Ilmenau University of Technology
Faculty of Electrical Engineering and Information Technology
Institute for Media Technology
Audio-Visual Technology

Kalman Filtering

Investigation Seminar

Laurits Hamm
Herderstrasse 1
98693 Ilmenau
Register Number: 32781

November 2005

Professor: Prof. Dr.-Ing. Hans-Peter Schade
Assistent: Dipl.-Ing. Arne Nowak

Abstract

The Kalman Filter is a linear estimator and used in many applications. It has proven to be a practical and effective tool in digital computing. This work gives an introduction to the Discrete Kalman Filter and the system model for stochastic processes that it is based on. It focuses on the use of the Kalman Filter for prediction and tracking. Finally the HiBall tracking system which uses the Extended Kalman Filter is reviewed.

Contents

List of Figures

1 Introduction and Background

In 1960 Rudolph E. Kalman (Kal60) described a recursive solution to the discrete-data linear filtering problem. His solution later became known as the Kalman Filter (KF). It showed to be very feasible and practical in digital computing and digital data processing and thus the KF has found many applications. Kalman Filtering is an ongoing field of research even today and many advanced algorithms for higher performance and special applications have been developed.

1.1 What is the Kalman Filter?

The KF is an optimal linear estimator for a dynamic system. In this point it differs from the Wiener Filter which is an optimal estimator for a stationary system. The KF can be used to predict or estimate the future state of the system. It is used to estimate a state in the presence of noise. This is known as the black-box problem where the true state is estimated with an indirect observation. We can only observe or measure the output of the system which will be disturbed by noise. To use the KF the dynamic system needs to be described in the state-space form which is derived from the ARMA model for stochastic processes. The KF is optimal in the sense that is minimizes the mean-square of the estimation error. The algorithm is expressed in only five equations which have a recursive nature. This recursive description leads to an easy and practical implementation in digital computing.

The KF filter uses knowledge of the system in the form of statistics to calculate an estimate. It is important to note that all knowledge of the system is of statistical kind. The covariance of the stochastic processes is used in the calculation of the filter and describes this knowledge of the system. Every stochastic process consists of random variables and these relate a certain state with a certain probability. We only know that for a certain probability a certain state will occur. To start the algorithm of the KF the system needs to be initialized with a known state. The initialization can be more or less favourable. In any case the KF needs some time at the beginning to approach the true state of the system. One good property of the KF is that it converges fast towards the true state.

Why is it called a filter? Actually the Kalman filter is an adaptive filter. In 1994 Sayed and Kailath (AHS94) showed the one-to-one correspondence between the KF and the well known RLS (recursive least-squares adaptive filter). In this light the KF provides a framework for the derivation of the complete family of RLS algorithms. The RLS filter uses the method of least-squares. Like the KF the RLS is a recursive data processing algorithm. In adaptive filtering three basic forms of estimation can be distinguished: filtering, smoothing and prediction. The difference between these three forms is the span of time used in the processing. Filtering at time t includes data up to and including to t. Smoothing at time t' includes data that was measured after t. Prediction uses data up to the time t to estimate a state in the future.

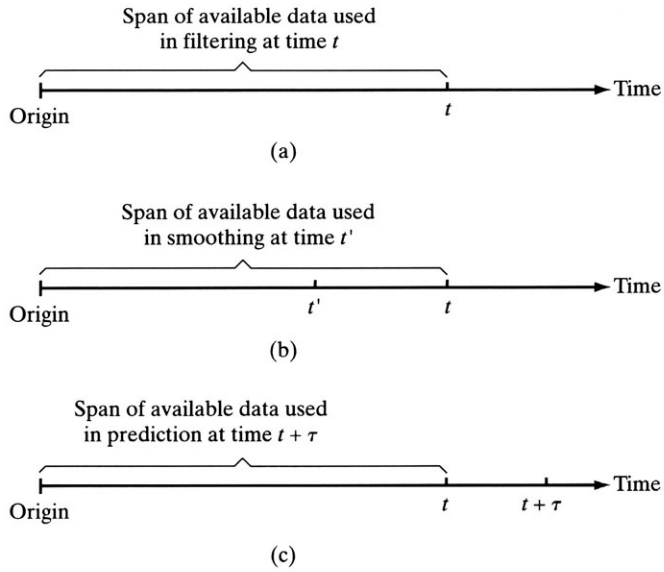

Figure 1: Three basic forms of estimation: (a) filtering, (b) smoothing and (c) prediction (Hay02)

The application of adaptive filters can be distinguished in four classes: identification (e.g. system identification), inverse modeling (e.g. equalization), prediction (e.g. tracking) and interference cancellation (e.g. beamforming).

