Object Recognition Based on Impulse Restoration Using the Expectation-Maximization Algorithm

Ahmad Abu-Naser[1], Nikolas P. Galatsanos[1,†], Miles N. Wernick[1], and Dan Schonfeld[2]

[1] Department of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago, IL 60616
Tel: (312) 567-5259
E-mail: npg@ece.iit.edu

[2] Department of Electrical Engineering
and Computer Science (m/c 154)
University of Illinois at Chicago
851 South Morgan Street - 1120 SEO
Chicago, IL 60607-7053

ABSTRACT

It has recently been demonstrated that object recognition can be formulated as an image-restoration problem. In this approach, which we term impulse restoration, the objective is to restore a delta function indicating the detected object's location. Here we develop solutions based on impulse restoration for the Gaussian noise case. We propose a new iterative approach, based on the expectation-maximization (EM) algorithm, that simultaneously estimates the background statistics and restores a delta function at the location of the template. We use a Monte-Carlo study and localization-receiver-operating characteristics (LROC) curves to evaluate the performance of this approach quantitatively and compare it with existing methods. We present experimental results that demonstrate that impulse restoration is a powerful approach for detecting known objects in images severely degraded by noise. Our numerical experiments point out that the proposed EM-based approach is superior to all tested variants

of the matched filter. This demonstrates that accurate modeling and estimation of the background and noise statistics are crucial for realizing the full potential of impulse restoration-based template-matching.

---

†Corresponding author

1. INTRODUCTION

Template matching, which consists of the identification and localization of known objects (templates) in a scene, is an important image-processing task with application in novelty detection, motion estimation, target recognition, industrial inspection, and sensor fusion. The classical approach to this problem is the matched filter (MF) which is optimal in the sense that it is the filter for which the signal-to-noise ratio (SNR) of the output is maximized.[1] However, the MF is known to have several important drawbacks, including: 1) it is very sensitive to small variations of the object; 2) it can produce very broad correlation peaks that do not localize the object well;[2] and 3) it fails in the presence of background areas of high intensity because these areas produce high peaks in the output.

Attempts to overcome these limitations have led investigators to propose a number of modifications to the classical MF, among the most important being the phase-only MF (POMF) and its variants (see, for example, Refs. 3,4,5). It has been widely recognized that the phase of the Fourier transform is more important to describing an image than is its magnitude (see, for example, Ref. 6). Phase-only matched filters (POMF) and symmetric phase-only matched filters (SPOMF) have been proposed, and have been shown to provide better discrimination ability and noise robustness than the MF (see, for example, Refs. 2,3,7 and the references within).

Filters in use in joint transform correlators have been proposed that further improve the capabilities of the MF.[8] These filters are based on normalization of the correlation signal by the

background power spectrum. Methods for estimation of the power spectrum of the background are given in Ref. 8.

A different way of viewing the design of template-matching filters has recently emerged. In this new approach the problem of template matching is framed as one of image restoration, in which the image to be restored is an impulse at the location of the object to be detected. To our knowledge the first explicit application of this principle was reported in Ref. 9, in which the use of expansion matching, i.e., self-similar non-orthogonal basis decomposition, was proposed. More recently, the relationship between linear-minimum-mean-squared-error (LMMSE) impulse restoration and template matching was recognized and impulse-restoration filters were proposed for template-matching applications.[10,11,12] In these papers it was demonstrated that impulse restoration is a very powerful approach to template matching and has the potential to provide better localization, class discrimination, and error tolerance than the MF and its variants. Independently, the relationship between template matching and impulse restoration was recognized in previous papers.[13,14,15] In these papers LMMSE-based impulse restoration was used to find the displacement vector field (DVF) from noisy and/or blurred image sequences.

A critical shortcoming of the earlier work on impulse restoration has been the assumption that the background and/or noise statistics are known, which is not the case in practice. In realistic applications, one is required to estimate these statistics from the observed image. Without accurate estimates, the performance of impulse-restoration methods is severely compromised.

In Ref. 8 three different approaches to estimate the backgound power spectrum were introduced. In Refs. 13,14,15 one-step periodogram-based approaches were proposed to estimate the background statistics; in this paper, a new approach is proposed based on the expectation-maximization (EM) algorithm.[16]

In this paper we propose a new technique for template matching based on the approach of restoring an impulse indicating the object location. We describe an expectation-maximization (EM) algorithm [16] that simultaneously estimates the impulse along with the statistics of the noise and background. This algorithm is similar in philosophy to the application of EM to the image-restoration problem in Ref. 17. A major advantage of the impulse-restoration viewpoint is that it permits the substantial base of knowledge gained in the image-restoration field to be brought to bear on the template-matching problem.

Another important issue not often addressed in the object-recognition literature is the quantitative evaluation of task performance. Frequently, figures of merit are used that only indirectly measure the ability of an algorithm to detect and locate objects with accuracy. Such performance measures include the mean-square error (MSE) between the filter output and the desired impulse, and the peak-to-sidelobe (PTS) ratio of the filter output.[18] While these measures certainly have some relationship with detection performance, they provide only indirect evidence.

To make quantitative evaluations of the performance of various techniques, we use an extension of the receiver operating characteristic (ROC) curve, called the localization ROC (LROC)

curve.[19] The ROC curve, which plots the probability of detection versus the probability of false alarm for the continuum of possible decision thresholds, is a comprehensive description of the detection performance of an algorithm (see, for example, Ref. 20). The LROC curve plots the probability of detection *and* correct localization of an object versus the false-alarm probability, thus it also measures the ability of the detection algorithm to locate objects correctly. The LROC curve thoroughly and directly describes the detection and localization capability of an algorithm, thus it is much more informative than criteria such as MSE or PTS.

The remainder of this paper is organized as follows. In Section 2 we review the impulse-restoration formulation for the template-matching problem. In Section 3 we introduce a maximum-likelihood (ML) approach based on the EM algorithm that simultaneously estimates the background statistics and restores an impulse at the location of the template. In Section 4 we briefly review the LROC curve. In Section 5 we present numerical experiments and LROC-based comparisons demonstrating the proposed impulse-restoration template-matching algorithm. Finally, in Section 6 we present our conclusions and suggestions for future work.

2. BACKGROUND

A. Object Recognition as Impulse Restoration

In this section, the formulation of object recognition as an impulse-restoration problem is reviewed. For notational simplicity, images are represented throughout the paper by one-

dimensional signals. Consider the observed image $g(n)$, $n = 0, 1, \ldots, N' - 1$ in which the known object $f(n)$ is located at the unknown position n_0. One can then represent the image by

$$g(n) = f(n - n_0) + b(n), \tag{1}$$

where $b(n)$ is the combined background and noise. The goal of template matching is to find the location n_0 of the known object $f(n)$ within the image $g(n)$. Equation (1) can be rewritten as

$$g(n) = f(n) \star \delta(n - n_0) + b(n), \tag{2}$$

where $\delta(n - n_0)$ represents a discrete impulse function located at n_0 and $\star$ denotes convolution.

Using a matrix-vector notation and circular convolutions we write Eq. (2) as

$$\vec{g} = F\vec{\delta} + \vec{b}, \tag{3}$$

where $\vec{g}$, $\vec{b}$ and $\vec{\delta}$ are $N \times 1$ vectors representing $g(n)$, $b(n)$ and $\delta(n)$, respectively. In the rest of this paper, small letters with arrows represent vectors and capital letters represent matrices. In Eq. (3) $\vec{\delta}$ is a vector having zeros everywhere except in element n_0 which has a value of 1. The matrix F is an $N \times N$ circulant matrix constructed from the template $f(n)$ such that $F\vec{\delta}$ represents the convolution $f(n) \star \delta(n - n_0)$. In general, the equivalency of Eqs. (2) and (3) can be guaranteed by choosing $N \geq N'$.[21] Note that if each image vector is obtained by lexicographic ordering of a two-dimensional image, then F is a block-circulant matrix.

In this formulation it becomes apparent that the problem of detecting and locating the template can be viewed as a restoration problem in which $\delta(n - m)$ is the signal to be restored

and the template $f(n)$ is the point spread function (PSF).

B. LMMSE Formulations

An LMMSE framework can be used to estimate $\vec{\delta}$ from the observation $\vec{g}$ in Eq. (3). We will assume signals $\vec{\delta}$ and $\vec{b}$ to have zero mean with covariance matrices C_δ and C_b, respectively. Moreover, we assume signals $\vec{\delta}$ and $\vec{b}$ to be stationary; thus, C_b is an $N \times N$ circulant covariance matrix with eigenvalues equal to the power-spectrum coefficients $S_b(k)$, $k = 0, 1, \ldots, N-1$. Since $\vec{\delta}$ is a discrete impulse function, $C_\delta = I$, where I is the $N \times N$ identity matrix and $S_\delta(k) = \frac{1}{N}$, $k = 0,\ 1, \ldots, N-1$. We also make the assumption that $\vec{b}$ and $\vec{\delta}$ are uncorrelated.

In previous work, LMMSE estimates of $\vec{\delta}$ have been sought. To find the LMMSE estimate for $\vec{\delta}$ one minimizes with respect to $\vec{\delta}$ the Bayesian mean square error

$$Bmse(\hat{\vec{\delta}}) = E[(\vec{\delta} - \hat{\vec{\delta}})^2], \tag{4}$$

while constraining the estimator to be linear. One can find the LMMSE estimate for the linear observation model in Eq. (3) using the orthogonality principle which gives the following well-known solution [22]:

$$\hat{\vec{\delta}} = (F^t {C_b}^{-1} F + C_\delta^{-1})^{-1} F^t {C_b}^{-1} \vec{g} = (F^t {C_b}^{-1} F + I)^{-1} F^t {C_b}^{-1} \vec{g}, \tag{5}$$

where the superscript t denotes the transpose of a matrix. Note that $\vec{\delta}$ cannot be determined using Eq. (5) without accurate knowledge of C_b.

