

Adaptive Background Estimation using an Information Theoretic Cost for Hidden State Estimation

Goktug T. Cinar and José C. Príncipe

Abstract—Hidden state estimation in linear systems is a popular and broad research topic which became a mainstream research area after Rudolf Kalman’s seminal paper. The Kalman Filter (KF) gives the optimal solution to the estimation problem in a setting where all the processes are Gaussian random processes. However because of the sub-optimal behavior of the KF in non-Gaussian settings, there is a need for a new filter that can extract higher order information from the signals. In this paper we propose using an information theoretic cost function utilizing the similarity measure *Correntropy* as a performance index. We present the superior performance of the new filter on both synthetic data and on adaptive background estimation problem and discuss future research directions.

I. INTRODUCTION AND BACKGROUND

Hidden state estimation in linear systems is a popular and broad research topic which became a mainstream research area after Rudolf Kalman’s seminal paper [1]. It is well known that the KF is the optimal solution for hidden state estimation when the system is driven by Gaussian processes. The elegance of the approach lies in having an analytical solution to the state estimation and therefore having the optimal estimation at every iteration.

The problem setup is as follows. The system equation is given as:

$$x_k = F_{k|k-1}x_{k-1} + w_{k-1} \quad (1)$$

and the measurement equation is:

$$y_k = H_k x_k + v_k \quad (2)$$

where w_k and v_k are independent, zero mean, Gaussian noise processes with covariance matrices Q_k and R_k , respectively.

It is very important to understand the fundamental idea behind the formulation of Kalman Filter (KF). Among many ways that Kalman Filter formulation can be derived such as using orthogonality principle or the innovations approach, there is a derivation that can be achieved by the optimization of the cost function given in (3) [2].

$$J = \frac{1}{2}(y_k - H_k x_k)^T R_k^{-1}(y_k - H_k x_k) + \frac{1}{2}(x_k - x_k^-)^T M_k^{-1}(x_k - x_k^-) \quad (3)$$

where M_k is the covariance matrix of $(x_k - x_k^-)$ and:

Goktug T. Cinar and José C. Príncipe are with the CNEL, Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA (email: gcinar@ufl.edu, principe@cnel.ufl.edu).

This work was supported by NSF grant ECCS 0856441

$$x_k^- = F_{k|k-1}x_{k-1} \quad (4)$$

The value $\hat{x}_k = \arg\min_{x_k} J(x_k)$ will be the best estimate. In this paper the system parameters will be assumed fixed for simplicity however all the formulation can be extended easily to time-varying case. For this purpose from now on $F = F_{k|k-1}$ and $H = H_k$ will be used for system and measurement matrices respectively.

The formulation of KF can be derived by analytically solving (5).

$$\frac{\partial J}{\partial x_k} = 0 \quad (5)$$

Beyond the formulation that is achieved through this simple optimization process, there is an important message embedded in this cost function. It states that the main goal of the estimation process is to reduce the effect of the two sources of uncertainties in the dynamical system. The system noise (uncertainty) w_k and the measurement noise (uncertainty) v_k . Basically the filter tries to account for these uncertainties and as it propagates the covariance matrix of the error through iterations, it only makes use of the second order information in the measurements. Therefore it is optimal in the case of Gaussian uncertainties, however it is sub-optimal when the gaussianity condition is loosened.

However it is reasonable to say that there would be many applications of the filter where the gaussianity condition would not hold and we would be left with the sub-optimal solution the filter produces. Therefore there is a need for a method that would use the information available in the higher-order statistics of the signals. One such measure is *Correntropy* [3].