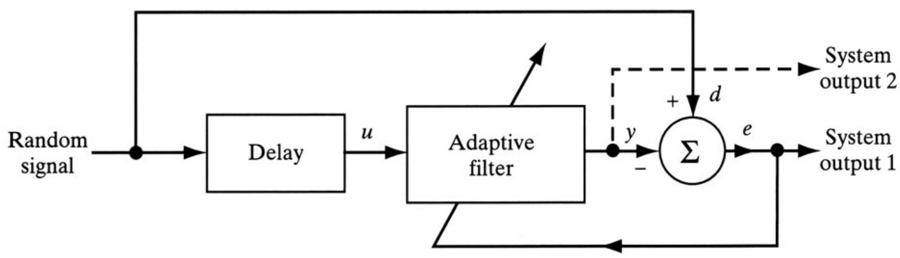

Figure 2: Prediction using an adaptive filter (Hay02)

This classification of adaptive filters gives a clearer view on the special use of the KF in prediction and tracking. The ideas behind adaptive filters help to understand the KF itself. For further reading on adaptive filters the reader is referred to (Hay02).

1.2 Advantages

The KF has various practical advantages which are the reason that it is found in a wide area of applications. The calculation of the filter algorithm is little complex because it only consists of five equations. In digital computing the recursive nature of the KF is its biggest advantage because it allows an easy and fast implementation. As digital computing and digital data processing are used in many fields the KF is a good solution to various problems in these fields. Because of its recursive nature the KF needs little memory. It only needs to store the last state of the system to calculate an estimate. Because the state and the covariance are always updated for the next iterative run of the KF the knowledge of all past states is contained in the corrected estimate and the updated error covariance. The KF has proven to be simple and robust. Robustness means it seldom looses track and seldom diverges from the true state. Thus it seldom needs to be reinitialized. Over years the KF has shown to be a reliable and good estimator. It exist a large and good knowledge of the KF, its function and application.

1.3 Applications

Most of the applications are found in the fields of automation and control. It has been used for a long time in radar and sonar systems as it can be used in target tracking in general. Today the most known application is GPS (Global Positioning System). GPS is a global satellite system that allows positioning and tracking of objects, e.g. for vehicle navigation. A new field of research is the use of the KF to estimate wireless radio channels. Estimation of the channel coefficients can be used to achieve higher transmission capacities in mobile communication. Other applications include for example weather forecasting and speech enhancement. The KF has found many applications in different fields. The little complexity and little memory requirement make the KF especially applicable in real time applications, e.g. for tracking in virtual reality.

1.4 Example: Vehicle Tracking

We want to develop a simple example of the KF. This example of vehicle tracking in a 2D plane (Kay93) provides a first idea how the KF is applied to a practical problem and how the system equations are designed. We assume a vehicle travelling at constant speed. Its movement is disturbed by small amounts of velocity correction and other influences, e.g. wind guts. The disturbances are seen in figure 3 as the true vehicle track differs from a straight line.

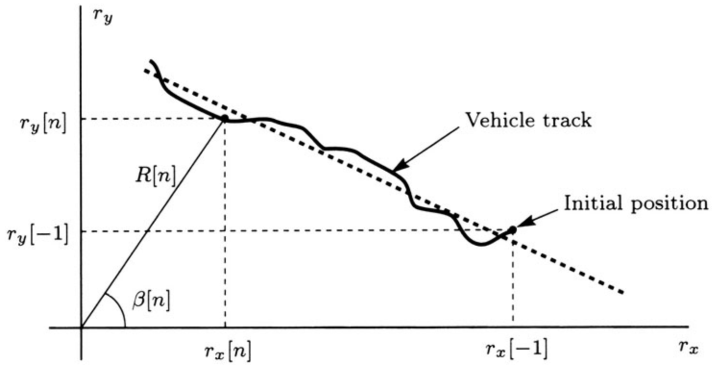

Figure 3: Track of a vehicle moving at constant speed (Kay93)