Because F and C_b are assumed circulant Eq. (5) can be written in the discrete Fourier transform (DFT) domain [1] as

$$\hat{\Delta}(k) = \frac{F^*(k)G(k)}{|F(k)|^2 + NS_b(k)}, \quad k = 0, 1, \ldots, N-1, \tag{6}$$

where $\hat{\Delta}(k)$ is the k^{th} DFT component of $\hat{\vec{\delta}}$, $F(k)$ and $G(k)$ represent the k^{th} DFT components of the template and the observed image, respectively, and $S_b(k)$ is the k^{th} eigenvalue of the covariance matrix of the background, and $*$ denotes conjugation. In the rest of this paper, indexed capital letters represent DFT quantities. Clearly, the computation of $\Delta(k)$ using Eq. (6) requires the knowledge of $S_b(k)$, the power spectrum of the combined background and noise. Filters identical to the LMMSE estimator in Eq. (6) were used in Ref. 11; in Ref. 9 the background was assumed white with variance ${\sigma_b}^2$, i.e., $S_b(k) = {\sigma_b}^2$,$k = 0,\ 1, \ldots, N-1$. But neither Ref. 9 nor Ref. 11 offered a systematic approach for estimating ${\sigma_b}^2$ or $S_b(k)$.

In Ref. 14 the following periodogram estimate for $S_b(k)$ was used for impulse restoration

$$S_b(k) = \frac{|G(k)|^2 - |F(k)|^2}{N}, \quad k = 0, 1, \ldots, N-1. \tag{7}$$

In Ref. 23 it was proven that this is equivalent to the maximum-likelihood (ML) estimate of $S_b(k)$, therefore, we refer to the use of Eqs. and (6) and (7) as the LMMSE-ML method. By definition the power spectrum $S_b(k)$ cannot be negative. However, owing to inconsistencies between the data and the assumed model, mainly due to the assumption that the signal and the background are uncorrelated, it was observed that the periodogram estimator yields a negative value for the power spectrum at many frequencies. These negative values were observed to

degrade the performance of the LMMSE estimate. So the following constrained ML (CML) estimate of $S_b(k)$ was introduced in Ref. 23

$$S_b(k) = \max\{0, \frac{|G(k)|^2 - |F(k)|^2}{N}\}, \quad k = 0, 1, \ldots, N-1. \tag{8}$$

The main difficulty in the impulse-restoration problem arises from the fact that, in most practical applications, the power spectrum $S_b(k)$ of the background in Eq. (6) is unknown and must be estimated. A solution is offered in the following section.

3. MAXIMUM-LIKELIHOOD IMPULSE RESTORATION

The purpose of impulse restoration is to obtain an estimate of $\vec{\delta}$ which is as close as possible to the true impulse $\vec{\delta}$. Restoration filters of the kind described in the previous section require knowledge of the background statistics $S_b(k)$ to accomplish this goal.

In this section we derive a new iterative restoration filter that does not require advance knowledge of $S_b(k)$. This filter estimates a set of unknown parameters which is defined by

$$\theta = \{\vec{\delta},\ C_b\} \tag{9}$$

where $\vec{\delta}$ denotes the impulse and C_b represents the background covariance matrix.

The ML estimator of θ is given by

$$\hat{\theta}_{ML} = \arg\max_{\theta}\ log\ p(\vec{g}|\theta) \tag{10}$$

where $p(\vec{g}|\theta)$ is the likelihood function of $\vec{g}$ given θ. We now assume the background to obey a Gaussian probability density function (PDF) with zero mean and covariance matrix C_b, i.e., $\vec{b} \sim N(0, C_b)$. We further assume that $\vec{\delta} \sim N(0, C_\delta)$. Assuming $\vec{\delta}$ and $\vec{b}$ to be stationary signals, C_b is an $N \times N$ circulant covariance matrix, the eigenvalues of which are the power-spectrum coefficients $S_b(k)$, $k = 0,\ 1,\ \ldots N-1$. Because $\vec{\delta}$ is a delta function, $C_\delta = I$, where I is the $N \times N$ identity matrix. This implies that $S_\delta(k) = \frac{1}{N}$, $k = 0,\ 1,\ \ldots\ N-1$. We also make the assumption that $\vec{b}$ and $\vec{\delta}$ are uncorrelated; therefore, because they are Gaussian, they are also independent.

In this formulation the number of parameters to be estimated is $2N$: N power-spectrum coefficients of $\vec{b}$ and N elements of $\vec{\delta}$. Therefore, the number of parameters far exceeds the number of observations N (the number of pixels in the observed image). In an effort to reduce the number of parameters to be estimated, we model the background as a low-order two-dimensional autoregressive (AR) process with causal support,[1] i.e.,

$$b(n) = b(n) \star \alpha(n) + v(n) \tag{11}$$

where $\alpha(n)$ represents the background AR-model coefficients that minimize $E(v(n)^2)$. The modeling error is a zero-mean Gaussian random process with covariance matrix $C_v = \sigma_v^2 I$. Using a matrix-vector notation and circular convolutions we write Eq. (11) as

$$\vec{b} = A\vec{b} + \vec{v}. \tag{12}$$

Thus,

$$C_b = \sigma_v^2 (I - A)^{-1} (I - A)^{-t} \tag{13}$$

where A is a circulant matrix containing the model parameters $\alpha(n)$ such that the matrix multiplication is equivalent to convolution. Using the AR model for the background, we can rewrite our set of unknown parameters as $\theta = \{\vec{\delta},\ A\}$. Hence the likelihood function becomes

$$p(g|\theta) = N(F\vec{\delta}, F^t F + \sigma_v^2 (I - A)^{-1} (I - A)^{-t}). \tag{14}$$

By substituting Eq. (14) into Eq. (10), the ML estimate of θ is obtained as

$$\hat{\theta}_{ML} = \arg \max_{\theta} \{-log\ (|F^t F + C_b|) - (\vec{g} - F\vec{\delta})^t [F^t F + C_b]^{-1} (\vec{g} - F\vec{\delta})\} \tag{15}$$

where C_b is the covariance matrix of $\vec{b}$ as in Eq. (13). Unfortunately, this is a complicated nonlinear function which cannot be optimized directly. Therefore, we propose to employ the EM algorithm to obtain a solution for Eq. (15). The EM algorithm is a general iterative method to compute ML estimates if the observed data can be regarded as incomplete.[16]

We rewrite the matrix-vector form of the imaging equation (1) as follows:

$$\vec{g} = F\vec{\delta} + \vec{b} = \left[\ F\ ;\ I\ \right] \left[\ \begin{array}{c} \vec{\delta} \\ \vec{b} \end{array}\ \right] = H\vec{z}, \tag{16}$$

where we have defined $\vec{z} = [\vec{\delta}^t,\ \vec{b}^t]^t$ and I is an $N \times N$ identity matrix.

In applying the EM algorithm, one defines a set of complete data and a set of incomplete data.[16] Here, the vector $\vec{z}$ represents the complete data, consisting of the background $\vec{b}$ and the

impulse $\vec{\delta}$; the observation vector $\vec{g}$ represents the incomplete data. As its name suggests, the EM algorithm consists of two steps: the expectation step (E-step) and the maximization step (M-step). In the E-step, one computes the conditional expectation of the likelihood function of the complete data parameterized by the observed data $\vec{g}$ and the current estimate of the relevant parameters. In the M-step this expectation is maximized. The EM algorithm can be expressed as the alternating iterative computation of the following equations:

E-step:

$$Q(\theta;\theta^{(l)}) \;=\; E[ln(p(\vec{z};\theta))|g,\theta^{(l)}] \tag{17}$$

M-step:

$$\theta^{(l+1)} = \arg\max_{\theta} \,[Q(\theta;\theta^{(l)})]. \tag{18}$$

We will assume the complete data and the observation to be stationary Gaussian random signals. The Gaussian assumption results in linear equations for the E-step and the M-step. Moreover, the stationarity that we assume yields circulant covariance matrices. Using the diagonalization properties of the DFT for circulant matrices the EM algorithm can be represented in the DFT domain as a set of scalar equations.[17]

By evaluating the conditional expectation of the complete-data likelihood function as shown in Appendix A, The E-step can be written, for $k = 0,\ 1,\ \ldots N-1$, as

$$M_{\delta|g}^{(l)}(k) = \frac{F^*(k)G(k)}{|F(k)|^2 + N\tilde{S}_b^{(l)}(k)} \tag{19}$$

$$M_{b|g}^{(l)}(k) = \frac{N\tilde{S}_b^{(l)}(k)G(k)}{|F(k)|^2 + N\tilde{S}_b^{(l)}(k)} \tag{20}$$

$$S_{b|g}^{(l)}(k) = \frac{|F(k)|^2\tilde{S}_b^{(l)}(k)}{|F(k)|^2 + N\tilde{S}_b^{(l)}(k)}, \tag{21}$$

where $F(k)$ and $G(k)$ represent the k^{th} DFT components of the template and the observed image, respectively; $M_{\delta|g}^{(l)}(k)$ and $M_{b|g}^{(l)}(k)$ represent the k^{th} DFT components of the conditional mean of the impulse and the background, respectively; $S_{b|g}^{(l)}(k)$ is the k^{th} eigenvalue of the conditional covariance matrix of the background parameterized by the observation; and $\tilde{S}_b^{(l)}(k)$ is the k^{th} eigenvalue of the autocorrelation matrix of the background based on the AR model. The index l is the iteration index of the algorithm.

