j)$, assuming that the structures of the models (1) and (4) are known *a priori*. The maximum likelihood estimator of θ is given by

$$
\hat{\theta}_{ml} = \arg\left\{ \max_{\theta \in \Theta} \mathcal{L}(\theta) \right\} = \arg\left\{ \max_{\theta \in \Theta} \log p(g; \theta) \right\}
\tag{7}
$$

where $\mathcal{L}(\theta)$ denotes the log-likelihood function of θ, where Θ specifies the range of the elements of θ, and where $p(g; \theta)$ is the pdf of g for a given θ. In order to

compute $p(g; \theta)$, we combine (2) and (5), yielding

$$g = Df + w = D(I - A)^{-1} v + w. \qquad (8)$$

Since v and w are independent Gaussian processes, $p(g; \theta)$ is Gaussian as well with zero mean, and covariance P given by:

$$
\begin{aligned}
P &= \text{cov } (g; \theta) \\
&= E\left\{ \left(D(I - A)^{-1} v + w\right)\left(D(I - A)^{-1} v + w\right)' \right\} \\
&= D(I - A)^{-1} Q_v (I - A)^{-t} D' + Q_w. \qquad (9)
\end{aligned}
$$

By substituting this result for $p(g; \theta)$ into (7), the ML blur identification problem can be expressed as follows:

$$\hat{\theta}_{ml} = \arg \max_{\theta \in \Theta} \left\{ -\log \left(\det |P|\right) - g' P^{-1} g \right\}. \qquad (10)$$

Unfortunately, (10) specifies a complicated nonlinear optimization problem in several variables (PSF, image model, and covariances), mainly because of the nonquadratic behavior of $\log (\det |P|)$. In the next section, we propose to employ the expectation-maximization algorithm to avoid the direct optimization in (10).

We note here, that in general the blur identification problem cannot be solved uniquely by using (10), because P is independent of the phases of $a(k, l)$ and $d(m, n)$. We therefore need to have *a priori* information about $a(k, l)$ and $d(m, n)$. Because the image model (1) should be (bibo) stable, its poles should lie within the unit bicircle. This requirement immediately determines the phase of $a(k, l)$ in a unique way. Although PSF's may theoretically have any shape, many PSF's found in image processing applications are symmetric. Following [9], [10], [12], the ML estimator of $d(m, n)$ is enforced to be symmetric, i.e., to have zero phase. Furthermore, image formation systems do normally not absorb or generate energy. Hence the solution to (10) should satisfy

$$\sum_{m, n \in S_d} d(m, n) = 1.0. \qquad (11)$$

Finally, the PSF has a finite support in nearly all practical applications, i.e., $d(m, n)$ is nonzero only over a (prescribed) finite window S_d. Due to the above conditions on the PSF, (10) becomes a constrained maximum likelihood problem, which is harder to solve than the original unconstrained optimization, but typically shows a unique optimum.

III. Simultaneous Identification and Restoration Algorithm

In this section, we will first briefly review the general theory of the expectation-maximization (EM) algorithm. Next, the EM-algorithm will be applied to ML blur identification, and the derivation of the various steps in the algorithm will be discussed in detail.

A. Review of the EM Algorithm

It has long since been recognized that computing maximum likelihood (ML) parameter estimates can be a highly complicated task in many relevant estimation problems. The EM algorithm, presented by Dempster *et al.* in [15], is a general iterative method to compute ML estimates if the observed data can be regarded as "incomplete." Since the EM algorithm was presented, similar or related algorithms have appeared in the literature, such as in [20]–[23]. It has been shown in various signal processing applications that the use of the EM-like algorithms leads to computationally efficient estimation algorithms [20], [22], [24]–[27].

Let $\mathcal{Y}$ denote the observed "incomplete" data which possesses the pdf $p(\mathcal{Y}; \theta)$, where θ again is the vector of parameters to be estimated. The ML estimator of θ, based on the available incomplete data, is given by

$$\hat{\theta}_{ml, \mathcal{Y}} = \arg \left\{ \max_{\theta \in \Theta} \mathcal{L}_{\mathcal{Y}}(\theta) \right\} = \arg \left\{ \max_{\theta \in \Theta} \log p(\mathcal{Y}; \theta) \right\} \qquad (12)$$

which is assumed to be a complicated problem. The incomplete data is related to some complete data, denoted by $\mathcal{X}$, through a noninvertible many-to-one transformation $\mathfrak{I}$:

$$\mathcal{Y} = \mathfrak{I}(\mathcal{X}). \qquad (13)$$

We restrict ourselves here to cases where $\mathfrak{I}$ represents a linear projection; for each given $\mathcal{Y}$ there exists a hyperplane $\mathcal{X}(\mathcal{Y})$ whose elements satisfy (13). The pdf of $\mathcal{X}$, which is also indexed by θ, is related to $\mathcal{Y}$ as follows:

$$p(\mathcal{Y}; \theta) = \int_{\mathcal{X}(\mathcal{Y})} p(\mathcal{X}; \theta) \, d\mathcal{X}. \qquad (14)$$

At this point it is assumed that the complete data have been chosen in such a way that computing the ML estimator of θ from the complete data, i.e., solving

$$\hat{\theta}_{ml, \mathcal{X}} = \arg \left\{ \max_{\theta \in \Theta} \mathcal{L}_{\mathcal{X}}(\theta) \right\} = \arg \left\{ \max_{\theta \in \Theta} \log p(\mathcal{X}; \theta) \right\} \qquad (15)$$

is significantly simpler than solving (12). However, the complete data is not available, but only observed via the noninvertible relation (13). Starting out with an estimate of the parameter vector, called $\hat{\theta}^{(k)}$, the EM algorithm finds the conditional expectation of the log-likelihood of complete data, denoted by $\mathcal{L}(\theta; \hat{\theta}^{(k)})$, given the observed incomplete data and $\hat{\theta}^{(k)}$:

$$
\begin{aligned}
\mathcal{L}(\theta; \hat{\theta}^{(k)}) &= E\left\{ \log p(\mathcal{X}; \theta) / \mathcal{Y}; \hat{\theta}^{(k)} \right\} \\
&= \int_{\mathcal{X}(\mathcal{Y})} \log p(\mathcal{X}; \theta) \, p(\mathcal{X}/\mathcal{Y}; \hat{\theta}^{(k)}) \, d\mathcal{X}
\end{aligned}
\qquad (16)
$$

where $p(\mathcal{X}/\mathcal{Y}; \hat{\theta}^{(k)})$ is the conditional pdf of the complete data, given the incomplete data and the estimate $\hat{\theta}^{(k)}$. Equation (16) is called the E step of the EM algorithm. In the M step, $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ is maximized with respect to θ. This leads to a new parameter estimate $\hat{\theta}^{(k+1)}$:

$$\hat{\theta}^{(k+1)} = \arg \left\{ \max_{\theta \in \Theta} \mathcal{L}(\theta; \hat{\theta}^{(k)}) \right\}. \qquad (17)$$

By alternating (16) and (17), the iterative EM algorithm is obtained, which converges to a *stationary point of* $\mathcal{L}_\mathcal{Y}(\theta)$ [15]. Sufficient convergence conditions are that $\mathcal{L}(\theta; \hat{\theta}^{(k)})$, which forms the basis of the EM algorithm, is continuous in both θ and $\hat{\theta}^{(k)}$ [15], [22], [28]. At each iteration cycle, the maximization in (17) ensures that $\mathcal{L}_\mathcal{Y}(\theta)$ increases. This maximization requirement can be relaxed somewhat. Namely, the EM algorithm converges for as long as $\hat{\theta}^{(k+1)}$ is chosen in such a way that $\mathcal{L}_\mathcal{Y}(\theta)$ is increased [15]. This variant is called a generalized EM (GEM) algorithm.

