A New Extension of Kalman Filter to Non-Gaussian Priors

G.R.K.S. Subrahmanyam, A.N. Rajagopalan, and R. Aravind

IIT Madras, India

Abstract. In the Kalman filter, the state dynamics is specified by the state equation while the measurement equation characterizes the likelihood. In this paper, we propose a generalized methodology of specifying state dynamics using the conditional density of the states given its neighbors without explicitly defining the state equation. In other words, the typically strict linear constraint on the state dynamics imposed by the state equation is relaxed by specifying the conditional density function and using it as the prior in predicting the state. Based on the above idea, we propose a sampling-based Kalman Filter (KF) for the image estimation problem. The novelty in our approach lies in the fact that we compute the mean and covariance of the prior (possibly non-Gaussian) by importance sampling. These apriori mean and covariance are fed to the update equations of the KF to estimate the aposteriori estimates of the state. We show that the estimates obtained by the proposed strategy are superior to those obtained by the traditional Kalman filter that uses the auto-regressive state model.

Keywords: Dynamic state space models, Kalman filter, Auto-regressive models, Importance sampling, Markov random fields.

1 Introduction

The problem of image estimation involves recovering the original image from its noisy version. The image estimation problem can be cast in to a state estimation from noisy measurements in state space representation of the image. When the state transition and measurement equations are both linear, and the state and measurement noises are independent and additive Gaussian, the Kalman filter gives the minimum mean square error (MMSE) estimate of the state. Extension of the 1-D KF to 2-D was first proposed by Woods and Radewan [7]. They considered the local neighborhood in updation of the state vector and arrived at a suboptimal filter known as the reduced update Kalman filter (RUKF) [8]. The reduced order model Kalman filter (ROMKF) proposed in [9] includes only local states in its state vector, but performs on par with RUKF. Effects of any distortion resulting from blur and noise can be removed by Kalman filtering, provided the appropriate image and blur parameters are completely known. In general, however, such parameters are apriori unknown, and furthermore can vary spatially as a function of the image coordinates. Hence, adaptive identification/filtering procedures are necessary for satisfactory restoration. A rapid

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 162–171, 2006.

edge adaptive filter for restoration of noisy and blurred images based on multiple models has been presented in [12].

A primary issue with all image estimation methods is about how they handle noise smoothing versus preservation of edges since the two requirements are contradictory. Geman and Geman [13] approach the edge preservation problem using line fields. The smoothness constraint is switched off at points where the magnitude of the signal derivative exceeds certain thresholds. For a thorough survey of techniques for image estimation, we refer the reader to [11].

To preserve edges, one must look beyond Gaussianity. Increasingly, for many application areas, it is becoming important to include elements of non-linearity and non-Gaussianity, in order to model accurately the underlying dynamics of a physical system. In this paper, we propose an interesting extension to the traditional Kalman filter to tackle discontinuities by incorporating non-Gaussianity within the Kalman filtering framework. This is achieved by modeling the prior as a discontinuity adaptive Markov random field and proposing sampling-based approaches to derive necessary statistical parameters required for the update stage of the Kalman filter. If the state transition equation is not known but an assumption on the state transition density (possibly non-Gaussian) can be made we can still use the Kalman filter update equations in the proposed frame work. The edge preservation capability is implicitly incorporated using the discontinuity adaptive state conditional density. Importance sampling is used to obtain the statistics of this PDF and the Kalman filter is used to update the prior estimates.

We use the discontinuity adaptive function given by Li [2] to construct the prior conditional density and show how the edges are better retained in our method. This is in addition to obtaining better overall estimates of the entire image. It may be noted that the proposed approach is different from the Ensemble Kalman filter [6,5] which is based on Monte Carlo simulation of the state probability distribution. It works by creating and propagating the ensemble through model operator. The mean and the error covariance are obtained by the analysis of the ensemble. In contrast, we use the Monte Carlo approach only to determine the mean and covariance of the conditional PDF. It is possible to extend the proposed approach to nonlinear filtering problem.