In Appendix A we show that the M-step is obtained by taking the partial derivative of the expectation of the likelihood function with respect to the background AR-model parameters which are contained in the matrix A and setting it to zero. This leads to the set of normal equations

$$\bar{R}_b^{(l)}\vec{\alpha}^{(l+1)} = \bar{R}_b^{(l)}\vec{1} = \vec{r}_b^{(l)}. \tag{22}$$

where $\bar{R}_b^{(l)}$ is an $N \times M$ matrix containing the first M columns of the $N \times N$ $(N > M)$ conditional autocorrelation matrix of the background given the observation $R_b^{(l)}$, where M is the order of the AR model and $\vec{1} = [1\ ,0\ \ldots,0]^t$. Thus, the normal equations can be written as

$$\vec{r}_b^{(l)} = \bar{R}_b^{(l)}\vec{\alpha}^{(l+1)} \tag{23}$$

where $\vec{r}_b^{(l)}$ is an $N \times 1$ vector representing the inverse DFT of $S_b^{(l)}(k)$ and $\vec{\alpha} = [\alpha(1), \alpha(2), \ldots, \alpha(M)]$. In Appendix A we show that the DFT of $\vec{r}_b^{(l)}$ is found by

$$S_b^{(l)}(k) = S_{b|g}^{(l)}(k) + \frac{1}{N}|M_{b|g}^{(l)}(k)|^2, \ k = 0, \ 1, \ \ldots \ N-1. \tag{24}$$

Consequently $S_b^{(l)}(k)$ is the k^{th} eigenvalue of the circulant matrix $R_b^{(l)}$.

Taking Eq. (13) to the DFT domain the AR-modeled power spectrum of the background $\vec{b}$ is given by

$$\tilde{S}_b^{(l)}(k) = \frac{(\sigma_v^{(l)})^2}{|1 - A^{(l)}(k)|^2}, \quad k = 0, 1, \ldots, N-1. \tag{25}$$

4. LOCALIZATION RECEIVER OPERATING CHARACTERISTICS CURVE

The receiver operating characteristics (ROC) curve is a comprehensive way to describe the detection performance of a human or machine observer. The ROC curve is a plot of the probability of correct detection P_D versus the probability of false alarm P_F for the continuum of possible decision thresholds. Thus, it summarizes the range of trade-offs between missed detections and false alarms. In comparing two detection algorithms, the one having a strictly higher ROC curve can be said to be superior in the sense that for any specified false alarm probability, the detection probability is higher (likewise, for a specified detection probability, the false alarm rate is lower).

The limitation of the ROC curve in characterizing an object-recognition algorithm is that it only captures the ability of the algorithm to decide correctly whether an object is present somewhere in a given scene. Thus, it fails to account for errors in localization which can be quite significant. For example, if the algorithm mistakenly identifies a non-target object as a target, this is treated by the ROC curve as a correct decision as long as a target happens to be present somewhere else in the scene.

The localization ROC (LROC) curve remedies this shortcoming by taking into account localization performance as well as detection performance. An LROC curve is a plot of the probability of detection *and* correct localization versus the probability of false alarm. Thus, a decision is said to be correct only if the object is both detected and located correctly by the object-recognition algorithm.

The experimental method we used to compute LROC curves is described in the following section. For a detailed description of recent results and extensions of the LROC curve see Ref. 19.

5. EXPERIMENTAL RESULTS

In this section we describe numerical experiments that compare the proposed EM-based impulse-restoration filter with previous implementations of the LMMSE impulse-restoration filter and different types of MFs. Monte-Carlo studies were performed to evaluate the performance of the

different algorithms. Multiple synthetic images were generated by randomly varying the object locations and noise realizations.

We compared the phase only matched filter POMF, [3] given by

$$\hat{\Delta}(k) = \frac{F^*(k)G(k)}{|F(k)|}, \ k = 0, 1, \ldots, N-1, \tag{26}$$

the symmetric phase only matched filter SPOMF,[2] given by

$$\hat{\Delta}(k) = \frac{F^*(k)G(k)}{|F(k)||G(k)|}, \ k = 0, 1, \ldots, N-1. \tag{27}$$

and previous LMMSE filters given in Eq. (6) that use the ML and the CML estimates of the background statistics defined in Eqs. (7) and (8), respectively.

We also evaluated the performance of the MF variants described in Ref. 8. The first approach in Ref. 8 models the backgound as

$$b(n) = g(n) - f(n - n_0), \tag{28}$$

where $b(n)$ represents the backgound, $g(n)$ represents the observation image, and $f(n - n_0)$ is the template located at the location of the target in the observation. Since the location of the target is not normally known *a priori*, a uniform distribution of the target location is assumed. Using this assumption, the following form for the power spectrum of the background is obtained:

$$S_b(k) = |G(k)|^2 + |F(k)|^2. \tag{29}$$

We will refer to this filter in the experiments as MF-Model(1). Notice that this is identical to an LMMSE filter in which the power spectrum of the observation is used as an approximation

of the background power spectrum. The second MF approach in Ref. 8 models the background as

$$b(n) = w(n - n_0)g(n), \tag{30}$$

where $w(n - n_0)$ is a window function centered at the location of the object that has a value of 0 within the template and 1 elsewhere. By again by assuming a uniform distribution of the target location the background power spectrum ia approximated as

$$S_b(k) = |G(k)|^2 \star |W(k)|^2, \tag{31}$$

where $\star$ denotes convolution and $W(k)$ represents the k^{th} DFT component of the window function. We will refer to this filter as MF-Model(2).

The third MF approach in Ref. 8 uses the power spectrum of the observation as as estimate for the background power spectrum. This approach is based on the assumption that the size of the template is much smaller than the scene. This matched filter turns out to be identical in form to the LMMSE-ML filter defined in Sect. 2.B.

To establish the upper performance bound of the impulse-restoration template-matching approach, we implemented an exact LMMSE filter which assumes perfect knowledge of the noise and background statistics was also used. Of course, such information is never available in practice. The exact LMMSE filter serves only as a basis for judging the other more realistic approaches. The true statistics were only provided to the exact LMMSE filter. For purposes of implementing the exact LMMSE filter the actual background was determined based on the

imaging model of Eq. (3) by placing zeros at the location of the object (see, for example, Fig. 1). The sum of the periodogram power-spectrum estimate of the true background, plus the known variance of added white noise, was used as the background statistics for the implementation of the exact LMMSE filter in Eq. (6).

As stated before, the EM algorithm uses an AR model for the background. A 2×2 AR image model with causal support as shown in Fig. 2 was used to model the background. In our experience it was observed that the performance of the EM algorithm is not very sensitive to the order of the AR model in this application.

A. Experiment 1: LROC Analysis

In this experiment, LROC analysis was used to quantify the performance of various algorithms in detecting and locating objects within a scene. Test scenes of dimension 256×256 were generated by embedding the three objects shown in Fig. 3 at random locations within a background terrain image. These scenes were then degraded by blur and white Gaussian noise to represent the effects of the imaging system. The scenes were blurred with a Gaussian-shaped point spread function with full width at half maximum (FWHM) of 3 pixels. White Gaussian noise with $\sigma = 20$ was used, where σ represents the standard deviation. Examples of the noisy blurred scenes used in the experiment are shown in Fig. 4.

In each experimental trial, an object-recognition filter was applied to the generated scene.

The peak of the output was taken to be the restored impulse, and the value at the peak was used as a decision variable x. If x exceeded a decision threshold T, then the object was said to be present at the location of the peak; otherwise, the object was said to be absent. The performance of the algorithm for various thresholds is summarized by the LROC curve which plots the probability of detection *and* correct localization (P_{DL}) versus the probability of false alarm (P_F). These probabilities were obtained by numerical evaluation of the following integrals

$$P_{DL} = \int_T^\infty p(x|H_1)dx$$
$$P_F = \int_T^\infty p(x|H_0)dx, \tag{32}$$

where $p(x|H_j)$ is the conditional probability density function (PDF) of x given hypothesis H_j. The null hypothesis H_0 is that the object is absent; the alternative hypothesis H_1 is that the object is present.

In this study the conditional PDFs $p(x|H_1)$ and $p(x|H_0)$ were obtained using the following procedure. Two sets of 100 images were generated each with a different noise realization and with different random object locations. In one set of images, all three objects were present in each scene (see, for example, Fig. 4); in the other set, the object of interest was absent (see, for example, Fig. 5). In the case where the object of interest was present, each algorithm was applied to every image, and the magnitude x of the peak value of the restored impulse was recorded, provided that the peak correctly indicated the target location. The normalized histogram of these values was then used as $p(x|H_1)$. To find $p(x|H_0)$ the set of images was

used in which the object of interest was absent. Again, the algorithms were applied to each image, but the magnitude of the peak value of the restored impulse was recorded regardless of its location. The normalized histogram of the recorded peak values of the restored impulse obtained was used as $p(x|H_0)$. Different points of the LROC curves were obtained by varying the decision threshold T and computing the integrals in Eq. (32).