It is now observed that $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ and $\mathcal{L}_\mathcal{X}(\theta)$ have the same dependence on θ [15], [22], and that where $\mathcal{L}_\mathcal{X}(\theta)$ is defined on the true complete data, $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ uses the conditional expectation of the complete data. The maximization of $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ with respect to θ is therefore of the same complexity as the maximization of $\mathcal{L}_\mathcal{X}(\theta)$. Because of this, the EM algorithm is an attractive alternative to the direct evaluation of (12) only if the solution to (15) can be computed relatively easily. However, the EM algorithm does not necessarily converge to the global optimum of $\mathcal{L}_\mathcal{Y}(\theta)$, but instead it may stabilize at a local optimum. In such situations, various starting points $\hat{\theta}^{(0)}$ may be needed.

B. EM Algorithm Applied to Blur Identification

In image restoration and blur identification problems, the noisy blurred image is the only data available, which therefore establishes the incomplete data. Here we specify the complete data as the stacked lexicographically ordered images f and g:

$$\mathcal{X} = \begin{bmatrix} f \\ g \end{bmatrix}. \tag{18}$$

The complete and incomplete data are now related via

$$\mathcal{Y} = g = \begin{bmatrix} O & I \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix} = \Im\mathcal{X} \tag{19}$$

where O and I are an $N^2 \times N^2$ identity and zero matrix, respectively. The particular choice of this complete data satisfies the requirement that solving (15) should be easy, namely:

- The image model coefficients and the related modeling error variance σ_r^2 can be obtained immediately by the (linear) 2-D Yule–Walker equations, because the autocorrelation coefficients can be computed directly from the original image f.
- The point-spread function and the related observation noise variance σ_w^2 follow directly from a classical system identification problem, namely determining the impulse response of a linear system with known input (the original image) and noisy output (the observed image).

It is pointed out that it is not necessary to include the noise processes v and w into the definition of the complete data, because these do not provide any additional useful information for the ML identification. Other choices for $\mathcal{X}$ and $\mathcal{Y}$ are considered in [29]. Substituting the above

choices for $\mathcal{X}$ and $\mathcal{Y}$ into (16), yields the following E step:

$$\mathcal{L}(\theta; \hat{\theta}^{(k)}) = E\{\log p(f, g; \theta)/g; \hat{\theta}^{(k)}\}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \log p(f, g; \theta)$$

$$\cdot p(f/g; \hat{\theta}^{(k)}) \, df(1) \cdots df(N^2). \tag{20}$$

We now need to evaluate $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ for the case at hand. The joint probability density function $p(f, g; \theta)$ is readily found from (3) and (6):

$$p(f, g; \theta) = p(g/f; \theta) p(f; \theta) = \sqrt{\frac{\det |I - A|^2}{2\pi^{2N^2} \det |Q_r Q_w|}}$$

$$\cdot \exp \left\{ -\frac{1}{2} (g - Df)' Q_w^{-1} (g - Df) \right.$$

$$\left. -\frac{1}{2} f'(I - A)' Q_r^{-1} (I - A) f \right\}. \tag{21}$$

By combining $p(f, g; \theta)$ and $p(g; \theta)$, which is Gaussian with covariance given by (9), we get the following expression for the conditional pdf $p(f/g; \hat{\theta}^{(k)})$:

$$p(f/g; \hat{\theta}^{(k)})$$

$$= \frac{p(f, g; \hat{\theta}^{(k)})}{p(g; \hat{\theta}^{(k)})}$$

$$= \frac{1}{\sqrt{2\pi^{N^2} \det |\hat{V}^{(k)}|}}$$

$$\cdot \exp \left\{ -\frac{1}{2} (f - \hat{f}^{(k)})' \hat{V}^{(k)^{-1}} (f - \hat{f}^{(k)}) \right\}. \tag{22}$$

Here $\hat{f}^{(k)}$ and $\hat{V}^{(k)}$ denote the conditional mean and covariance of f, respectively, at the kth iteration (we have dropped the index (k) of $\hat{f}$, $\hat{V}$, D, A, Q_w, and Q_r in order to keep these equations readable):

$$\hat{f} = E(f/g; \hat{\theta}^{(k)}) = \hat{V} D' Q_w^{-1} g \tag{23}$$

$$\hat{V} = \text{cov}(f/g; \hat{\theta}^{(k)})$$

$$= [(I - A)' Q_r^{-1} (I - A) + D' Q_w^{-1} D]^{-1}. \tag{24}$$

Finally, substitution of $p(f, g; \theta)$ and $p(f/g; \hat{\theta}^{(k)})$ into (20) yields (see Appendix A):

$$\mathcal{L}(\theta; \hat{\theta}^{(k)}) = C - N^2 \log (\sigma_r^2 \sigma_w^2) + \log \det |I - A|^2$$

$$- \frac{1}{\sigma_w^2} g'g + \frac{2}{\sigma_w^2} \text{tr} \left\{ D\hat{\Re}_{fg}^{(k)} \right\}$$

$$+ -\frac{1}{\sigma_w^2} \text{tr} \left\{ D\hat{\Re}_{ff}^{(k)} D' \right\}$$

$$- \frac{1}{\sigma_r^2} \text{tr} \left\{ (I - A) \hat{\Re}_{ff}^{(k)} (I - A)' \right\} \tag{25}$$

where C is an additive constant term, tr $\{A\}$ denotes the trace of the matrix A, and where the conditional autocorrelation matrix $\hat{\Re}_{ff}^{(k)}$ and cross-correlation matrix $\hat{\Re}_{fg}^{(k)}$ are defined by

$$\hat{\Re}_{ff}^{(k)} = E(ff'/g; \hat{\theta}^{(k)}) = \hat{V}^{(k)} + \hat{f}^{(k)}\hat{f}^{(k)'} \quad (26)$$

$$\hat{\Re}_{fg}^{(k)} = E(fg'/g; \hat{\theta}^{(k)}) = \hat{f}^{(k)}g'. \quad (27)$$

Under the conditions that $\sigma_w^2 > 0$, $\sigma_v^2 > 0$, and that $(I - A)$ is nonsingular, $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ is continuous in θ and $\hat{\theta}^{(k)}$, and the EM algorithm (20), (17) will converge. Note that these conditions were already assumed in the development of (3) and (6), and thus do not impose additional constraints.

C. The E Step of the Algorithm

Equation (25) depends on $\hat{\theta}^{(k)}$ only through the conditional autocorrelation matrix $\hat{\Re}_{ff}^{(k)}$ and cross-correlation matrix $\hat{\Re}_{fg}^{(k)}$. Because of this, the E step of the algorithm consists of the mere evaluation of (23) and (24), and the computation of the conditional correlation matrices. Observe that this requires solving linear equations only.