2 The Kalman Filter

The Kalman filter, rooted in the state-space formulation of linear dynamical systems, provides a recursive solution to the linear optimal filtering problem [10]. It applies to stationary as well as non-stationary environments. The solution is recursive in that each updated estimate of the state is computed from the previous estimate and the new input data.

2.1 Dynamic State-Space Model

The general state space model can be broken down into a state transition model and measurement model. In linear Gaussian regression, the state space representation is as follows:

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{w}_k \tag{1}$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{2}$$

where $\mathbf{y}_k \in \mathtt{R}^{n_y}$ denotes the output observations, $\mathbf{x}_k \in \mathtt{R}^{n_x}$ is the state of the system, $\mathbf{w}_k \in \mathtt{R}^{n_w}$ is the process noise and $\mathbf{v}_k \in \mathtt{R}^{n_v}$ is the measurement noise. The mappings $\mathbf{F}_k : \mathtt{R}^{n_x} \to \mathtt{R}^{n_x}$ and $\mathbf{H}_k : \mathtt{R}^{n_x} \to \mathtt{R}^{n_y}$ represent the deterministic process and measurement models. To complete the specification of the model, the prior distribution is denoted by $p(\mathbf{x}_0)$. The process noise $\mathbf{w}_k$ is assumed to be additive white Gaussian, with zero mean and with covariance matrix defined by $\mathbf{Q}_k$. The measurement noise $\mathbf{v}_k$ is additive white Gaussian with covariance matrix $\mathbf{R}_k$. The process and measurement noise are assumed to be uncorrelated. The states are assumed to follow a first-order Markov model and the observations are assumed to be independent given the states.

For the state space model given above, the minimum mean squared error (MMSE) estimate of the state $\mathbf{x}_k$ can be derived using the following Kalman recursive equations [3]:

State estimate propagation:- $\quad \widehat{\mathbf{x}}_{k/k-1} = \mathbf{F}_k \widehat{\mathbf{x}}_{k-1}$

Error covariance propagation:- $\quad \mathbf{C}_{k/k-1} = \mathbf{F}_k \mathbf{C}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_{k-1}$

Kalman gain matrix:- $\quad \mathbf{K}_k = \mathbf{C}_{k/k-1} \mathbf{H}_k^T \left[\mathbf{H}_k \mathbf{C}_{k/k-1} \mathbf{H}_k^T + \mathbf{R}_k \right]^{-1}$

State estimate update:- $\quad \widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_{k/k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \widehat{\mathbf{x}}_{k/k-1})$

Error covariance update:- $\quad \mathbf{C}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{C}_{k/k-1}$

Here, $\widehat{\mathbf{x}}_{k-1}$ and $\mathbf{C}_{k-1}$ are the posteriori estimates of the state and error covariance of the previous step available at time k, $\widehat{\mathbf{x}}_{k/k-1}$ and $\mathbf{C}_{k/k-1}$ are the apriori estimates of the state and error covariance at time k, $\mathbf{y}_k$ is the new measurement at time k, $\mathbf{F}_k$ and $\mathbf{H}_k$ are the state transition and measurement matrices at time k, $\mathbf{K}_k$ is the Kalman gain, and the $\widehat{\mathbf{x}}_k$ and $\mathbf{C}_k$ are the posterior state and error covariance of the present step.

For the image estimation problem, $\mathbf{x}_k$ corresponds to the true image pixels and $\mathbf{y}_k$ are the observations of the degraded image pixel. Matrix $\mathbf{F}_k$ contains the auto-regressive (AR) coefficients of the image. For example, if a_1, a_2, a_3 are the AR coefficients of the original image (i.e., coefficients of a three pixel neighborhood with non-symmetric half plane support (NSHP)), then $\mathbf{F}_k = \begin{bmatrix} a_1 & a_2 & a_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

Since we do not assume any blurring, we have $\mathbf{H}_k = [1\ 0\ 0]$. The above filter is referred to as the Auto-Regressive Kalman Filter (ARKF). Note that the filter imposes a strong (linear) constraint on the state equation. It is important to observe that linear dependence implies statistical dependence but not vice-versa. Our idea is to arrive at a more general framework wherein pixel dependencies can be expressed statistically.