In Fig. 6 we show the conditional PDFs $p(x|H_0)$ and $p(x|H_1)$ for the exact LMMSE, EM-AR, MF-Model(1), MF-Model(2), LMMSE-CML, SPOMF, LMMSE-ML, and POMF filters for the case where the tank was the object of interest. These figures illustrate that good discriminating capability results from non-overlapping $p(x|H_0)$ and $p(x|H_1)$. For example the hypotheses are almost perfectly distinguished by the exact LMMSE if the decision threshold T is 175.

The LROC curves are shown in Fig. 7. The LROC curves in Fig.7 indicate that the proposed EM-based approach with AR modeling performes best overall among the template matching filters tested. Other methods worked better in some cases, but were inconsistent and so did not perform as well overall.

B. Experiment 2: Robustness Comparison

In this experiment, we compare the robustness of the proposed algorithms as the noise increases using the 256×256 Lena image and the 21×21 template, both shown in Fig. 8. White additive noise was used to degrade the scene. Twenty different noise levels, with noise standard deviation

ranging from 10 to 100 in increments of 10, were used to degrade the scene (see, for example, Fig. 9). For each noise level the experiment was repeated 50 times with different noise realizations. The probability P_{DL} was found as $P_{DL} = \frac{N}{50}$ where N is the number of times that the maximum of the output of the restored delta was positioned correctly at the object location.

In Fig. 10 plots of P_{DL} versus the noise standard deviation are shown for the exact LMMSE, EM-AR, MF-Model(1), MF-Model(2), LMMSE-CML, SPOMF, LMMSE-ML, and POMF algorithms. This plot shows that the proposed EM-AR filter is the best among all filters that do not require prior knowledge of the location of the object. In Fig. 11 the restored impulses for the exact LMMSE, EM-AR, MF-Model(1), MF-Model(2), LMMSE-CML, SPOMF, LMMSE-ML, and POMF algorithms are shown for a specific noise realization when the standard deviation is 30. Notice that the proposed EM-AR filter gives the closest approximation to an ideal delta function among all filters that do not require prior knowledge of the location of the object. A mean square error metric MSE for the restored impulse was used to compare different algorithms. MSE is defined by

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{\delta}(i) - \delta(n_0))^2 \tag{33}$$

where $\hat{\delta}(i)$ is the normalized restored impulse as shown in Fig. 11 and n_0 represents the location of the object. In Table 1, the mean square error (MSE) values for the restored impulses are tabulated for different filters for the case of $\sigma = 30$.These values represent the average MSE over all the 50 cases for $\sigma = 30$. The table shows that the EM-AR filter gives the lowest average

MSE among all filters that do not require prior knowledge of the location of the object.

6. CONCLUSIONS AND FUTURE WORK

In this paper restoration-based template-matching filters were used to detect objects within a noisy, blurred scene. The main challenge involved in the impulse-restoration approach is good estimation of the background statistics. To address this, we propose a new iterative approach based on the EM algorithm that simultaneously estimates the background statistics and restores a delta function at the location of the template. Our algorithm models the background as an AR process with causal support.

Our numerical experiments verified the importance of accurately estimating the background statistics to the success of the template-matching impulse-restoration approach. The proposed EM-AR approach was superior to the variants of the matched filter that we tested. When the background statistics are not estimated accurately, impulse restoration-based template matching cannot realize its full potential. Our numerical experiments also show that the advantage of the proposed approach over previous methods increases with the strength of the noise in the scene.

We are currently studying application of this approach to photon-limited images and to the more difficult problems of object rotation and scaling.

7. APPENDIX A: THE EXPECTATION MAXIMIZATION ALGORITHM

Let us assume the complete-data set $\vec{z}$ to obey the following Gaussian likelihood function:

$$p(\vec{z};\theta) = \mid 2\pi C_z \mid^{-\frac{1}{2}} exp[\ -\frac{1}{2}\vec{z}\ ^t {C_z}^{-1}\vec{z}\ ], \qquad (A-1)$$

in which θ is the set of the unknown parameters in Eq. (9).

Taking the logarithm of Eq. (A-1) we get

$$ln(p(\vec{z};\theta)) = -\frac{1}{2}[\ ln(\mid 2\pi C_z \mid) + \vec{z}\ ^t {C_z}^{-1}\vec{z}\ ] = K - \frac{1}{2}\{\ ln(\mid C_z \mid) + \vec{z}\ ^t C_z^{-1}\vec{z}\ \}. \qquad (A-2)$$

Then, the expected value of the logarithm of the conditional PDF required by the E-step of the EM algorithm in Eq. (17) is given by

$$Q(\theta;\theta^{(l)}) = K - \frac{1}{2}\{\ E[\ ln(\mid C_z \mid)\ /\ \vec{g}\ ;\ \theta^{(l)}\ ] + E[\ \vec{g}\ ^t C_z^{-1}\vec{z}\ /\ \vec{g}\ ;\ \theta^{(l)}\ ] = K - Y(\theta;\theta^{(l)}). \qquad (A-3)$$

It is easy to show that

$$Y(\theta;\theta^{(l)}) = ln(\mid C_z \mid) + trace[\ C_z^{-1} {R_{z|g}}^{(l)}\ ], \qquad (A-4)$$

with

$${R_{z|g}}^{(l)} = E[\ \vec{z}\vec{z}\ ^t\ /\ \vec{g}\ ;\ \theta^{(l)}\ ] = {C_{z|g}}^{(l)} + \vec{\mu}_{z|g}^{(l)}\ (\vec{\mu}_{z|g}^{(l)})^t. \qquad (A-5)$$

Thus, we can write

$$Y(\theta;\theta^{(l)}) = ln(\ \mid C_z \mid\ ) + trace[\ C_z^{-1} C_{z|g}^{(l)}\ ] + (\vec{\mu}_{z|g}^{(l)})^t {C_z}^{-1}\vec{\mu}_{z|g}^{(l)}. \qquad (A-6)$$

Because of the definition of $\vec{z}$ in Eq. (16) and since $\vec{\delta}$ and $\vec{b}$ are assumed uncorrelated we get

$$C_z = \begin{bmatrix} I\ ,\ 0 \\ 0\ ,\ C_b \end{bmatrix},\quad C_{z/g} = \begin{bmatrix} C_{\delta|g}\ ,\ 0 \\ 0\ ,\ C_{b|g} \end{bmatrix},\quad \text{and} \quad \vec{\mu}_{z|g} = [\ (\vec{\mu}_{\delta|g})^t,\ (\vec{\mu}_{b|g})^t\ ]^t. \tag{A-7}$$

Using the definition of C_b in Eq. (13) we can write Eq. (A-6) as

$$\begin{aligned} Y(\theta;\theta^{(l)}) = \quad & ln|\ \sigma_v^2(I-A)^{-1}(I-A)^{-t}\ | + trace\{C_{\delta|g}^{(l)} + \frac{1}{\sigma_v^2}(I-A)^t(I-A)C_{b|g}^{(l)}\} \\ & +(\vec{\mu}_{\delta|g}^{(l)})^t\vec{\mu}_{\delta|g}^{(l)} + \frac{1}{\sigma_v^2}(\vec{\mu}_{b|g}^{(l)})^t(I-A)^t(I-A)\vec{\mu}_{b|g}^{(l)}. \end{aligned} \tag{A-8}$$

Since we assume an AR model with causal support, A is a strictly upper-triangular matrix. Hence,

$$ln|\ \sigma_v^2(I-A)^{-1}(I-A)^{-t}\ |\ =\ ln1\ = 0 \tag{A-9}$$

Thus, we can write

$$Y(\theta;\theta^{(l)}) = trace\{C_{\delta|g} + \frac{1}{\sigma_v^2}(I-A)^t[C_{b|g}^{(l)} + \vec{\mu}_{b|g}^{(l)}(\vec{\mu}_{b|g}^{(l)})^t](I-A)\} + (\vec{\mu}_{\delta|g}^{(l)})^t\vec{\mu}_{\delta|g}^{(l)}. \tag{A-10}$$

A is a circulant matrix containing the model parameters, and $C_{b|g}^{(l)}$ is a circulant matrix since we assume stationarity. For circulant matrices it is easy to show that

$$trace\{X^tAX\} = N(\vec{x}^tA\vec{x}), \tag{A-11}$$

and

$$trace\{X\vec{a}\vec{a}^tX^t\} = trace\{A\vec{x}\vec{x}^tA\} = \vec{a}^tX^tX\vec{a}, \tag{A-12}$$

where X and A are $N \times N$ circulant matrices and $\vec{x}$ and $\vec{a}$ are $N \times 1$ vectors representing the first columns of X and A, respectively. Hence we obtain

$$Y(\theta;\theta^{(l)}) = trace\{C_{\delta|g} + \frac{1}{\sigma_v^2}(\vec{1} - \vec{\alpha'})^t[N \cdot C_{b|g}^{(l)} + M_{b|g}^{(l)}(M_{b|g}^{(l)})^t](\vec{1} - \vec{\alpha'})\} + (\vec{\mu}_{\delta|g}^{(l)})^t\vec{\mu}_{\delta|g}^{(l)}. \tag{A-13}$$

in which $\vec{1} = [1\ 0 \ldots 0]^t$; $\vec{\alpha'} = [\alpha(1) \ldots \alpha(M), 0, 0 \ldots 0]^t$, where M is the order of the AR model and $\alpha(k)$ is the k^{th} parameter of the AR model; and $M_{b|g}$ is an $N \times N$ circulant matrix constructed by circular shifting of $\vec{\mu}_{b|g}$.