We need to design a state-space model of the vehicle movement. It is easy and illustrative to use a position-velocity model with two equations. Equation (1) describes the movement of the vehicle and with this equation the future position of the vehicle can be calculated. Δ is the time step from one state to the next one. Equation (2) describes the disturbance of the velocity by adding a random variable.

$$r[k] = r[k-1] + \Delta * v[k-1] \tag{1}$$

$$v[k] = v[k-1] + u[k] \tag{2}$$

The velocity disturbance is described by the covariance matrix Q. Disturbances and errors in the system description of the KF are always described by their covariance matrices.

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_u^2 & 0 \\ 0 & 0 & 0 & \sigma_u^2 \end{bmatrix} \tag{3}$$

Both equations are applied to the x- and y-coordinate which gives a matrix and vector equation of the complete vehicle movement.

$$\begin{bmatrix} r_x[k] \\ r_y[k] \\ v_x[k] \\ v_y[k] \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} r_x[k-1] \\ r_y[k-1] \\ v_x[k-1] \\ v_y[k-1] \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ u_x[k] \\ u_y[k] \end{bmatrix} \tag{4}$$

$$x_k = Ax_{k-1} + Bu_k \tag{5}$$

This matrix equation is a state-space model, which will be discussed in more detail in the next chapter 2. Equation (4) is used to predict a future position of the vehicle. The actual position of the vehicle is observed, i.e. it is measured. The distance R and the angle β which are seen from the origin of the coordinate plane are used to describe the observation vector (fig. 3). In figure 4 we see an observation which is disturbed by additive noise.

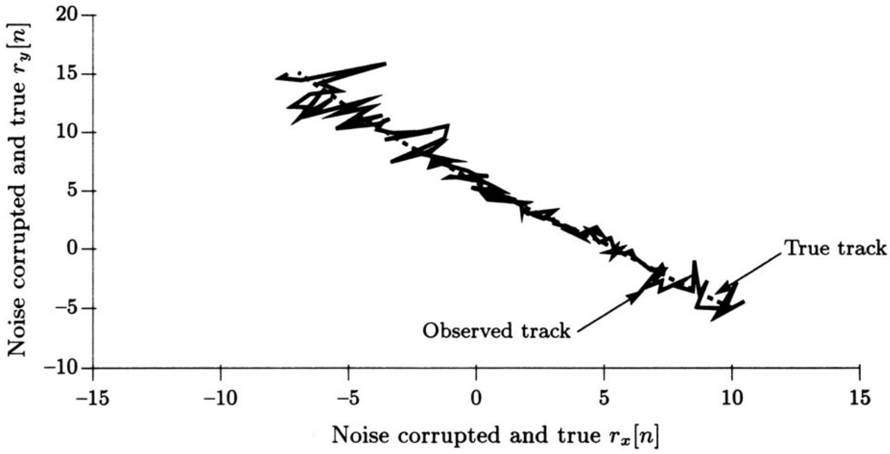

Figure 4: Observed vehicle track (Kay93)

The calculation of the distance R and the angle β lead to the following measurement model.

$$h(x_k) = \begin{bmatrix} R[k] \\ \beta[k] \end{bmatrix} = \begin{bmatrix} \sqrt{r_x^2[k] + r_y^2[k]} \\ \arctan \frac{r_y[k]}{r_x[k]} \end{bmatrix} \tag{6}$$

$$\tilde{z}_k = h(x_k) + v_k \tag{7}$$

It is clearly seen that the measurement model is non-linear, because the functions of the distance R and the angle β are non-linear. The disturbance of the observation is described by the covariance matrix R.

$$R = \begin{bmatrix} \sigma_R^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix} \tag{8}$$

In the case of non-linear system equations the Extended Kalman Filter (EKF) has to be used. In figure 5 we see the EKF estimate of the vehicle track.