3 Discontinuity Adaptive Prior

A realization of a random field is generated when we perform a random experiment at each spatial location and assigns the outcome of the random experiment to that location. A Markov random field (MRF) possesses Markovian property: i.e., the value of a pixel depends only on the values of its neighboring pixels and on no other pixel [4,2]. More details of MRF can be found in Li [2].

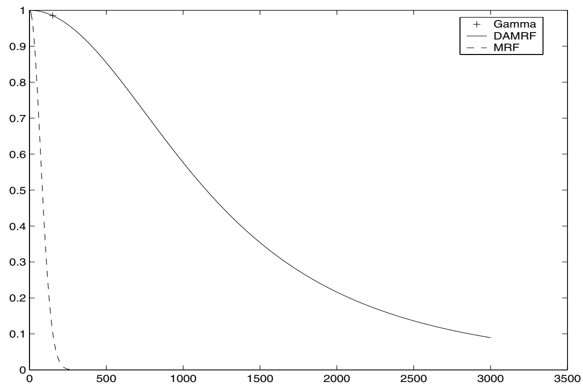


Fig. 1. Plot shows how MRF and DAMRF differ in the weighing with respect to η

Smoothness is a property that underlies a wide range of physical phenomena. However, it is not valid at discontinuities. How to apply the smoothness constraint while preserving edges has been an active research area within the MRF framework. Li [2] identifies that the fundamental difference among different models for dealing with discontinuities lies in the manner of controlling the interaction among neighboring points. Li then proposes a discontinuity adaptive (DA) model based on the principle that whenever a discontinuity occurs, the interaction should diminish. One such interaction function is $h_\gamma(\eta) = \frac{1}{1+\frac{\eta^2}{\gamma}}$ and its corresponding adaptive potential function is $g_\gamma(\eta) = \gamma \log(1 + \frac{\eta^2}{\gamma})$. The function is such that the smoothing strength $|\eta h_\gamma(\eta)|$ increases monotonically as η increases [2] within a band $B_\gamma = (-\sqrt{\gamma}, \sqrt{\gamma})$. Outside the band, smoothing decreases as η increases and becomes zero as $\eta \rightarrow \infty$. This enables it to preserve image discontinuities. It differs from the quadratic (Gaussian) regularizer which smoothes edges as $\eta \rightarrow \infty$. In Fig. 1 we show that for large η the Gaussian MRF assigns zero weight while the discontinuity adaptive MRF (DAMRF) allows edges with finite weight.

In the case of a simple GMRF model, the state conditional probability density function (PDF) is given by $\exp(-\eta^2)$ where $\eta^2(x) = ((x - c_1)^2 + (x - c_2)^2 + (x - c_3)^2)/2\beta^2$. Pixels c_1, c_2, c_3 denote the previously (estimated) pixels in the NSHP support. This can be shown to be equivalent to a Gaussian (PDF) with mean $(c_1+c_2+c_3)/3$ and variance $\Gamma = \beta^2/3$. We assume the state conditional density to be non-Gaussian and of the form $\exp(-g_\gamma(\eta))$ where $g_\gamma(\eta) = \gamma \log(1 + \frac{\eta^2}{\gamma})$ and η is as defined in the simple MRF case which leads to the DAMRF model [2].

4 Importance Sampling

It is not analytically possible to compute the mean and covariance of the non-Gaussian DAMRF distribution. Hence, we resort to Monte Carlo techniques. An efficient way of doing this is to adopt the importance sampling method. Our aim is to obtain the conditional mean and variance of the distribution corresponding to the DAMRF at every pixel, using importance sampling.