Consider two scalar random variables $X, Y \in \Re$. Cross-Correntropy [3] is a generalized similarity measure between two arbitrary scalar random variables X and Y defined by:

$$v(X, Y) = E_{XY}[\kappa(X, Y)] = \int \int \kappa(x, y) p_{X,Y}(x, y) dx dy \quad (6)$$

where $\kappa(\cdot, \cdot)$ is any continuous positive definite kernel. In our formulation the Gaussian kernel $G_\sigma(\cdot, \cdot)$ will be employed. Therefore (6) becomes:

$$v(X, Y) = E_{XY}[G_\sigma(X, Y)] = \int \int G_\sigma(x, y) p_{X,Y}(x, y) dx dy \quad (7)$$

where σ is the kernel size or bandwidth. As we have only limited amount of data and the joint PDF is unknown, we use the sample estimator instead of the expectation operator.

$$\hat{v}_{\sigma,N}(X, Y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(x_i - y_i) \quad (8)$$

One nice property of correntropy is that it is positive and bounded and with Gaussian kernel it reaches its maximum if and only if $X = Y$. Another important property is that for Gaussian kernel, correntropy is a weighted sum of all the even moments of the random variable $Y - X$. The proof follows from the Taylor series expansion of the Gaussian function in (7):

$$v_{\sigma}(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E[(X - Y)^{2n}] \quad (9)$$

It is seen in (9) that correntropy is the sum of all even moments when the Gaussian kernel is used. The kernel size appears as a parameter weighting the second-order moment and higher order moments. With very large σ (compared to the dynamic range of the data) this measure approaches correlation.

Having the property of correntropy being the weighted sum of all even order moments of the random variable, it might be useful to employ it in our cost function. Therefore the cost function to be optimized in the Correntropy Filter is:

$$J_c = v(\|y_k - Hx_k\|) + \frac{1}{L} v(\|x_k - Fx_{k-1}\|) \quad (10)$$

$$J_c = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(\|y_i - Hx_i\|) + \frac{1}{LN} \sum_{i=1}^N G_{\sigma}(\|x_i - Fx_{i-1}\|) \quad (11)$$

The reason behind using the Euclidean norm is to enable the usage of the definition of Correntropy (6) on a multiple-state multiple-output system. We choose $N=1$ to use the instantaneous gradient, which makes the formula equal to:

$$J_c = G_{\sigma}(\|y_i - Hx_i\|) + \frac{1}{L} G_{\sigma}(\|x_i - Fx_{i-1}\|) \quad (12)$$

To make the derivation easily followed we are going to split our cost into two parts where:

$$J_1 = G_{\sigma}(\|y_i - Hx_i\|)$$

$$J_2 = \frac{1}{L} G_{\sigma}(\|x_i - Fx_{i-1}\|)$$

$$\frac{\partial J}{\partial x_i} = \frac{\partial J_1}{\partial x_i} + \frac{\partial J_2}{\partial x_i}$$

$$\frac{\partial J_1}{\partial x_i} = \frac{1}{\sigma^2} G_{\sigma}(\|y_i - Hx_i\|)(y_i - Hx_i)^T H$$

$$\frac{\partial J_2}{\partial x_i} = \frac{-1}{L\sigma^2} G_{\sigma}(\|x_i - Fx_{i-1}\|)(x_i - Fx_{i-1})$$

II. RESULTS

The filter is tested on artificial data first. To create the data we have used a two dimensional rotation matrix, the system has no inputs and is driven by a Gaussian mixture noise. The mixture consists of two Gaussian distributions centered at 0 and 2 with variances 0.1 respectively. The observed signal is the sum of the hidden states. We see the superior performance of the Correntropy filter in Fig. 1. Also one can observe the error *pdf*s of both correntropy filter and KF in Fig. 2. The new filter places the error distributions at 0 which is very desirable. It also results in smaller error variance compared to KF. In the measurement estimation the new filter fails to place the error distribution around 0 however the mean is close to 0 and the variance is of the same order as KF. However it should be noted that as we assume the measurements are available to us, we actually care about the state estimation.

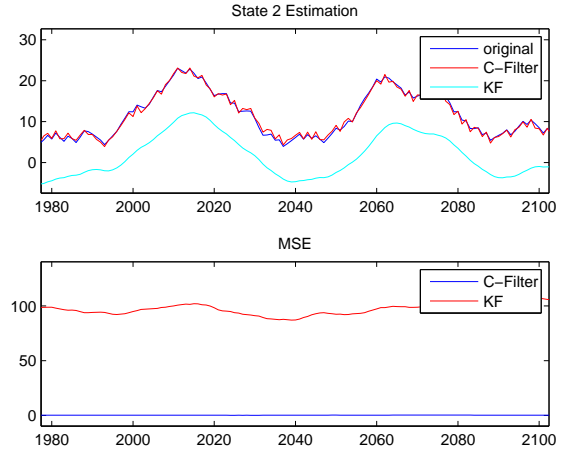


Fig. 1. The figure shows the state estimation of both filters and the true states.