It is seen that the EKF is initialized quite far from the true track but quickly approaches it and follows closely. It is usual to initialize the filter far away from the true track because at the beginning we do not have any knowledge about the true

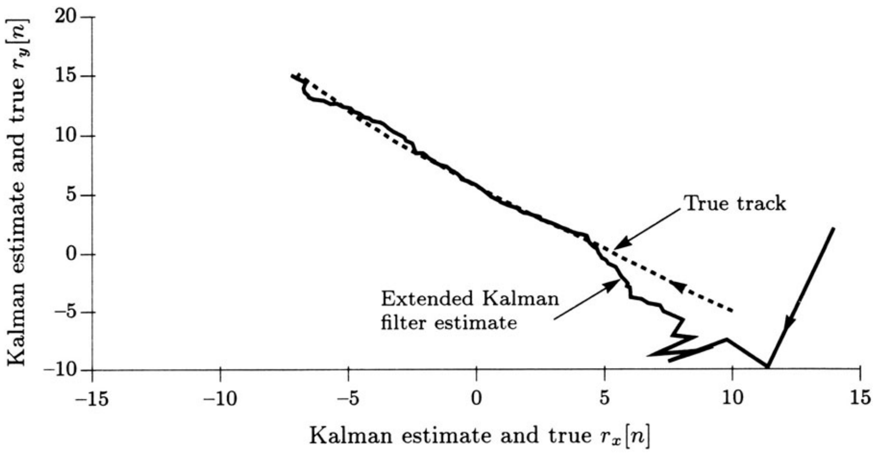

Figure 5: EKF estimate (Kay93)

state. It is a good property of the KF that it approaches the true track quickly.

2 System Model

The KF is based on the state-space model for stochastic processes. Actually the state-space form is a matrix and vector description of the ARMA model. ARMA means autoregressive moving average. The ARMA model has the same structure as an IIR filter and describes a system in the form of a transfer function and its coefficients a and b. A white discrete-time random process with zero mean is assumed to be the input $f(t)$ of the system and the coefficients can be calculated to obtain a desired output. The output $y(t)$ is a stochastic process and described by its autocorrelation function. The structure of the ARMA model using the state-space form is shown in figure 6.

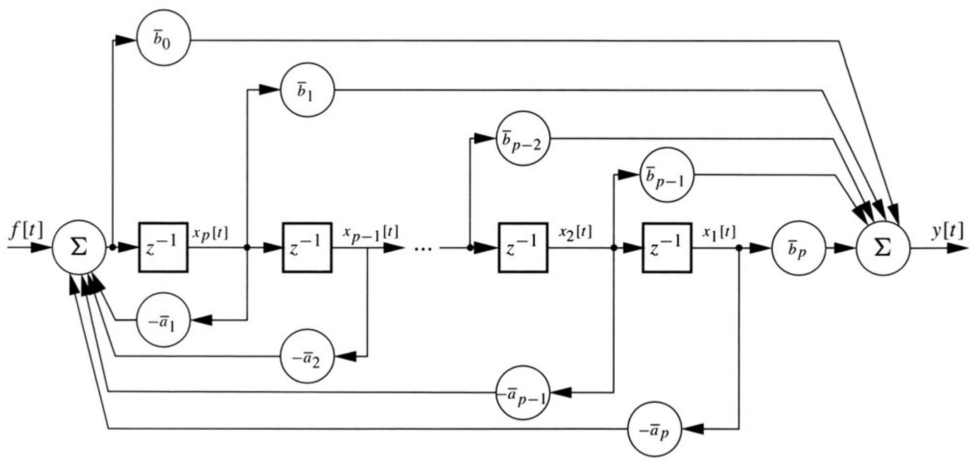

Figure 6: State-space form of the ARMA model (TKM00)

Using the ARMA and the state-space model the KF has to be discrete and to apply the KF to a practical problem it needs to be described in the form of the state-space model. This system design can be a difficult task as the model must represent the true behavior of the system accurately to obtain useful and realistic results. For a tracking problem it is straight forward to use a position-velocity (ch. 1.4) model which is quite simple and demonstrative.