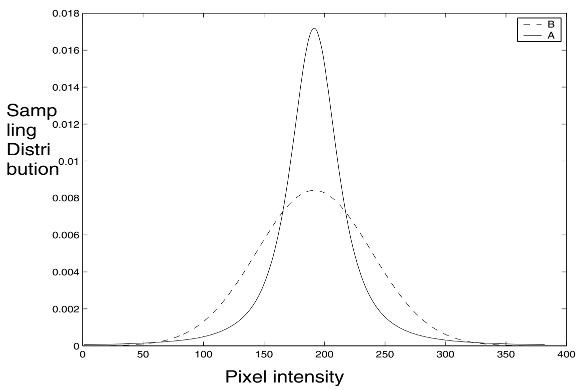


Fig. 2. Importance sampling: A is the PDF whose moments are to be estimated, while B is the sampler density

Importance Sampling (IS) is a Monte Carlo method to determine the estimates of a (non-Gaussian) target PDF, provided its functional form is known up to a multiplication constant [1]. Let us consider a PDF $A(s)$ which is known up to a multiplicative constant but it is very difficult to make any estimates of its moments. However, from the functional form, we can estimate its support (region where it is non-zero). Consider a different distribution $B(s)$ which is known up to a multiplicative constant, is easy to sample, and is such that the (non-zero) support of $B(s)$ includes the support of $A(s)$. Such a density $B(s)$ is called a sampler density. A typical plot showing the PDFs of B (solid line) and A (dashed line) is given in Fig. 2.

Our aim is to determine the first two central moments of the PDF A. Since it is difficult to draw samples from the non-Gaussian PDF A, we draw L samples, $\{s^{(l)}\}_{l=1}^{L}$ from the sampler PDF B. If these were under A, we can determine the moments of A with these samples. In order to use these samples to determine the estimates of the moments of A, we proceed as follows.

When we use samples from B to determine any estimates under A, in the regions where B is greater than A, these estimates are over-represented. In the regions where B is less than A, they are under-represented. To account for this, we use correction weights $w^l = \frac{A(s^{(l)})}{B(s^{(l)})}$ in determining the estimates under A. For example, to determine the mean of the distribution A we use $\widehat{\mu}_a = \frac{\sum_l w^l s^{(l)}}{\sum_l w^l}$. If $L \to \infty$ the estimate $\widehat{\mu}_a$ tends to the actual mean of A. This methodology of estimating moments of A by sampling from an importance function B forms the core of importance sampling.

5 The Proposed Kalman Filter

In this section, we present a new algorithm for estimating an image from its degraded version using the state conditional PDF and Kalman filter update equations. In section 3, we showed how to construct a DAMRF PDF using a discontinuity adaptivity MRF function. In section 4, we explained how to determine the estimates of a PDF using importance sampling. We now present a novel strategy which integrates the above steps within the Kalman filter framework to restore images degraded by additive white Gaussian noise. In the proposed strategy, only the assumption on the conditional PDF needs to be made. The parameters of the PDF are a function of the already estimated pixels and the values of Γ and γ. This implicitly generalizes the state transition equation. The steps involved in the proposed method are as follows:

1. At each pixel, construct the state conditional PDF using the past three pixels from its NSHP support, and the values of Γ and γ in the DAMRF model (section 3) . Using the DAMRF function given by [2] we construct the state conditional PDF as
$$P(X(m,n)/\widehat{X}(m-i,n-j)) = \exp\left(-\gamma \log(1 + \frac{\eta^2(X(m,n))}{\gamma})\right); \qquad (3)$$
where $(i,j) = (0,1), (1,0), (1,1)$ and
$$\eta^2(X(m,n)) = ((X(m,n) - \widehat{X}(m,n-1))^2 + (X(m,n) - \widehat{X}(m-1,n))^2 + (X(m,n) - \widehat{X}(m-1,n-1))^2)/(2\beta^2), \text{ and } \beta^2 = 3\Gamma.$$
Here, X and $\widehat{X}$ refers to the original image and the estimated image, respectively. The pixels $\widehat{X}(m,n-1)$,$\widehat{X}(m-1,n)$ and $\widehat{X}(m-1,n-1)$ are the (estimated) past three pixels of the NSHP support.
2. Obtain the mean and covariance of the above PDF using importance sampling as described in section 4. Explicitly, we draw samples $\{s^l\}$ from a Gaussian sampler [1] The sampler $B(s)$ has mean $\mu_b = (\widehat{X}(m,n-1) + \widehat{X}(m-1,n) + \widehat{X}(m-1,n-1))/3$ and variance $\sigma_b^2 = 15\beta^2$. We weight these samples through the importance weights $w^l = \frac{A(s^l)}{B(s^l)}$. The mean $\widehat{\mu}_a$ and variance $\widehat{\sigma}_a^2$ of A are computed as
$$\widehat{\mu}_a = \frac{\sum_l w_l s^{(l)}}{\sum_l w_l} \qquad \widehat{\sigma}_a^2 = \frac{\sum_l w_l (s^{(l)} - \mu_a)^2}{\sum_l w_l} \qquad (4)$$