Since A and D have block-circulant structures, (23) and (24) can be implemented efficiently by using 2-D discrete Fourier transforms [19]. As a consequence, $\hat{\Re}_{ff}^{(k)}$ and $\hat{\Re}_{fg}^{(k)}$ represent correlation matrices of weakly jointly stationary processes, and have block-circulant structures as well. We therefore replace (26) and (27) by

$$\hat{r}_{ff}^{(k)}(p, q) = \hat{V}^{(k)}(p, q) + \frac{1}{N^2} \sum_{i,j=1}^{N} \hat{f}^{(k)}(i, j)$$
$$\cdot \hat{f}^{(k)}(i - p, j - q) \quad (28)$$

$$\hat{r}_{fg}^{(k)}(p, q) = \frac{1}{N^2} \sum_{i,j=1}^{N} \hat{f}^{(k)}(i, j) g(i - p, j - q) \quad (29)$$

where $\hat{r}_{ff}^{(k)}(p, q)$ and $\hat{r}_{fg}^{(k)}(p, q)$ are the defining bisequences [19] of $\hat{\Re}_{ff}^{(k)}$ and $\hat{\Re}_{fg}^{(k)}$, respectively, and where $\hat{V}^{(k)}(p, q)$ is the defining bisequence of $\hat{V}^{(k)}$.

It should be noted that in image restoration (23) is known as a (constrained) least-squares filter, but may also be regarded as the (conditional) Bayesian estimator of the original image if θ is known completely [1]. Hence, in the process of identifying the image and blur parameters θ, a restoration result of the blurred image is obtained simultaneously in each E step of the iterations. This can be considered one of the advantages of the EM algorithm based approach towards blur identification, since the quality of a restoration result using the current parameter estimates can be evaluated at each iteration step, and the iterative process can be truncated when a (visually) stationary or acceptable solution has been reached.

D. The M Step of the Algorithm

In the M step, $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ is maximized with respect to the parameters θ, yielding a new estimate $\hat{\theta}^{(k+1)}$. As can be seen directly from (25), the optimizations with respect to the image model coefficients and PSF coefficients are independent problems, and can be considered separately.

1) Image Model Identification: By substituting (25) into (17), and by dropping all constant terms that are image model independent we arrive at the following optimization problem:

$$\{\hat{a}(k, l), \hat{\sigma}_v^2\} \leftarrow \max_{a(k,l),\sigma_v^2} \mathcal{J}(A, \sigma_v^2)$$

$$= \max_{a(k,l),\sigma_v^2} \left\{ \log \det |I - A|^2 - N^2 \log \sigma_v^2 \right.$$

$$\left. - \frac{1}{\sigma_v^2} \text{tr} \left\{ (I - A) \hat{\Re}_{ff}^{(k)} (I - A)' \right\} \right\}. \quad (30)$$

If we assume that the image model boundary effects are negligible, $(I - A)$ is a lower triangular matrix (because the image model support S_a is causal), and det $|I - A| = 1$. Therefore, (30) becomes quadratic in the image model coefficients $a(k, l)$, and its solution is given by the 2-D Yule–Walker equations:

$$\hat{r}_{ff}^{(k)}(p, q) = \sum_{k,l \in S_a} \hat{a}(k, l) \hat{r}_{ff}^{(k)}(p - k, q - l),$$

$$\forall(p, q) \in S_a \quad (31)$$

$$\hat{\sigma}_v^2 = \hat{r}_{ff}^{(k)}(0, 0) - \sum_{k,l \in S_a} \hat{a}(k, l) \hat{r}_{ff}^{(k)}(k, l). \quad (32)$$

2) Blur Model Identification: The PSF coefficients and the observation noise variance are computed by maximizing (25) with respect to $d(m, n)$ and σ_w^2. Substituting (25) into (17), and dropping all blur model independent terms, yields

$$\{\hat{d}(m, n), \hat{\sigma}_w^2\} \leftarrow \max_{d(m,n),\sigma_w^2} \mathcal{J}(D, \sigma_w^2)$$

$$= \max_{d(m,n),\sigma_w^2} \left\{ -N^2 \log \sigma_w^2 - \frac{1}{\sigma_w^2} g'g \right.$$

$$\left. + \frac{2}{\sigma_w^2} \text{tr} \{D\hat{\Re}_{fg}^{(k)}\} - \frac{1}{\sigma_w^2} \text{tr} \{D\hat{\Re}_{ff}^{(k)}D'\} \right\}. \quad (33)$$

In order to obtain a unique solution to (33), the PSF is enforced to be symmetric (cf. Section II-B):

$$\hat{d}(m, n) = \hat{d}(-m, -n), \quad \forall(m, n) \in S_{\bar{d}}. \quad (34)$$

To satisfy (11), we choose $d(0, 0)$ as:

$$\hat{d}(0, 0) = 1.0 - 2 \sum_{m,n \in S_{\bar{d}}} \hat{d}(m, n). \quad (35)$$

Here $S_{\bar{d}}$ is that subset of S_d which contains the unique, PSF defining coefficients $d(m, n)$ (e.g., $S_{\bar{d}}$ is a nonsymmetric half-plane). All other PSF coefficients can be derived from these defining coefficients via (34) and (35). Even with the above (linear) constraints, $\mathcal{J}(D, \sigma_w^2)$ is quadratic in $d(m, n)$. Solving (33) is tedious but straightforward, and leads to the following set of linear equa-

tions:

$$\hat{r}_{fg}^{(k)}(p, q) + \hat{r}_{fg}^{(k)}(-p, -q) - 2\hat{r}_{fg}^{(k)}(0, 0)$$
$$\quad - 2\hat{r}_{ff}^{(k)}(p, q) + 2\hat{r}_{ff}^{(k)}(0, 0)$$
$$= 2 \sum_{m,n \in S_{\bar{d}}} \hat{d}(m, n) \left\{ \hat{r}_{ff}^{(k)}(p - m, q - n) \right.$$
$$\quad + \hat{r}_{ff}^{(k)}(p + m, q + n) - 2\hat{r}_{ff}^{(k)}(m, n)$$
$$\left. \quad - 2\hat{r}_{ff}^{(k)}(p, q) + 2\hat{r}_{ff}^{(k)}(0, 0) \right\}, \forall (p, q) \in S_{\bar{d}}$$

$$(36)$$

$$\hat{\sigma}_w^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} g(i, j)^2 - \sum_{m,n \in S_d} \hat{d}(m, n) \hat{r}_{fg}^{(k)}(-m, -n).$$

$$(37)$$

We note here that other relevant linear constraints, such as additional symmetry properties of the PSF, can be incorporated into (33) as well, without significantly increasing the computational complexity of the blur model identification. It is also straightforward to incorporate the decomposition of PSF's into 4 quarter-plane convolutional factors as proposed in [9], [10], or the dyadic factoring of PSF's as in [11].