Letting

$$R_b^{(l)} = C_{b|g}^{(l)} + \frac{1}{N} M_{b|g}^{(l)} (M_{b|g}^{(l)})^t, \tag{A-14}$$

we can write

$$Y(\theta; \theta^{(l)}) = trace\{C_{\delta|g} + \frac{N}{\sigma_v^2}(\vec{1} - \vec{\alpha'})^t R_b^{(l)} (\vec{1} - \vec{\alpha'})\} + (\vec{\mu}_{\delta|g}^{(l)})^t \vec{\mu}_{\delta|g}^{(l)}. \tag{A-15}$$

Using the diagonalization properties of the DFT for circulant matrices, from Eq. (A-14) we can write

$$S_b^{(l)}(k) = S_{b|g}^{(l)}(k) + \frac{1}{N} |M_{b|g}^{(l)}(k)|^2 \text{for, } k = 0,\ 1,\ \ldots\ N-1. \tag{A-16}$$

where $S_b^{(l)}(k)$ is the k^{th} eigenvalue of $R_b^{(l)}$; $S_{b|g}^{(l)}(k)$ is the k^{th} eigenvalue of the covariance matrix of the background parameterized by the observation obtained by Eq. (21); and $M_{b|g}^{(l)}(k)$ represents the k^{th} eigenvalue of the matrix $M_{b|g}^{(l)}$, found as the k^{th} DFT component of the vector $\vec{\mu}_{\delta|g}$.

A. The M-step

Taking the partial derivative of $Y(\theta; \theta^{(l)})$ in Eq. (A-15) with respect to $\vec{\alpha}$ and setting it to zero we obtain

$$R_b^{(l)} \vec{\alpha'}^{(l+1)} = R_b^{(l)} \vec{1} = \vec{r}_b^{(l)}. \tag{A-17}$$

So the AR parameters are found by solving a set of normal equations. Since $\vec{\alpha'}$ contains $M < N$ non zero unknown coefficients we calculate the least squares estimate of $\vec{\alpha}^{(l+1)}$

$$\vec{\alpha}^{(l+1)} = [(\bar{R}_b^{(l)})^t \bar{R}_b^{(l)}]^{-1} (\bar{R}_b^{(l)})^t \vec{r}_b^{(l)} \qquad \text{(A-18)}$$

where $\bar{R}_b^{(l)}$ is an $N \times M$ matrix containing the first M columns of $R_b^{(l)}$.

From Eq. (13), the update equation of the power spectrum of the background $\vec{b}$ under the AR model is then given by

$$\tilde{S}_b{}^{(l)}(k) = \frac{(\sigma_v{}^{(l)})^2}{|1 - A^{(l)}(k)|^2}, \qquad k = 0, 1, \ldots, N-1. \qquad \text{(A-19)}$$

where σ_v is the standard deviation of the driving noise of the AR model. It is calculated from $\vec{r}_b$ and $\vec{\alpha}$ [24], according to

$$(\sigma_v^{(l)})^2 = r_b^{(l)}(0) - \sum_{k=1}^{M} \alpha^{(l)}(k) r_b^{(l)}(k), \qquad \text{(A-20)}$$

and $A^{(l)}(k)$ represents the k^{th} DFT component of $\vec{\alpha'}^{(l)}$, where $\vec{\alpha'}^{(l)}$ is a version of $\vec{\alpha}^{(l)}$ zero-padded to length N.

B. The E-step

When the vectors $\vec{z}$ and $\vec{g}$ are related by Eq. (16) it can be shown [22] that the the conditional mean is given by

$$\vec{\mu}_{z|g} = C_z H^t (H C_z H^t)^{-1} \vec{g} = \begin{bmatrix} F^t \\ C_b \end{bmatrix} (F F^t + C_b)^{-1} \vec{g} = \begin{bmatrix} \vec{\mu}_{\delta|g} \\ \vec{\mu}_{b|g} \end{bmatrix}. \qquad \text{(A-21)}$$

In the DFT domain Eq. (A-21) gives, for $k = 0,\ 1,\ \ldots N-1$,

$$M_{\delta|g}^{(l)}(k) = \frac{F^*(k)G(k)}{|F(k)|^2 + NS_b^{(l)}(k)}, \qquad (A-22)$$

$$M_{b|g}^{(l)}(k) = \frac{NS_b^{(l)}(k)G(k)}{|F(k)|^2 + NS_b^{(l)}(k)}. \qquad (A-23)$$

For Gaussian $\vec{z}$ and $\vec{g}$ with $\vec{g} = H\vec{z}$ it can be shown [22] that the conditional covariance

$$C_{z/g} = \begin{bmatrix} C_{\delta/g}\ ,\ 0 \\ 0\ ,\ C_{b/g} \end{bmatrix} = C_z - C_z H^t (HC_z H^t)^{-1} HC_z. \qquad (A-24)$$

Thus,

$$C_{b/g} = C_b - C_b(FF^t + C_b)^{-1}C_b. \qquad (A-25)$$

In the DFT domain Equation (A-25) becomes

$$S_{b|g}(k) = S_b(k) - \frac{S_b^2(k)}{\frac{1}{N}|F(k)|^2 + S_b(k)}. \qquad (A-26)$$

Thus, from Eq. (A-26) we obtain

$$S_{b|g}^{(l)}(k) = \frac{|F(k)|^2 S_b^{(l)}(k)}{|F(k)|^2 + NS_b^{(l)}(k)} \text{ for } k = 0,\ 1,\ \ldots N-1. \qquad (A-27)$$

In this EM formulation the conditional mean $\vec{\mu}_{\delta|g}$ is equivalent to the LMMSE estimate for $\vec{\delta}$ in Eq. (6). Since the AR model is assumed for the background statistics, $S_b^{(l)}(k)$ in Eqs. (A-22), (A-23) and (A-26) is replaced by $\tilde{S}_b^{(l)}(k)$ which yields Eqs. (19), (20) and (21), respectively.

References

[1] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice Hall 1989.

[2] Q. Chen, M. Defrise and F. Decorninck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Trans. on Pattern Recognition and Machine Intelligence,* **12**(12), 1156-1198, (1994).

[3] J. L. Horner and P. D. Gianino, "Phase-only matched filtering", *Applied Optics*, **23**(6), 812-816, (1984).

[4] J. L. Horner and R. L. Leger, "Pattern recognition with binary phase only matched filters", *Applied Optics*, **23**(6), 812-816, (1985).

[5] F. M. Dickey and B. D. Hansche, "Quad-phase correlation filters for pattern recognition," *Applied Optics*, **28**, 1611-1613, (1989).

[6] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *IEEE Proceedings* **69**(5), 529-541, (1981).

[7] O. K. Ersoy and M Zeng, "Nonlinear matched filtering", *Journal of the Optical Society of America-A,* **6**(5), 636-648, (1989).

[8] Leonid P. Yaroslavsky, and Emanuel Marom, *Nonlinearity optimization in nonlinear joint transform correlators.* App. Opt. **36**, 4816-4822 (1997).

[9] J. Ben-Arie, and K. R. Rao, "A novel approach to template matching by nonorthogonal image expansion," *IEEE Trans. on Circuits and Systems for Video Technology,* **3**(1), 71-84, (1993).

[10] C. R. Chatwin R. K. Wang, R. C. D. Young, "Assessment of A Wiener Filter - Synthetic Discriminant Function for Optical Correlation, " *Journal of Optics and Lasers in Engineering,* **22**(1),33-51, (1995).

[11] E. Marom and H. Inbar, "New interpretations of Wiener filters for image recognition," *Journal of the Optical Society of America-A,* **13**(7), 1325-1330, (1996).

[12] B. Javidi, F. Parchekani, and G. Zhang, "Minimum-mean-square error filters for detecting a noisy target in background noise," *Applied Optics*, **35**,6964-6975, (1996).

[13] M. Choi, N. Galatsanos, and D. Schonfeld, "Image restoration-based template-matching with application to motion estimation," R. Ansari, and M. Smith, ed., *Proc. SPIE, Visual Communications Image Processing-96,* SPIE-**2727**, 375-386, (1996).

[14] M. Choi, N. Galatsanos, and D. Schonfeld, "On the relation of image restoration and template-matching: application to block-matching motion estimation," *Proceeding IEEE International Conference on Acoustics Speech and Signal Processing 1996*, **IV**, 2112-2115, (1996).

[15] M. Choi, N. Galatsanos and D. Schonfeld, "Image restoration based template-matching for multi-channel Restoration of Image Sequences", *Proceedings of 1996 ASILOMAR conference*, Pacific Grove, November 1996.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data", *J. Roy. Statist. Soc. B*, **39**,1-38, (1977).

[17] A. K. Katsaggelos and K.-T.Lay *Digital Image Restoration*, Springer-Verlag Berlin Heidelberg, 1991.

[18] B. V. Kumar and L. Hassebrook, "Performance measures for correlation filters," *Applied Optics*, **29**(20), 2997-3006, (1990).

[19] R. Swensson, "Unified measurement of observer performance and localizing target objects on images, " *Medical Physics*, **23**, 1709-1725, (1996).

[20] H. L. Van Trees, *Detection Estimation, and Modulation Theory: Part-I.* New York: Wiley 1968.