[1] The idea is to have the support of the target density A included in the support of the sampler density B so that the mean 'μ_b' is near to the actual mean of the MRF, and the variance 'σ_b^2' is high enough.

3. The predicted mean and error covariance are fed to the update stage of the Kalman filter as follows:

   $\widehat{\mathbf{x}}_{k/k-1} = \widehat{\mu}_a$; $\mathbf{C}_{k/k-1} = \widehat{\sigma}_a^2$;

   Kalman gain matrix:- $\mathbf{K}_k = \mathbf{C}_{k/k-1}\mathbf{H}_k^T\left[\mathbf{H}_k\mathbf{C}_{k/k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1}$

   State estimate update:- $\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_{k/k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k/k-1})$

   This gives the estimated mean $\widehat{X}(m,n) = \widehat{\mathbf{x}}_k$; go to step 1 and repeat.

   Finally, the filtered image is $\widehat{X}$.

We note that in this case the state becomes a scalar, the matrix $\mathbf{H}_n = 1$, and $\mathbf{y}_n$ is the scalar observation pixel. This approach does not need the state equation (1).

In the proposed approach, based on the past three pixels of the NSHP support, the prior is constructed. Importance sampling is used to estimate the mean and covariance of the non-Gaussian prior. These estimates are effectively used by the Kalman filter update equations (Kalman gain and mean updation equations), to arrive at the posterior mean (the estimated pixel intensity). Note that in the proposed formulation, the prior is not restricted to be Gaussian. In other words, the process noise can have any distribution but with a known functional form.

6 Experimental Results

In this section, we compare the proposed importance sampling based Kalman Filter (ISKF) with the auto-regressive Kalman Filter (ARKF). In an AR based Kalman filter, the original image is used to determine the AR coefficients and the process noise. An alternative is to use the AR coefficients obtained from images of the same class or to use the observed image itself. But this will in general, degrade the performance of the algorithm. In contrast for the proposed algorithm, the image model parameters are not required. Since the conditional PDF has all the information. The proposed algorithm has two parameters γ and Γ which depend on the image. We have found that the optimum γ for most images is in the range of 1 to 2 while the required value of Γ is in the range of 50 to 150. For low values of γ and high values of Γ the estimated image will be noisy, and for high values of γ and low values of Γ the estimated image will be blurred. For a quantitative comparesion of ARKF and the proposed method we use the improvement-in-signal-to-noise-ratio (ISNR) which is defined as $ISNR = 10\log_{10}\left(\frac{\sum_{m,n}(Y(m,n)-X(m,n))^2}{\sum_{m,n}(\widehat{X}(m,n)-X(m,n))^2}\right) dB$. Here, (m,n) are over entair image. X, Y and $\widehat{X}$ represent the original image, degraded observation, and the estimated image, respectively.