In the preceding sections, we have established an expectation-maximization blur identification algorithm. Whereas the original ML blur identification formulation required solving a highly complex optimization problem, the proposed iterative procedure involves solving linear equations only (see Fig. 2). In addition, a restoration result of the noisy blurred image is obtained simultaneously in each cycle of the algorithm. Examples of the proposed technique will be given in Section V. Experiments have shown, however, that if θ contains a large number of unknown parameters, for example due to a large blur, these cannot be identified satisfactorily. In the next section, we will look at how to reduce the dimension of θ in such situations.

IV. PARAMETRIC MODELING

A. Initial Conditions

Point-spread functions of degradations encountered in practice may have a support S_d of considerable extent. Without additional knowledge about the relations between the PSF coefficients, this requires the estimation of a large number of independent parameters. The blur identification method described in the previous section turns out to be less suitable for the "larger" types of blur for the following reasons.

In the first place, the log-likelihood function $\mathcal{L}(\theta)$ becomes insensitive to variations in θ if the PSF and the image model contain more than a few free coefficients, or, in other words, the extrema in $\mathcal{L}(\theta)$ become less pronounced. Since the EM algorithm is basically a hill-climbing algorithm, inaccuracies in solving (30) and (33) will prohibit convergence to the correct ML estimator. Furthermore, the convergence speed drops dramatically because of th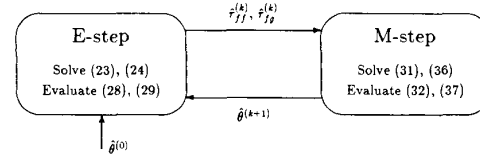is behavior of $\mathcal{L}(\theta)$. It is worthwhile noticing that this problem will occur in any maximum likelihood-based blur identification algorithm.

Fig. 2. Structure of the entirely linear expectation-maximization blur identification algorithm.

Second, in the development of the EM algorithm, the extent of the supports S_a and S_d have been assumed to be known *a priori*, cf. (31), (36). This is, however, hardly ever true for the support of the PSF. It is therefore usually suggested to overestimate S_d initially, and to remove small coefficients from the PSF after a preliminary identification, thus reducing S_d. This strategy increases, however, the number of parameters enormously, which again leads to the accuracy problems mentioned previously.

In the third place, the EM algorithm may converge only to a local optimum of $\mathcal{L}(\theta)$. For an increasing number of unknowns the number of suboptimal solutions, and hence the number of initial guesses $\hat{\theta}^{(0)}$ required, may grow unrealistically or unacceptably large.

Finally, it should be kept in mind that, in general, a blur identification process obtains useful information only from "edgy" areas in an image, because in the more gradual regions blur cannot be detected. Since edges appear only infrequently or, in other words, images contain relatively few high-frequency components, the total number of relevant data points is very limited. Increasing the support of the PSF will reduce the ratio between the number of relevant data and the number of unknown parameters even further, which leads inevitably to less reliable identification results.

We may circumvent some of the mentioned problems by initializing the EM algorithm with reasonable initial conditions $\hat{\theta}^{(0)}$. It is well known that the order of magnitude of the noise variance σ_w^2 can satisfactorily be estimated from a smooth image region. Based on the number of pixels involved in this estimation, confidence intervals on σ_w^2 can be obtained as well, which may be enforced during the identification process.

Since pixels in images are usually highly correlated, initial conditions for the image model coefficients are easily obtained, for example, by using an image model computed from a representative image. The log-variance of the modeling error, $\log \sigma_v^2$, is normally in the order of 1.5 to 2.5. It should be noted here that the image restoration result $\hat{f}$ in (23) depends only upon the ratio $\alpha = \sigma_w^2 / \sigma_v^2$, and is not very sensitive to the actual values of α and the image model coefficients.

The above reasoning cannot be applied to the initial conditions for $d(m, n)$, because, in the first place, $\hat{f}$ is very sensitive to variations in the PSF, and second, because the number of coefficients to be initialized (which may be very large) is unknown *a priori* (i.e., S_d is unknown). A solution to this dilemma might be to

perform an exhaustive search on the PSF coefficients, that is, to evaluate $\mathcal{L}(\theta)$ on a "rough" grid for fixed $\{\sigma_w^2, \sigma_v^2, a(k, l)\}$ and varying $d(m, n)$. Although this

Since $\mathcal{J}(\rho_v, \rho_h)$ is a quadratic expression in ρ_v for a fixed value of ρ_h, and vice versa, the following partial gradients based iteration is an efficient way to maximize (40):

$$\hat{\rho}_v^{(i)} = \frac{\hat{r}(1, 0) - \hat{\rho}_h^{(i-1)}\{\hat{r}(1, 1) + \hat{r}(1, -1)\} + \hat{\rho}_h^{(i-1)^2}\hat{r}(1, 0)}{\hat{r}(0, 0) - 2\hat{\rho}_h^{(i-1)}\hat{r}(0, 1) + \hat{\rho}_h^{(i-1)^2}\hat{r}(0, 0)} \qquad (41)$$

$$\hat{\rho}_h^{(i)} = \frac{\hat{r}(0, 1) - \hat{\rho}_v^{(i)}\{\hat{r}(1, 1) + \hat{r}(-1, 1)\} + \hat{\rho}_v^{(i)^2}\hat{r}(0, 1)}{\hat{r}(0, 0) - 2\hat{\rho}_v^{(i)}\hat{r}(1, 0) + \hat{\rho}_v^{(i)^2}\hat{r}(0, 0)}. \qquad (42)$$

approach works well for a relatively small number of PSF coefficients, it becomes unrealistic for an increasing number of unknowns due to the required amount of computation.

Motivated by the preceding discussion, we propose to reduce the number of unknown parameters in θ by assuming that we have some knowledge about the structure of the PSF and the image model. This *a priori* information is available in many cases of interest, simply because it is the key to concluding that an image is blurred indeed. In the following two sections, we will describe various parametric image and blur models, which can be incorporated directly into $\mathcal{L}(\theta; \hat{\theta}^{(k)})$. The E step of the algorithm is not influenced by using these parametric models; a modified M step will be given for each of the parametric models suggested.

B. Image Model

The structure of an original image is often such that its autocorrelation function resembles a 2-D separable exponentially decaying function [18]:

$$r_{ff}(p, q) = \sigma_f^2 \rho_v^{|p|} \rho_h^{|q|}. \qquad (38)$$

Here ρ_v and ρ_h denote the vertical and horizontal correlation coefficients, respectively. Since in image restoration, image models are merely used to stabilize the inversion of PSF's, it is not necessary to have the exact MSE coefficients $a(k, l)$ as obtained by (31). In order to reduce the number of unknown parameters, and to enforce an image model structure prior to the identification process, we assume that the image model coefficients fit the above autocorrelation function exactly. For the image model with quarter-plane support this leads to [18]

$$a(1, 0) = \rho_v, \quad a(0, 1) = \rho_h, \quad \text{and } a(1, 1) = -\rho_v \rho_h. \qquad (39)$$

Substitution of these values into (30) yields (we have dropped the subscripts and superscripts of $\hat{r}_{ff}^{(k)}(p, q)$ for simplicity):

$$\mathcal{J}(\rho_v, \rho_h) = -\hat{r}(0, 0) + 2\rho_v \hat{r}(1, 0) + 2\rho_h \hat{r}(0, 1)$$
$$- 2\rho_v \rho_h \{\hat{r}(1, 1) + \hat{r}(1, -1)\}$$
$$- \rho_v^2 \hat{r}(0, 0) - \rho_h^2 \hat{r}(0, 0)$$
$$+ 2\rho_v^2 \rho_h \hat{r}(0, 1) + 2\rho_v \rho_h^2 \hat{r}(1, 0)$$
$$- \rho_v^2 \rho_h^2 \hat{r}(0, 0). \qquad (40)$$

These iterations must be run in each M step, and converge usually within a few iteration steps. Observe that if $\mathcal{J}(\rho_v, \rho_h)$ is optimized only approximately, the EM algorithm turns into a GEM algorithm.