[21] H. Andrews and B. Hunt, *Digital Image Restoration.* Englewood Cliffs, New Jersey: Prentice-Hall, 1977.

[22] S. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory.* Englewood Cliffs, New Jersey: Prentice Hall, 1993.

[23] A. Abu-Naser, "Object recognition based on impulse restoration for images in gaussian noise", Master's thesis, ECE Dept., Ill. Inst. of Tech., Chicago, Dec. 1996.

[24] John G. Proakis, and Dimitris G. Manolakis, *Digital Signal Processing, Principles, Algorithms, and Applications.* Englewood cliffs, New Jersey: Prentice Hall, 1996.

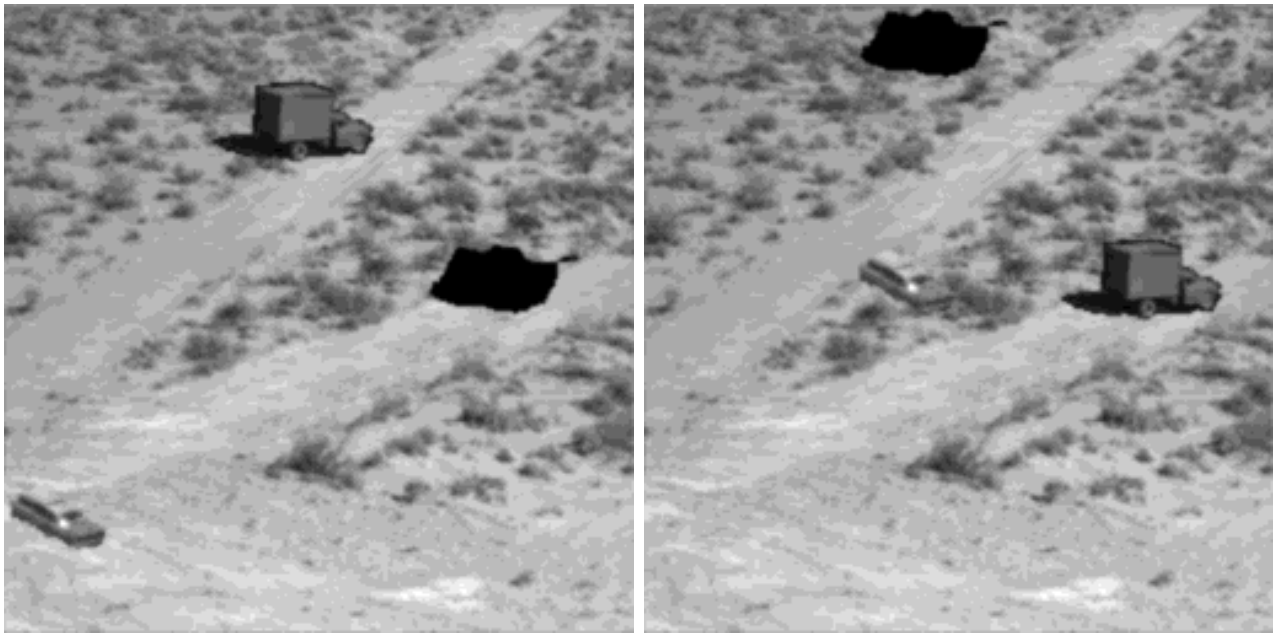

Figure 1: Examples of the backgrounds used by the exact LMMSE filter. The object of interest is the tank. The exact LMMSE filter is unrealistic. It was studied only to determine a theoritical upper bound on the performance.

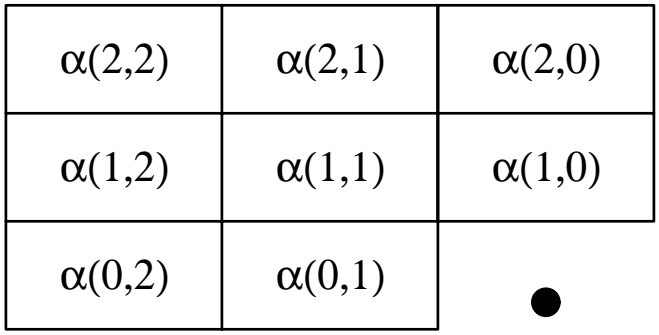

Figure 2: 2×2 autoregressive image model with causal support.

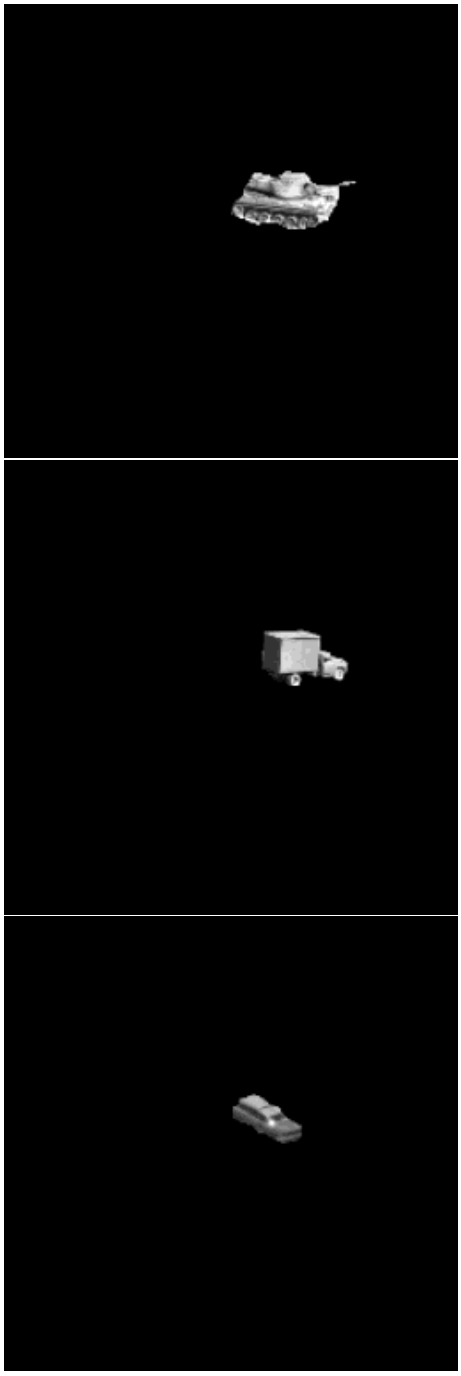

Figure 3: Object templates.

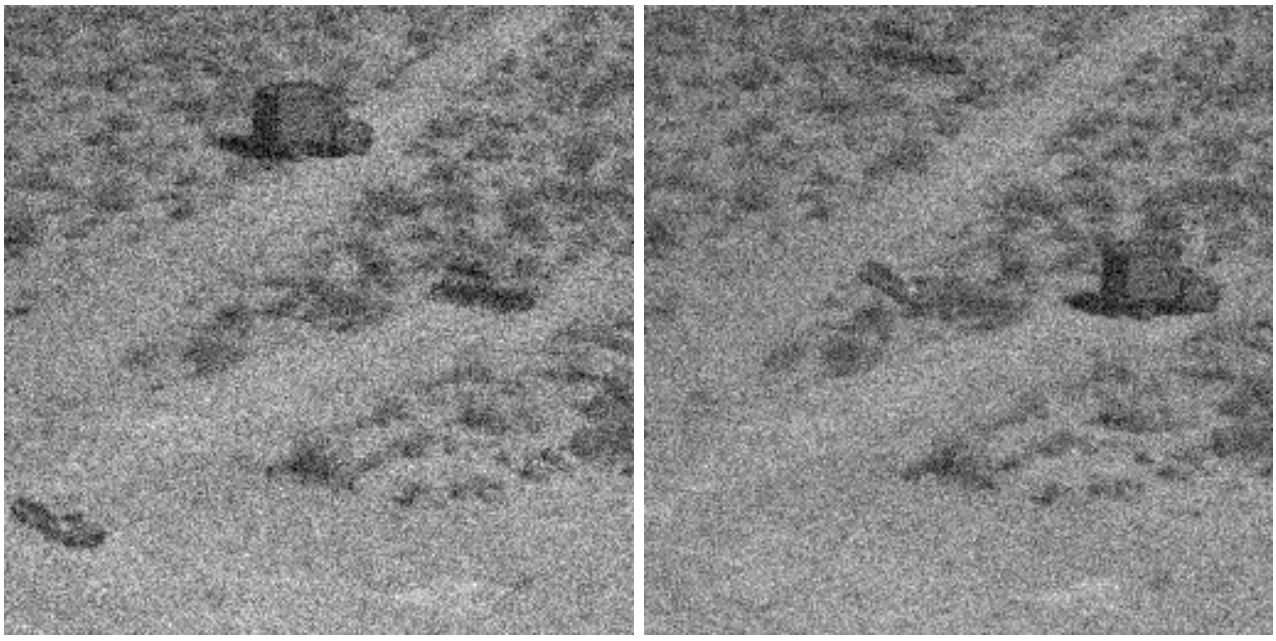

Figure 4: Examples of noisy test scenes with the object of interest present (in this case, tank). The noise level is $\sigma = 20$. In each realization, the objects are placed randomly within the scene.