A practical use of this filter might be beneficial in the following scenario. It is well known that in video surveillance applications it is very important to estimate background and separate foreground objects from the background. This will give possibility to detect sudden changes in the scene, or would enable the machine to track the moving objects/subjects. However most of the time this is achieved through human supervision, and a human operator finds a moment on the video track with smallest possible number of foreground objects and assigns it as background. This however is open to errors and will suffer from the changes that can occur through natural causes such as illumination changes and seasonal changes. Therefore it is desirable

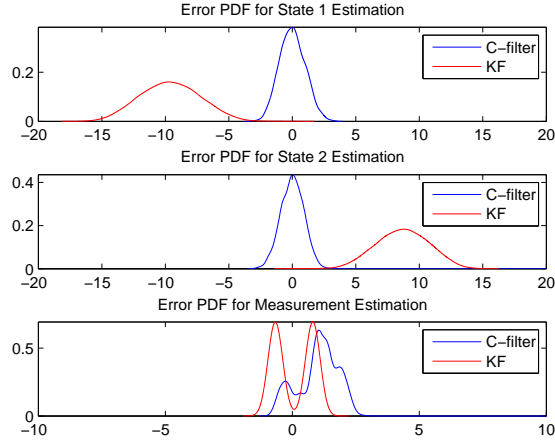


Fig. 2. The figure shows the error distributions of the estimations.

to have an adaptive background estimation. Usage of the Kalman Filter without a regularizer is problematic in this setting because of the well known properties of the MSE cost. KF would try to adapt more with respect to the sudden changes in the pixel values. On the other hand Correntropy cost function will tend to neglect the outliers and therefore neglecting the sudden changes in the pixel values. To implement the idea the following setup is used. We assume there is a filter for each pixel of the video sequence. The three hidden states are the RGB values of the pixels of background. We have an identity matrix for both F and H . Having an identity matrix for the system matrix means we are not expecting the states to change which makes sense as we are trying to find the background. As we have the measurement as the pixel values of the current frame the states directly relate to the measurements. We assume the noise term in the state equation refers to the moving foreground objects and the noise in the measurement equation refers to the noise introduced by the sensors, such as salt and pepper noise. We initialize the background as the first frame available to us and then let the filter work with each incoming frame unsupervised. As a result the filter manages to extract the background, eliminate the salt and pepper noise and adapt to the changes in the background scene; such as removing a vehicle that left the parking lot from the background or phasing a newly parked vehicle from foreground to background.

Figures 3,4 and 5 demonstrate the success of the proposed method. Please notice that the filter gradually learns the static background and therefore eliminates the salt and pepper noise as well as the moving foreground objects. It also manages to include new objects to the background and exclude the ones that left. This is obtained through the measurement error in the cost function and the adaptation is done very quickly.

When the Kalman filter is applied on the same problem the outcome is as expected. The estimated background is almost



Fig. 3. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 10th frame in the video sequence. Notice that the filter yet learning the background.



Fig. 4. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 1003rd frame in the video sequence. Notice that the filter is phasing out the car that is leaving the parking lot from the background.

exactly the current frame of the video stream. The KF can only detect the edges as they consist the sudden changes in the pixel values. When the salt and pepper noise added to the video stream, KF manages to reduce the effect of the noise but fails to eliminate it completely. This is because of the over-reaction of mean-squared cost function to the outliers (in this case foreground objects and salt and pepper noise). The results are shown in figures 6 and 7.

The next example shows a football game and is more challenging as there are many moving objects at the same time. No measurement noise is added this time. Again the initialization is done by setting the first frame as the initial background. It takes about 8 seconds for the filter to gather the background information and exclude the players from the first scene. From that point on when the players are moving the filter removes them from the background, however as the



Fig. 5. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 2808th frame in the video sequence. Notice that the filter adapted to the new scene with new parked car and completely discarded the bicycle rider.