Fig. 3(a) shows the "daisy" image. The image after degradation by additive white Gaussian noise of $SNR = 10$ dB is shown in Fig. 4(a). The images estimated by ARKF and the proposed importance sampling-based Kalman filter, are given in Figs. 4(b) and 4(c), respectively. Note that the image estimated by the proposed approach has sharp petals. At the same time, it is less noisy in

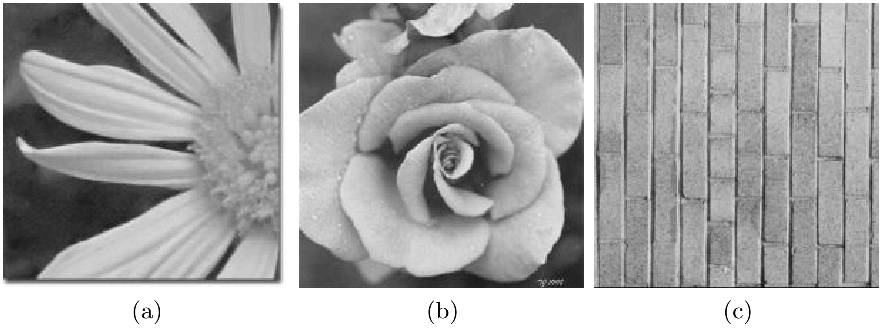


Fig. 3. Original images (a) Daisy image, (b) Flowers image and (c) Bric image

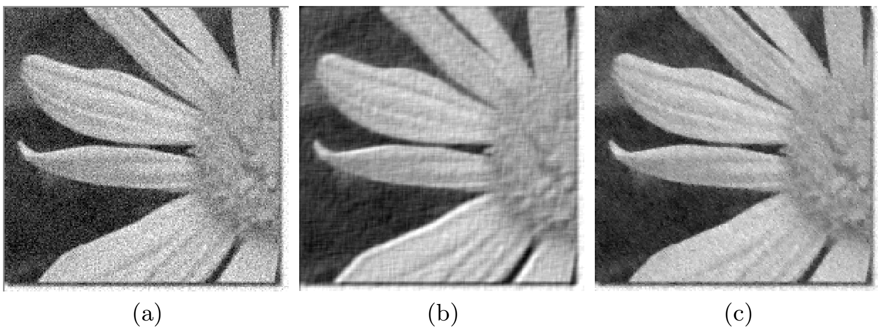


Fig. 4. Daisy (a) Image degraded by additive white Gaussian noise ($SNR = 10$ dB). Image estimated using (b) AR based KF ($ISNR = 3.42$ dB) and (c) Proposed method ($ISNR = 4.25$ dB, $\Gamma = 50$, $\gamma = 1.5$).

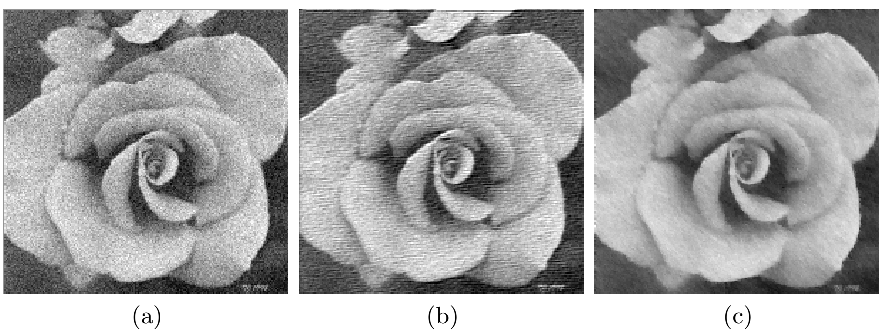
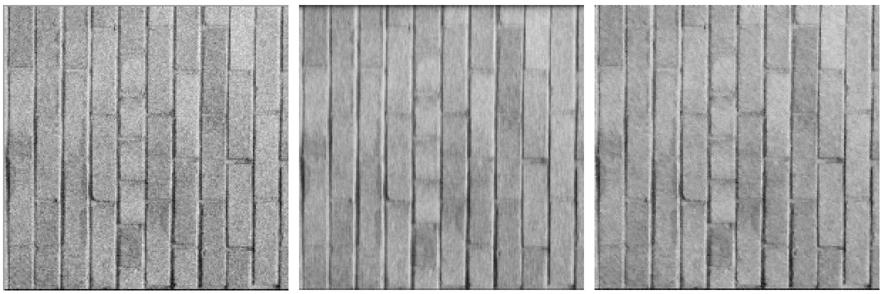


Fig. 5. Flower (a) Degraded image ($SNR = 10$ dB). Image estimated by (b) AR based KF ($ISNR = -1.39$ dB) (c) Proposed method ($ISNR = 3.73$ dB, $\Gamma = 50$, $\gamma = 1.5$).

homogeneous regions compared to the ARKF output. It has a superior improvement-in-signal-to-noise-ratio (ISNR) value over ARKF.