To reduce the number of image model parameters even more, we may choose $\rho = \rho_v = \rho_h$. By substitution of this condition into (40), optimizing $\mathcal{J}(\rho_v, \rho_h)$ becomes solving a third-order polynomial equation, which can be done analytically.

Causal image models have the obvious advantage that the determinant of the matrix $(I - A)$ is independent of the coefficients $a(k, l)$, hence optimizing $\mathcal{J}(A, \sigma_v^2)$ is always easy. However, this type of models may introduce artifacts in restored images which are characteristic for the causality definition used (e.g., smearing of edges in one specific direction). If, in a certain situation, these effects become dominant, the M step can easily be modified to account for noncausal image models which do not have any preference for a certain orientation. Parametric models, which are based on ρ_v and ρ_h, may again be employed to reduce the number of image model coefficients [18].

C. Blur Model

Besides the previously mentioned problems in identifying the PSF of the degrading system, it is in general not very realistic to model a PSF as a set of independent coefficients $d(m, n)$ considering the constraints enforced (e.g., (34) and (35)), and the fact that only a restricted subset of all possible $d(m, n)$ combinations will be accepted as representing realistic PSF's (e.g., $d(m, n)$ should represent a low-pass filtering action, should be "smooth," and $d(m, n) \geq 0.0$). For these reasons it is appropriate to consider a class of continuous parametric functions $\mathfrak{D}(x, y; \psi)$, which model realistic point-spread functions, and are indexed by ψ. Here ψ denotes the vector consisting of a few parameters which characterize the PSF. Relevant PSF's which can be modeled by such low-order parametric functions, are, for example:

• Linear motion blur over L pixels under an angle of ϕ radians:

$$\mathfrak{D}(x, y; L, \phi)$$

$$= \begin{cases} \dfrac{1}{C}, & \text{if } \sqrt{x^2 + y^2} \leq \dfrac{L}{2} \text{ and } \dfrac{y}{x} = \tan \phi \\ 0, & \text{elsewhere.} \end{cases} \qquad (43)$$

- Uniform out-of-focus blur with a radius of R pixels:

$$\mathfrak{D}(x, y; R) = \begin{cases} \dfrac{1}{C}, & \text{if } \sqrt{x^2 + y^2} \leq R \\ 0, & \text{elsewhere.} \end{cases} \quad (44)$$

- Blur due to atmospheric turbulence, Gaussian out-of-focus blur:

$$\mathfrak{D}(x, y; \sigma) = C \exp\left\{ -\frac{x^2 + y^2}{2\sigma^2} \right\}. \quad (45)$$

- Blur due to X-ray scatter [30]:

$$\mathfrak{D}(x, y; \beta) = \frac{C}{\left(\beta^2 + (x^2 + y^2)\right)^{3/2}}. \quad (46)$$

The constants C are chosen in such a way that the condition in (11) is satisfied, and some of the above PSF's have to be truncated properly to get a support of finite extent. Observe that spectral or cepstral techniques may be used as well to identify the linear motion and uniform out-of-focus blurs, but that this does not hold for the other blur models.

In order to compute the discrete PSF coefficients $d(m, n)$ from a continuous PSF $\mathfrak{D}(x, y; \psi)$, the discretization process has to be modeled. Assuming a simple square sensor array, we have

$$d(m, n) = \int_{\Omega(m,n)} \mathfrak{D}(x, y; \psi) \, dx \, dy \quad (47)$$

$$\Omega(m, n) = \left\{ x, y \,\middle|\, m - \frac{1}{2} \leq x \leq m + \frac{1}{2}, \right.$$
$$\left. n - \frac{1}{2} \leq y \leq n + \frac{1}{2} \right\}. \quad (48)$$

It is pointed out that since $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ has to be continuous in θ in order to guarantee convergence of the EM algorithm, each $d(m, n)$ needs to be continuously dependent of ψ. This condition restricts the possible set of parametric functions $\mathfrak{D}(x, y; \psi)$, and asks for an accurate (numerical) evaluation of (47).

By first substituting the selected parametric blur model into (47), and next substituting (47) into $\mathcal{G}(D, \sigma_w^2)$, (33) becomes a nonlinear optimization problem in ψ, to which an explicit solution can rarely be found. In general, (33) must be solved by numerical methods, which can be done efficiently since ψ is of low dimensionality. We have employed a straightforward steepest descent iteration to minimize $\mathcal{G}(D, \sigma_w^2)$, where the gradients were computed numerically. Convergence of these iterations is normally achieved within 5 to 10 iteration steps. Again, as in the case with the image model coefficients, exact minimization of $\mathcal{G}(D, \sigma_w^2)$ is not required as long as ψ is chosen in such a way that the likelihood function $\mathcal{L}(\theta)$ is increased in each M step (GEM algorithm).

V. EXPERIMENTAL RESULTS

A. Nonparametric Identification

In this first example we repeat the experiments of [10]. Whereas the results in [10] are based on noise-free data,

Fig. 3. Original cameraman image.

we added also observation noise to the blurred data. A "cameraman" image of size 256×256 pixels (Fig. 3), was synthetically blurred by two different types of out-of-focus blur, namely a truncated 3×3 Gaussian PSF and a 3×3 uniform PSF. Noise was added with $\sigma_w^2 = 3.6$, yielding an SNR of 30 dB:

$$\text{SNR} = 10 \log_{10} \left[\frac{\text{variance of } d(i, j) * f(i, j)}{\text{variance of } w(i, j)} \right] \quad \text{dB.} \quad (49)$$

The true and identified parameters using the nonparametric M step procedures in Section III-D, are summarized in Tables I and II for the two different cases considered. Convergence of the iterations was achieved within 50 EM iterations. It is observed that these identification results obtained from noisy blurred images are comparable to the ones given in [10] for noiseless data.

In these experiments, $S_{\bar{d}}$ contains 4 free coefficients (the extent of the PSF was assumed to be known in computing the above identification results); the total number of free parameters in θ is therefore equal to 9. Various other experiments indicate that blurs with larger support size (i.e., with more free PSF coefficients), cannot satisfactorily be identified without using additional knowledge about the structure of the PSF.

B. Parametric 2-D Gaussian Blurs

The image in Fig. 3 was synthetically blurred by Gaussian out-of-focus blurs with various standard deviations σ_G. The experiment included PSF's with supports ranging from very small ($\sigma_G = 0.30$ yields a PSF support of 3×3 pixels), to very large ($\sigma_G = 2.00$ yields a PSF with a support of 13×13 pixels). Noise was added up to the SNR level of 30 dB ($\sigma_w^2 \approx 3.5$).