Fig. 7. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 1303rd frame in the video sequence. Notice that the Kalman filter reduces the effect of the noise but fails to extract the background.



Fig. 6. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 1303rd frame in the video sequence. Notice that the Kalman filter fails to extract the background as expected.



Fig. 8. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 50th frame in the video sequence. Notice that the filter yet learning the background and there are artifacts left from the initial frame.

game stops and the players rest they start to appear faintly in the background which demonstrates the adaptive force of the method. Also if we pay attention to the difference of current frame and background this faint reappearance does not effect the quality and it would be very easy to track the players. Results are shown in figures 8,9 and 10.

III. CONCLUSION AND DISCUSSIONS

In this report we presented the performance of the new information theoretic filter, *Correntropy Filter*, for hidden state estimation in linear systems with non-Gaussian uncertainties. We motivated the new filter by the cost function of the well-known Kalman Filter. As it is well-known the KF uses the second order statistics of the observed signals and therefore provides the optimal solution when the uncertainties in the system are Gaussian. On the other hand this assumption and second order cost function gives the possibility of coming

up with an analytical solution. Therefore the convergence of the filter is very fast.

It is obvious that the filter has very nice properties when the environment has Gaussian uncertainties; but what about real life scenarios? It is astonishing that KF still performs considerably well in non-Gaussian scenarios. However this performance might not be considered enough and opens up the need for a new filter that uses higher order statistics. This is the reason we adapted the cost function to use the information theoretic measure *Correntropy*. When the properties of correntropy are analyzed, it is seen that correntropy is indeed a weighted sum of all even moments of its argument. This right away gave us a hint that this filter would be useful in non-zero mean and non-Gaussian environments; which proved true. Even though we gained the power of the



Fig. 9. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 1283rd frame in the video sequence. Notice that the filter almost perfectly separated the foreground.

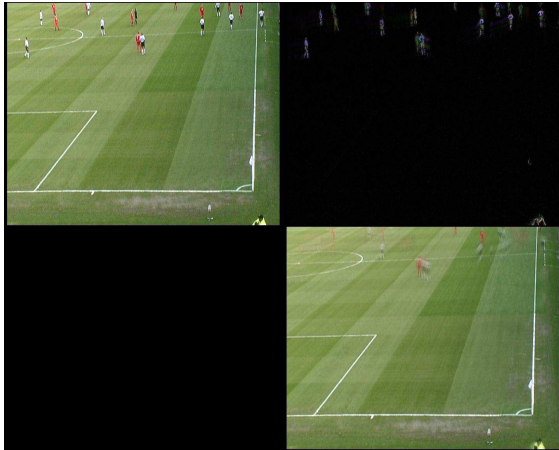


Fig. 10. The figure shows the current frame on top left, extracted foreground on top right, and the extracted background on bottom right. This is the 2363rd frame in the video sequence. Notice that as the game stopped for a throw the player start to appear faintly in the background scene.

update rule. This would eliminate the step size parameter as well as speeding up the convergence.

This paper should be considered as the presentation of a tool that can be applied to a broad range of problems. One such application is presented on the adaptive background problem. We show that the immunity of the Correntropy measure to outliers gives us the ability of extracting the background from a video sequence and adaptively change it with respect to the changes in the environment. It also copes with salt and pepper noise which is clearly not Gaussian.

REFERENCES

- [1] R. Kalman *et al.*, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [2] A. Bryson and Y. Ho, *Applied optimal control*. Wiley New York, 1975.
- [3] J. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. Springer Verlag, 2010.

information hidden in the higher order statistics, we lost the analytical solution which is an important feature of the KF and its variants. As the correntropy measure is a weighted sum of even moments, there might be a possibility to create an RLS type of algorithm to speed up the convergence.

The new filter has the free parameter called the kernel bandwidth, that controls the behavior of the filter. It is observed that the C-filter is behaving very similarly to the KF when it uses very large kernel bandwidth. When the kernel size is close to the dynamic range of the inputs, it uses higher order statistics and outperforms KF in non-Gaussian environments. For adaptive background problem we used a kernel bandwidth of 1.5 as the RGB values are between 0 and 1. The other free parameter is the step size and authors are currently working on the possibility of finding a fixed point