Next, we show in Fig. 3(b) a flower image. It is degraded by additive white Gaussian noise of $SNR = 10$ dB (Fig. 5 (a)). The image estimated by ARKF and the proposed approach are shown in Figs. 5(b) and 5 (c), respectively. The image estimated by the proposed approach has very little noise, retains the edges, and has higher ISNR value. Note the ringing-like artifact in the image estimated by ARKF. For the proposed method, the overall appearance of the estimated image is quite good.



(a) (b) (c)

Fig. 6. Brick (a) Degraded image ($SNR = 10$ dB). Image estimated using (b) AR based KF ($ISNR = -1.3$ dB) (c) Proposed method ($ISNR = 1.96$ dB, $\Gamma = 50$, $\gamma = 2$).

Fig. 3(c) shows a brick image while its degraded version is given in Fig. 6(a). The images estimated by ARKF and the proposed method are shown in Figs. 6(b) and Fig. 6(c), respectively. The proposed sampling-based Kalman filter again outperforms ARKF. The image retains the horizontal edges quite well and is much closer to the original image as compared to ARKF.

The above results show that the proposed approach is superior to ARKF in reducing noise, preserving edges, and yelding better ISNR values. Fixing the parameters for the proposed scheme is also quite simple as discussed in the beginning of this section.

7 Conclusions

We have proposed a novel importance sampling-based discontinuity adaptive Kalman filter. Instead of using the state transition equation to predict the mean and error covariance (as in traditional Kalman filter formulation), we use a DA non-Gaussian state conditional density function for prediction. Importance sampling is used to determine the apriori mean and covariance of a DAMRF model. These are then used in the Kalman filter update equations to obtain the a posteriori mean. The image estimates obtained by the proposed approach are superior to those obtained with the auto-regressive Kalman filter.

References

1. D. J. C. Mackay, "Introduction to Monte Carlo methods", In M. I. Jordan, editor, "Learning in graphical models", NATO Science Series, pp. 175–204. *Kluwer Academic Press*, 1998.

2. S. Z. Li, "Markov random field modeling in computer vision", *Springer Verlag*, 1995.
3. S. Haykin, "Kalman filtering and neural networks", *John Wiley and Sons*, 2001.
4. M. Petrou and P. G. Sevilla, "Image processing dealing with texture", *John Wiley and Sons*, 2006.
5. J. Drecourt, "Kalman filtering in hydrological modeling", *Technical report* (2003).
6. G. Evensen, "The ensemble Kalman filter: theoretical formulation and practical implementation", *Ocean Dynamics*, vol. 53, pp. 343-367, 2003.
7. J. W. Woods and C. H. Radewan, "Kalman filter in two dimensions", *IEEE Trans. on Information Theory*, vol. 23, pp. 473-482, 1977.
8. J. W. Woods and V. K. Ingle, "Kalman filtering in two dimensions-Further results", *IEEE Trans. Acoustics Speech and Signal Proc.*, vol. 29, pp. 188-197, 1981.
9. D. Angwin and H. Kaufman, "Image restoration using a reduced order model Kalman filter", *IEEE Trans. on Signal Processing*, vol. 16, pp. 21-28, 1989.
10. http://www.cs.unc.edu/~welch/kalman/
11. H. Kaufman and A. M. Tekalp, "Survey of estimation techniques in image restoration", *IEEE Control Systems Magazine*, vol. 11, pp. 16 - 24, 1991.
12. A. M. Tekalp, H. Kaufman and J. Woods, "Edge-adaptive image restoration with ringing suppression", *IEEE Trans. Acoustics Speech and Signal Proc.* vol. 37, pp. 892-899, 1989.
13. S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721-741, 1984.