The parametric EM identification algorithm was run on each of the cases tested with the following initial conditions: $\hat{\sigma}_G = 0.0$ (i.e., no blur), $\hat{\sigma}_w^2 = 1.0$, $\hat{\rho}_v = \hat{\rho}_h = 0.8$, and $\hat{\sigma}_r^2 = 200.0$. Convergence was achieved within 50 (small blurs) to 250 (large blurs) iterations. Table III lists the identification results. In this experiment it is obviously not necessary to fix the extent of the PSF *a priori*.

TABLE I
TRUE AND IDENTIFIED PARAMETERS FOR A TRUNCATED 3 × 3 GAUSSIAN
PSF AT SNR = 30 dB

$$d(m,n) = \begin{bmatrix} 0.075 & 0.124 & 0.075 \\ 0.124 & 0.204 & 0.124 \\ 0.075 & 0.124 & 0.075 \end{bmatrix} \quad \hat{d}(m,n) = \begin{bmatrix} 0.076 & 0.110 & 0.090 \\ 0.102 & 0.242 & 0.102 \\ 0.090 & 0.110 & 0.076 \end{bmatrix}$$

$$\sigma_w^2 = 3.60 \qquad\qquad \hat{\sigma}_w^2 = 3.34$$

$$a(k,l) = \begin{bmatrix} -0.506 & 0.727 \\ 0.759 & \bullet \end{bmatrix} \quad \hat{a}(k,l) = \begin{bmatrix} -0.578 & 0.771 \\ 0.792 & \bullet \end{bmatrix}$$

$$\sigma_v^2 = 237.10 \qquad\qquad \hat{\sigma}_v^2 = 206.11$$

TABLE II
TRUE AND IDENTIFIED PARAMETERS FOR A 3 × 3 UNIFORM PSF AT
SNR = 30 dB

$$d(m,n) = \begin{bmatrix} 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \end{bmatrix} \quad \hat{d}(m,n) = \begin{bmatrix} 0.102 & 0.113 & 0.114 \\ 0.106 & 0.131 & 0.106 \\ 0.114 & 0.113 & 0.102 \end{bmatrix}$$

$$\sigma_w^2 = 3.50 \qquad\qquad \hat{\sigma}_w^2 = 3.71$$

$$a(k,l) = \begin{bmatrix} -0.506 & 0.727 \\ 0.759 & \bullet \end{bmatrix} \quad \hat{a}(k,l) = \begin{bmatrix} -0.600 & 0.770 \\ 0.816 & \bullet \end{bmatrix}$$

$$\sigma_v^2 = 237.10 \qquad\qquad \hat{\sigma}_v^2 = 186.56$$

TABLE III
IDENTIFIED PARAMETERS FOR GAUSSIAN OUT-OF-FOCUS BLURS AT
SNR = 30 dB

σ_G	S_d	$\hat{\sigma}_G$	$\hat{\sigma}_w^2$	$\hat{\rho}_v$	$\hat{\rho}_h$	$\hat{\sigma}_v^2$	η_{snr}
0.300	3x3	0.365	20.30	0.768	0.825	247.3	0.42
0.500	5x5	0.530	5.08	0.768	0.817	232.2	5.66
0.750	7x7	0.774	3.71	0.782	0.829	219.6	4.45
1.000	9x9	0.978	3.46	0.829	0.827	168.3	3.14
1.500	11x11	1.475	3.34	0.873	0.837	133.3	2.32
2.000	13x13	1.988	3.20	0.898	0.843	112.7	2.08

Fig. 4. Blurred cameraman image.

Fig. 5. Restoration of Fig. 4 using the identified parameters.

The indicated size of S_d was directly derived from the identified value of σ_G by truncation of the PSF at coefficients smaller than approximately 0.1% of $d(0,0)$. From the original image we identified $\rho_v = 0.759$, $\rho_h = 0.796$, and $\sigma_{f}^2 = 243.3$. The performance measure η_{snr} denotes the improvement in signal-to-noise ratio as follows:

$$\eta_{snr} = 10 \log_{10} \frac{\sum_{i,j=1}^{N} \left(f(i,j) - g(i,j)\right)^2}{\sum_{i,j=1}^{N} \left(f(i,j) - \hat{f}(i,j)\right)^2} \quad \text{dB.} \quad (50)$$

It was observed that the SNR improvement of the restoration results using the identified parameters was in the same range as the SNR improvement of the restoration results using the true parameters.

It is worthwhile noticing that for small values of σ_G, the estimates $\hat{\sigma}_G$ become unreliable, and so does the estimation of σ_w^2. This is due to the fact that small blurs do not degrade the image spectrum enough to allow their identification, which is known as the "identifiability problem." On the other hand, for large values of σ_G the estimations of the correlation coefficients and the modeling error becomes more inaccurate due to the severe loss of

high image frequencies. Figs. 4 and 5 show the blurred image with $\sigma_G = 1.000$, and the restored image with $\hat{\sigma}_G = 0.978$, respectively.

In order to show the robustness of the blur identification procedure with respect to the image model coefficients, we selected the following fixed image model:

$$a(k,l) = \begin{bmatrix} -0.72 & 0.85 \\ 0.85 & \bullet \end{bmatrix}$$

and a fixed modeling error variance $\sigma_f^2 = 200.0$. Table IV lists the identification results using this fixed image model. Fig. 6 shows the restoration result of the blurred image in Fig. 4 using the fixed image model and the identified σ_G and σ_w^2. Observe that, in general, the estimates σ_G are less accurate and that the signal-to-noise-ratio improvements are smaller than the corresponding entries in Table III.

Finally we consider the breakdown of the identification procedure for decreasing signal-to-noise ratio. The original cameraman image was blurred with $\sigma_G = 1.00$, and noise was added up to the levels of 50, 40, 30, 20, 10, and 3 dB. Table V lists the identification results for this experiment. From these results and from a variety of other

TABLE IV
IDENTIFIED PARAMETERS FOR GAUSSIAN OUT-OF-FOCUS BLURS AT
SNR = 30 dB USING A FIXED IMAGE MODEL

σ_G	$\hat{\sigma}_G$	$\hat{\sigma}_w^2$	η_{snr}
0.300	0.280	21.7	-1.31
0.500	0.467	5.55	5.01
0.750	0.732	3.80	4.38
1.000	0.958	3.50	2.98
1.500	1.493	3.34	2.35
2.000	2.041	3.20	2.11

Fig. 6. Restoration of Fig. 4 using a fixed image model and identified blur parameters

TABLE V
IDENTIFIED PARAMETERS FOR GAUSSIAN OUT-OF-FOCUS BLUR WITH
$\sigma_G = 1.00$ AT VARIOUS SNR's

SNR	$\hat{\sigma}_G$	$\hat{\sigma}_w^2$	$\hat{\rho}_v$	$\hat{\rho}_h$	$\hat{\sigma}_v^2$
3	1.251	1759	0.907	0.861	116.7
10	1.095	350.9	0.895	0.869	97.9
20	0.980	35.2	0.860	0.847	133.4
30	0.978	3.46	0.829	0.827	168.3
40	0.983	0.34	0.804	0.819	185.5
50	0.985	0.033	0.786	0.813	191.9

experiments it turns out that the identification performs well even for relatively low SNR's, particularly with respect to the identified PSF. However, for a decreasing SNR the identified image model parameters become inaccurate, indicating that the proposed method may be too restrictive if the amount of noise exceeds the value of approximately 20 dB.