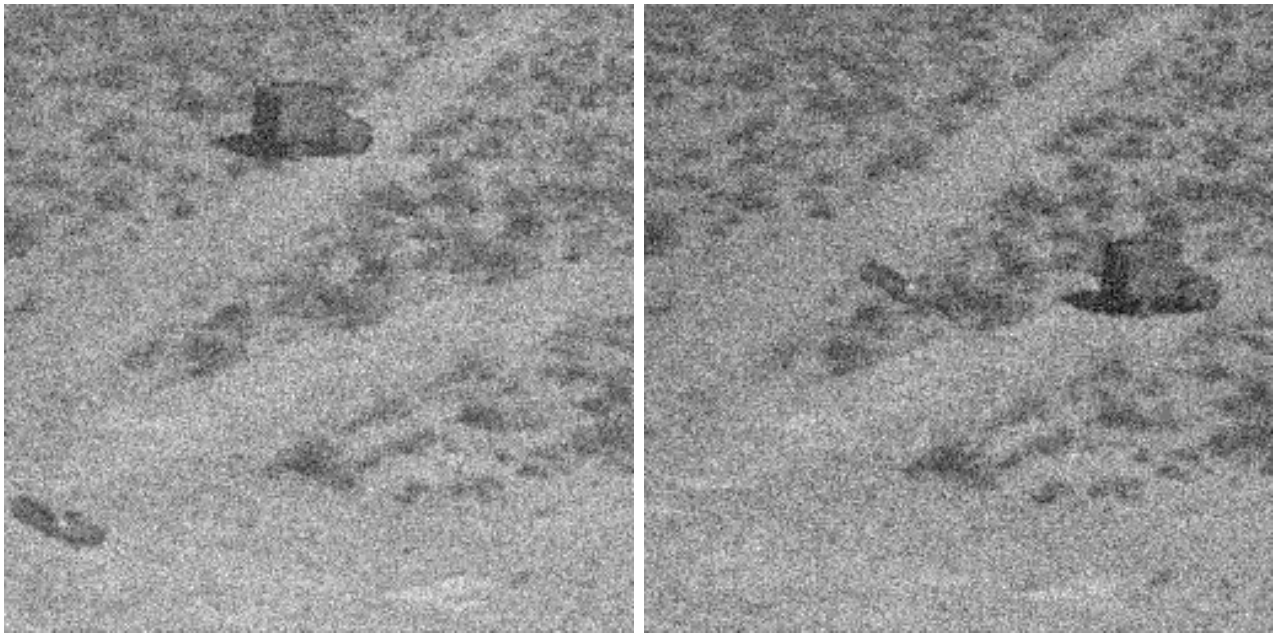

Figure 5: Examples of noisy scenes with the object of interest absent (in this case, the tank). The noise level is $\sigma = 20$.

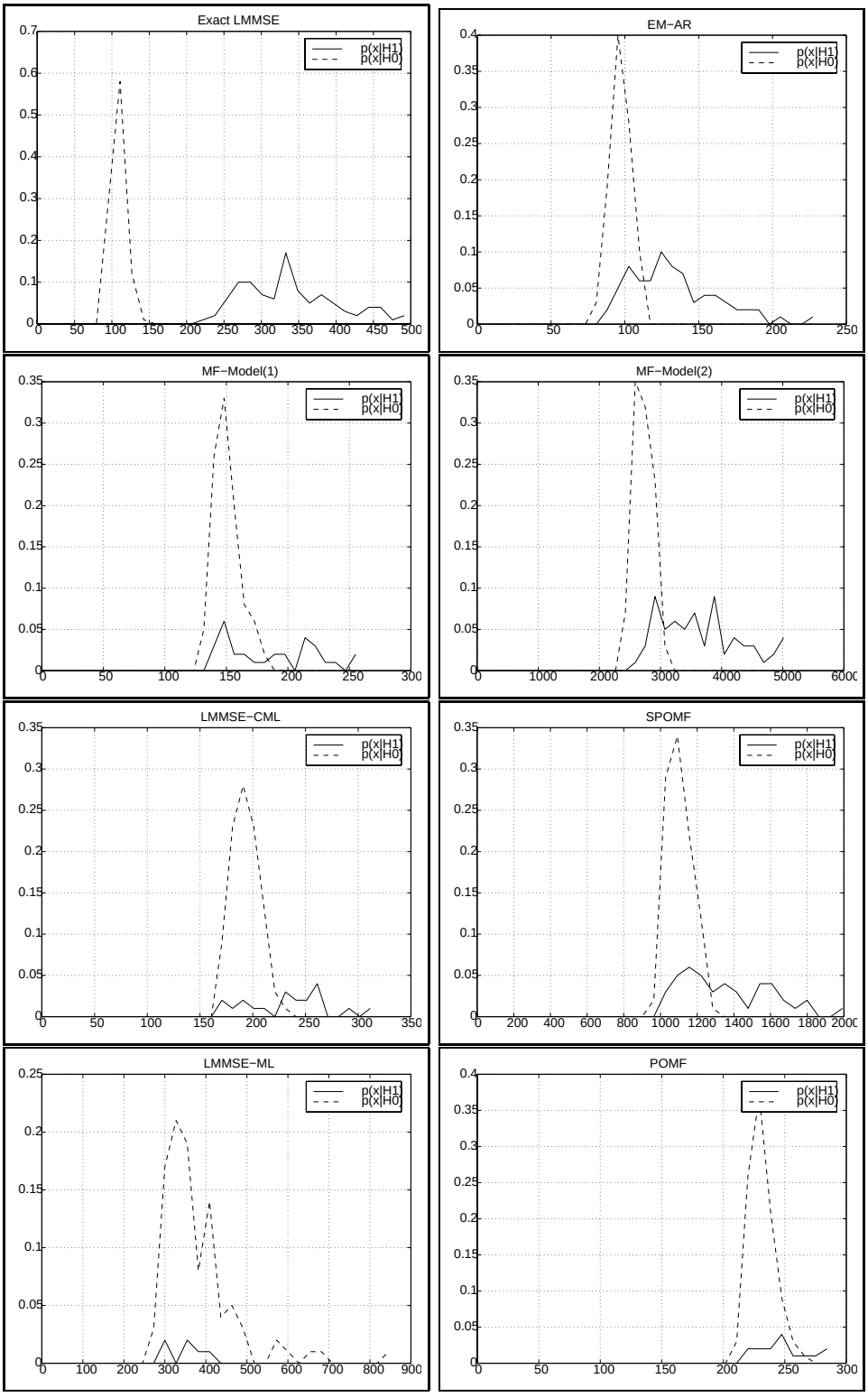

Figure 6: The conditional PDFs for $\sigma = 20$ when the tank is the object of interest.

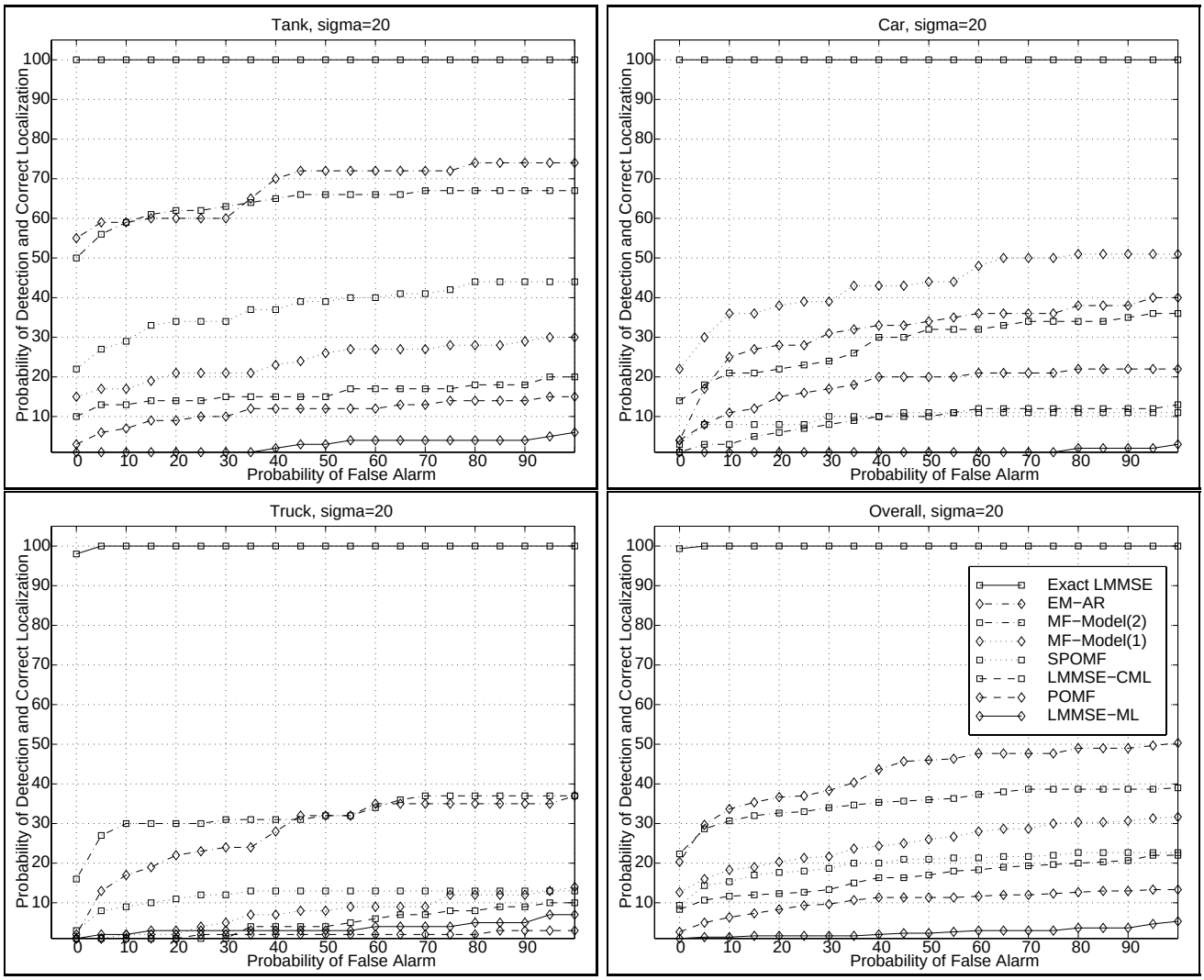

Figure 7: LROC curves for $\sigma = 20$. Overall, the EM-AR algorithm outperforms all but the exact (unrealistic) LMMSE filter which makes use of the unknown object location. Other methods did outperform the EM-AR algorithm in some specific situations, but were inconsistent and did not perform as well overall.

Figure 8: Lena scene and the template.

Figure 9: Examples of Noisy Lena Scenes for Different Noise Levels. Clockwise from top left, $\sigma = 5$, $\sigma = 30$, $\sigma = 50$, and $\sigma = 100$.

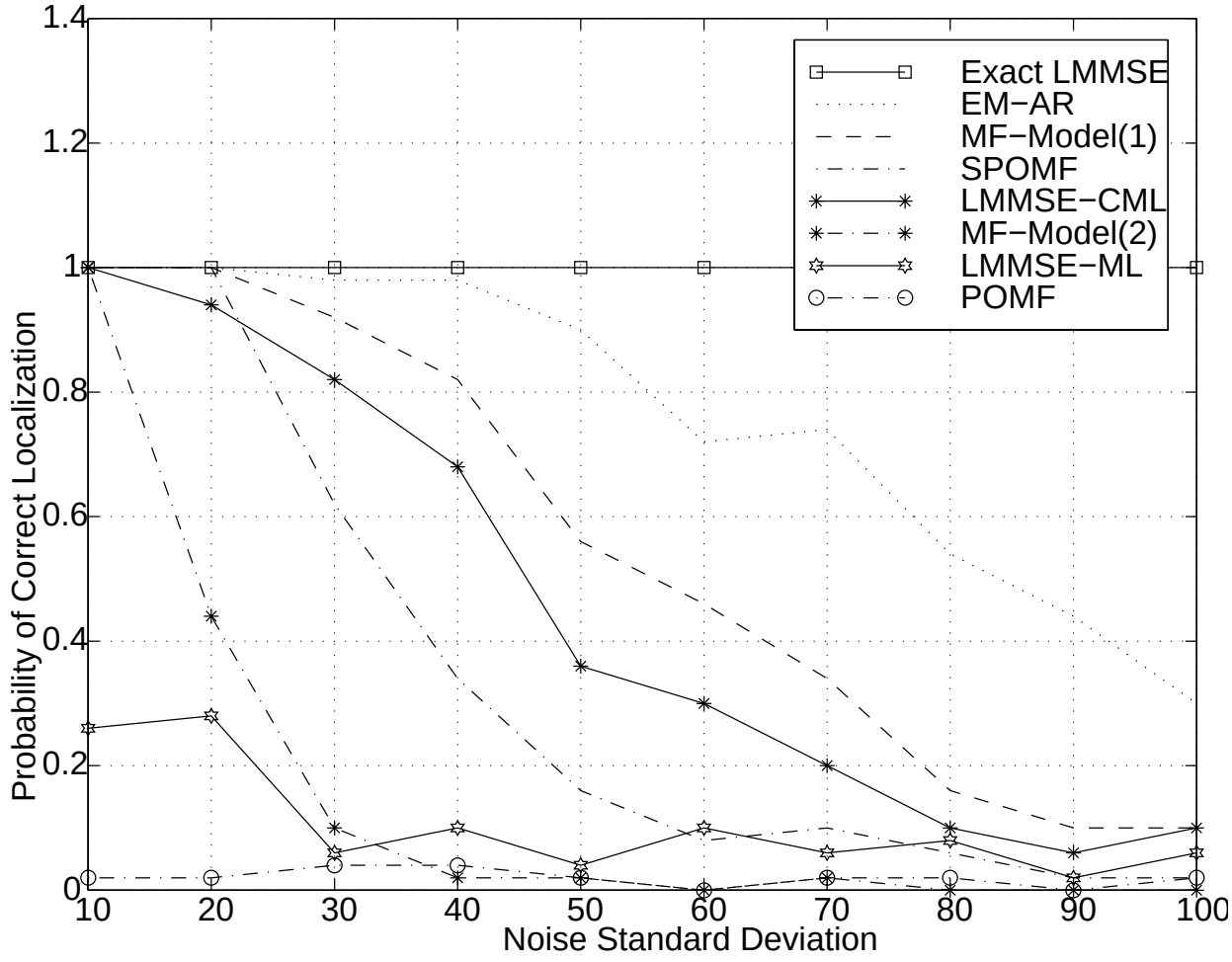

Figure 10: Plot of the probability of correct localization as a function of noise standard deviation.

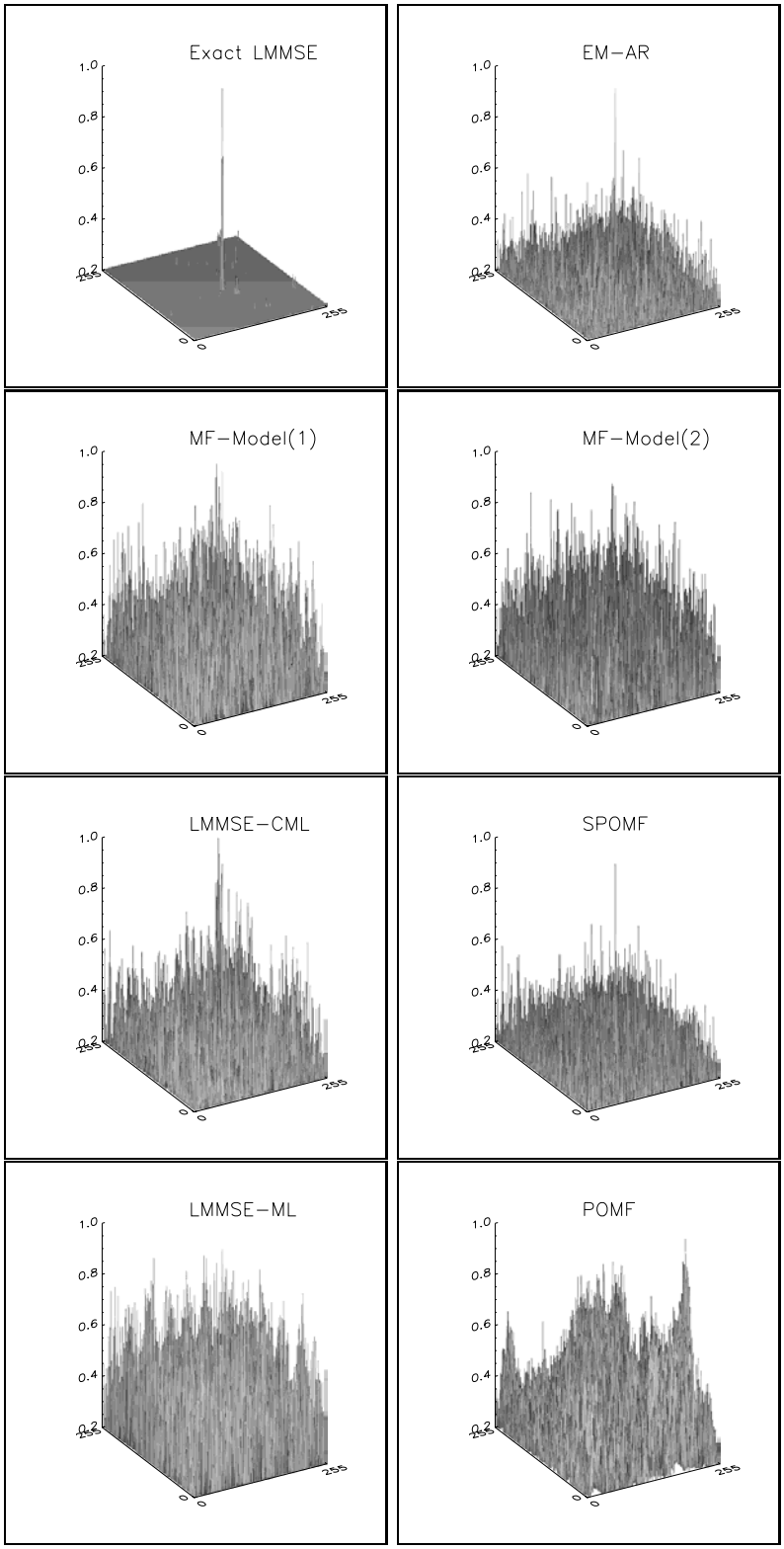

Figure 11: Examples of the restored impulse for $\sigma = 30$.

Table 1. Mean Squared Error values for Different Filters for $\sigma = 30$

	MSE
Exact LMMSE	4.4×10^{-3}
EM-AR	1.6×10^{-2}
MF-Model(1)	3.7×10^{-2}
MF-Model(2)	1.2×10^{-1}
LMMSE-CML	4.2×10^{-2}
SPOMF	5.5×10^{-2}
LMMSE-ML	1.2×10^{-1}
POMF	1.3×10^{-1}