C. 1-D Linear Motion Blur

The last experiment deals with the photographically blurred train image[1] in Fig. 7. Because the blur is space variant in this case, only the blurred section of this image was used for identification and restoration purposes. A parametric model for horizontal linear motion blur was used, yielding an estimated length of motion $\hat{L} = 8.31$ pixels, $\hat{\sigma}_w^2 = 0.40$, $\hat{\rho}_v = 0.784$, $\hat{\rho}_h = 0.767$, and $\hat{\sigma}_v^2 = 59.16$. This result for the length of motion confirms the result of $\hat{L} = 8$ or $\hat{L} = 9$ pixels reported in the literature

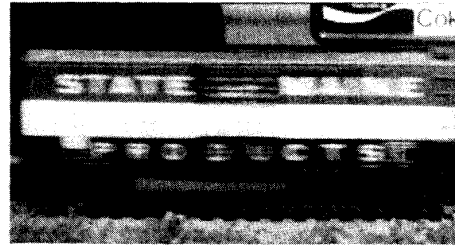[1]Courtesy of Eastman Kodak Company, Rochester, NY.
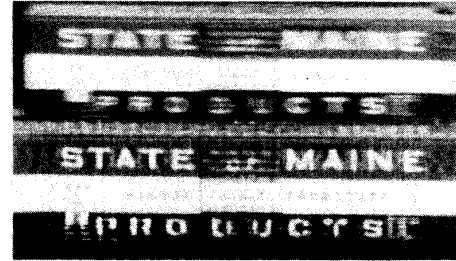

Fig. 7. Blur introduced by real motion.

Fig. 8. Blurred section of Fig. 7, and its restoration results using identified parameters.

(for example [6]) using a cepstral technique of [8]. The restoration result of the blurred section of the image using the identified parameters is shown in Fig. 8.

VI. DISCUSSION

In this paper, we have proposed the application of the EM algorithm to maximum likelihood blur identification. As opposed to earlier work in blur identification, the presence of noise in the observed blurred image is directly incorporated into the identification technique. As a result, the proposed method is not restricted to data in which the amount of noise is negligible, but performs well even in noisy data. A practical bound for the amount of noise the method is still able to handle, seems to be an SNR of approximately 20 dB. Low-order parametric image and blur models were incorporated into the identification scheme in order to make the identification algorithm applicable to more realistic blurs and to improve the identification results. An additional advantage of these parametric models is that they can be initialized more easily than the nonparametric ones. It can be considered to be one of the major advantages of the EM algorithm that the use of parametric models affects the M step only, without changing the E step. Furthermore, within the M step, the blur and image model identification remain independent problems.

In some situations of practical interest, such as the image in Fig. 7, the observation model (4) cannot be used because the PSF of the blur varies over the image. This requires space-variant blur identification, which can be done with the proposed algorithm only in a sliding-window approach. This is one of the subjects of current research in image identification and restoration.

APPENDIX A
DERIVATION OF EQUATION (25)

First, (21) is substituted into (20), yielding

$$\mathcal{L}(\theta; \hat{\theta}^{(k)}) = E\{ \log p(f, g; \theta)/g; \hat{\theta}^{(k)} \}$$

$$= -\frac{1}{2} \log (2\pi^{2N^2}) - \frac{1}{2} \log \det |Q_w Q_v|$$

$$+ \frac{1}{2} \log \det |I - A|^2$$

$$- \frac{1}{2} E\{ (g - Df)^t Q_w^{-1}(g - Df)$$

$$+ f^t(I - A)^t Q_v^{-1}(I - A)f/g; \hat{\theta}^{(k)} \}$$

$$= C - \frac{1}{2} N^2 \log (\sigma_w^2 \sigma_v^2) + \frac{1}{2} \log \det |I - A|^2$$

$$- \frac{1}{2} E\{ (g - Df)^t$$

$$\cdot Q_w^{-1}(g - Df)/g; \hat{\theta}^{(k)} \}$$

$$- \frac{1}{2} E\{ f^t(I - A)^t Q_v^{-1}(I - A)f/g; \hat{\theta}^{(k)} \}. \tag{51}$$

If we have a closer look at the fourth term of (51), we find

$$E\{ (g - Df)^t Q_w^{-1}(g - Df)/g; \hat{\theta}^{(k)} \}$$

$$= E\{ g^t Q_w^{-1} g/g; \hat{\theta}^{(k)} \} - 2E\{ f^t D^t Q_w^{-1} g/g; \hat{\theta}^{(k)} \}$$

$$+ E\{ f^t D^t Q_w^{-1} Df/g; \hat{\theta}^{(k)} \}$$

$$= \frac{1}{\sigma_w^2} g^t g - \frac{2}{\sigma_w^2} \text{tr} \{ D E(fg^t/g; \hat{\theta}^{(k)}) \}$$

$$+ \frac{1}{\sigma_w^2} \text{tr} \{ D E(ff^t/g; \hat{\theta}^{(k)}) D^t \} \tag{52}$$

where tr (A) denotes the trace of the matrix A. A similar derivation holds for the last term of (51):

$$E\{ f^t(I - A)^t Q_v^{-1}(I - A)f/g; \hat{\theta}^{(k)} \}$$

$$= \frac{1}{\sigma_v^2} \text{tr} \{ (I - A) E(ff^t/g; \hat{\theta}^{(k)})(I - A)^t \}. \tag{53}$$

Now, we define the $N^2 \times N^2$ conditional correlation matrices $\hat{\mathcal{R}}_{ff}^{(k)}$ and $\hat{\mathcal{R}}_{fg}^{(k)}$ as follows:

$$\hat{\mathcal{R}}_{ff}^{(k)} = E(ff^t/g; \hat{\theta}^{(k)})$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} ff^t p(f/g; \hat{\theta}^{(k)}) df(1) \cdots df(N^2)$$

$$= \text{cov}(f/g; \hat{\theta}^{(k)}) + E(f/g; \hat{\theta}^{(k)}) E(f/g; \hat{\theta}^{(k)})^t$$

$$= \hat{V}^{(k)} + \hat{f}^{(k)} \hat{f}^{(k)t} \tag{54}$$

$$\hat{\mathcal{R}}_{fg}^{(k)} = E(fg^t/g; \hat{\theta}^{(k)})$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} fg^t p(f/g; \hat{\theta}^{(k)}) df(1) \cdots df(N^2)$$

$$= E(f/g; \hat{\theta}^{(k)}) g^t = \hat{f}^{(k)} g^t. \tag{55}$$

Here $\hat{f}^{(k)}$ and $\hat{V}^{(k)}$ are, respectively, the conditional mean and covariance of f, given the observed image g and the estimate of θ at the kth iteration. By first substituting $\hat{\mathcal{R}}_{ff}^{(k)}$ and $\hat{\mathcal{R}}_{fg}^{(k)}$ into (52) and (53), next substituting (52) and (53) into (51), and finally multiplying the result by a factor 2 for simplicity, we arrive at (25).

REFERENCES

[1] H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* Englewood Cliffs, NJ: Prentice-Hall, 1977.
[2] R. C. Gonzalez and P. Wintz, *Digital Image Processing.* Reading, MA: Addison Wesley, 1987.
[3] J. W. Woods and V. K. Ingle, "Kalman filtering in two-dimensions: Further results," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 29, no. 2, pp. 188-197, 1981.
[4] J. Biemond, "Stochastic linear image restoration," in *Advances in Computer Vision and Image Processing,* vol. 2. London: JAI Press, 1986, pp. 213-273.
[5] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Regularized iterative image restoration with ringing reduction," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 36, no. 12, pp. 1874-1888, 1988.
[6] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, "Iterative methods for image deblurring," *Proc. IEEE,* June 1990.
[7] T. G. Stockham, T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. IEEE,* vol. 64, no. 4, 1975, pp. 678-692.
[8] M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-24, pp. 58-63, 1976.
[9] A. M. Tekalp, H. Kaufman, and J. W. Woods, "Identification of image and blur parameters for the restoration and noncausal blurs," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 34, pp. 963-972, 1986.
[10] A. M. Tekalp and H. Kaufman, "On statistical identification of a class of linear space-invariant blurs using nonminimum-phase ARMA models," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 36, no. 8, pp. 1360-1363, 1988.
[11] G. R. Wagner, "Self-tuning algorithms for two-dimensional signal processing," Ph.D. dissertation, Univ. of Manchester, Institute of Science and Technology, Control Systems Centre, Manchester, U.K., 1987.
[12] J. Biemond, F. G. van der Putten, and J. W. Woods, "Identification and restoration of images with symmetric noncausal blurs," *IEEE Trans. Circuits Syst.,* vol. 23, no. 3, pp. 385-394, 1988.
[13] R. L. Lagendijk, A. K. Katsaggelos, and J. Biemond, "Iterative identification and restoration of images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing 1988* (New York, NY), 1988, pp. 992-995.
[14] R. L. Lagendijk, D. L. Angwin, H. Kaufman, and J. Biemond, "Recursive and iterative methods for image identification and restoration," in *Proc. Fourth Euro. Signal Processing Conf. EUSIPCO'88,* J. G. Lacoume, Ed. (Grenoble, France), Sept. 1988, pp. 235-238.
[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data," *J. Roy. Statist. Soc. B,* vol. 39, 1977, pp. 1-38.
[16] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Simultaneous image identification and restoration using the EM-algorithm," in *Proc. SPIE Int. Conf. Visual Comm., Image Processing 1988* (Cambridge, MA), Nov. 1988, pp. 2-9.
[17] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Blur identification using the expectation-maximization algorithm," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing 1989* (Glasgow, Scotland), May 1989, pp. 1397-1400.
[18] A. K. Jain, "Advances in mathematical models for image processing," *Proc. IEEE,* vol. 69, no. 5, pp. 502-528, 1981.
[19] W. K. Pratt, "Vector space formulation of two-dimensional signal

processing operations," *Comput. Graphics and Image Processing*, no. 4, pp. 1-24, 1975.

[20] B. R. Musicus and J. S. Lim, "Maximum likelihood parameter estimates of noisy data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing 1979*, (Washington DC), pp. 224-227.

[21] Y. Bar-Shalom, "Optimal simultaneous state estimation and parameter identification in linear discrete-time systems," *IEEE Trans. Automat. Contr.*, vol. 17, no. 3, pp. 308-319, 1972.

[22] M. Feder, "Statistical signal processing using a class of iterative estimation algorithms," Ph.D. dissertation, Res. Lab. Electron., Dep. of EECS, Massachusetts Institute of Technology, 1987.

[23] M. Segal and E. Weinstein, "The cascade EM algorithm," *Proc. IEEE*, vol. 76, no. 10, pp. 1388-1390, 1988.

[24] A. Dembo and O. Zeitouni, "Maximum *a posteriori* estimation of time-varying ARMA processses from noisy data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 471-476, 1988.

[25] A. Dembo, "Signal reconstruction from noisy partial information of its transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 1, pp. 65-72, 1989.

[26] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 95-103, 1988.

[27] M. V. Ranganath, A. P. Dhawan, and N. Mullani, "A multigrid expectation-maximization reconstruction algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 7, no. 4, pp. 273-277, 1988.

[28] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95-103, 1983.

[29] A. K. Katsaggelos and K. T. Lay, "Simultaneous identification and restoration of images using maximum likelihood estimation," in *Proc. 1989 IEEE Conf. Contr. Applic.* (Jerusalem, Israel), pp. TP-5-1.

[30] F. C. Wagner, A. Macovski, and D. G. Nishimura, "A characterization of the scatter point-spread function in term of air gaps," *IEEE Trans. Med. Imaging*, vol. 7, no. 4, pp. 337-344, 1988.

Reginald L. Lagendijk (S'87) was born in Leiden, The Netherlands, on April 22, 1962. He received the M.Sc. and Ph.D. degrees from the Delft University of Technology, The Netherlands, in 1985 and 1990, respectively.

He is currently on the Faculty of the Department of Electrical Engineering at the Delft University of Technology. His research interests include information theory and multidimensional signal processing, with emphasis on image identification and restoration, and data compression of image sequences.

Jan Biemond (M'80-SM'87) was born in De Kaag, The Netherlands, on March 27, 1947. He received the M.S. and Ph.D. degrees from the Delft University of Technology, Delft, The Netherlands, in 1973 and 1982, respectively.

He is currently Professor in the Laboratory for Information Theory of the Department of Electrical Engineering, University of Technology. His research interests include multidimensional signal processing, image enhancement and restoration, data compression of images, and motion estimation with application in image coding and computer vision. He has authored and coauthored over 40 papers in these fields. In 1983 he was Visiting Researcher at Rensselaer Polytechnic Institute, Troy, NY, and at the Georgia Institute of Technology, Atlanta.

Dr. Biemond is a member of the IEEE Signal Processing Technical Committee on Multidimensional Signal Processing and of the IEEE Circuits and Systems Technical Committee on Visual Signal Processing and Communication. He served as the General Chairman of the Fifth ASSP/EURASIP Workshop on Multidimensional Signal Processing, held at Noordwijkerhout, The Netherlands, in September 1987. He is coeditor of the *International Journal on Multidimensional Systems and Signal Processing*.

Dick E. Boekee was born in the Hague, The Netherlands, in 1943. He received the M.Sc. and Ph.D. degrees in electrical engineering in 1970 and 1977, respectively, from the Delft University of Technology, Delft, The Netherlands.

In 1981 he became a Professor of Information Theory at the Delft University of Technology. During 1979-1980 he was a Visiting Professor at the Department of Mathematics, Katholieke Universiteit Leuven, Heverlee, Belgium. His research interests include information theory, image coding, cryptology, and signal processing.