2.1 General Assumptions

Designing the system and developing the KF some general assumptions have to be made. A dynamical system means a time-varying system. The state-space model is a stochastic and discrete model. Solutions for the continuous case have been developed but in practice the discrete case is of a much higher relevance due to digital data processing. First we want to develop the KF for the linear case only. It will later be extended to the non-linear case. Disturbances are always additive white Gaussian noise with zero mean. The noise is uncorrelated in time and different noises, e.g. in the process and measurement equations, are uncorrelated as well. With these assumptions the covariance matrices of the noise are just diagonal matrices with the variances of the different random variables on the main diagonal. The covariance matrices are calculated as follows.

$$Q_k = E(w_k w_k^T) \tag{9}$$

$$R_k = E(v_k v_k^T) \tag{10}$$

2.2 State-Space Model

For the KF two models are used, the process model and the measurement model. The process model (eq. (11)) describes the transformation from one state of the system to the next state. The state vector contains the different system variables and the matrix A can be interpreted as the transformation matrix from one state to the next one. In the tracking problem (ch. 1.4) this is the movement in time from the actual position to the future one. The matrix Bu_k is a controller input. With it a general influence on the system and the KF can be exerted. In the vehicle tracking example (ch. 1.4) we designed the velocity disturbance with this input. Finally the noise term w is added to the system which describes disturbances of different kinds. In tracking for example vibrations from outside sources or the tracking object itself may disturb the system.

$$x_k = Ax_{k-1} + Bu_k + w_{k-1} \tag{11}$$

The measurement model (eq. (12)) describes the relation of the state and the measurement. The matrix H depends on how the state is measured, e.g. optical or electrical. In tracking often an optical measurement of the position is used. The measurement model gives an estimate of the measurement and this estimate is compared with the actual true measurement in the KF. Every electrical measurement is disturbed by thermic noise and for this reason an additive noise term v is added to the equation.

$$z_k = Hx_k + v_k \tag{12}$$

2.3 Optimization

The KF solves a mathematical optimization problem. It optimizes the mean-square error of the estimation. The estimation error e_k is the difference between the estimated state $\hat{x}_k$ and the true state x_k.

$$e_k = x_k - \hat{x}_k \tag{13}$$

If the estimated state equals the true state of the system the error is zero. The estimation error is described by its covariance matrix P.

$$P_k = E(e_k e_k^T) \tag{14}$$

For the optimization the following cost function is used.

$$J_k = E[(x_k - \hat{x}_k)^2] \tag{15}$$

There are different ways to solve this optimization problem. Deriving its solution is a quite difficult task and the reader may be referred to (Hay02) and (TKM00). The solution of the optimization is the Kalman gain (KG). The KG is the weighting of the measurement residual and optimizes the estimate. It is used to correct the estimate and update the error covariance P for the next iterative run of the KF.

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \tag{16}$$

3 Discrete Kalman Filter

3.1 Time- & Measurement Update Equations

The equations of the Discrete Kalman Filter (DKF) fall into two groups: the time update equations and the measurement update equations. These two groups are also called the predictor and the corrector and compose a circular structure. The predictor projects forward the current state using the process model.

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k \tag{17}$$

It gives the a priori estimate $\hat{x}_k^-$ of the state. Also the error covariance P_k is projected forward to the next state.

$$P_k^- = AP_{k-1}A^T + Q \tag{18}$$

The corrector corrects the a priori estimate $\hat{x}_k^-$ and the error covariance P_k^-. For this correction it uses the actual measurement z_k which can be seen as a form of feedback control.

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \tag{19}$$

The corrected estimate is called the a posteriori estimate $\hat{x}_k$ and is used in the next iterative run of the KF to predict a new state. The corrector also updates the error covariance matrix for the next run.

$$P_k = (I - K_kH)P_k^- \tag{20}$$

Figure 7 shows the complete picture of the equations. It illustrates the recursive circular structure of the algorithm and the two groups of equations for the predictor and the corrector. In the first run the state and the error covariance need to be initialized. The corrector uses the a priori estimate $\hat{x}_k^-$ and the a priori error covariance P_k^- of the predictor for the correction. The predictor uses the a posteriori estimate $\hat{x}_k$ and the a posteriori error covariance P_k of the corrector for the estimate in the next run.

3.2 Measurement Residual

The corrector uses the measurement residual or measurement innovation to update the a priori estimate $\hat{x}_k^-$. The residual is the discrepancy between the estimated

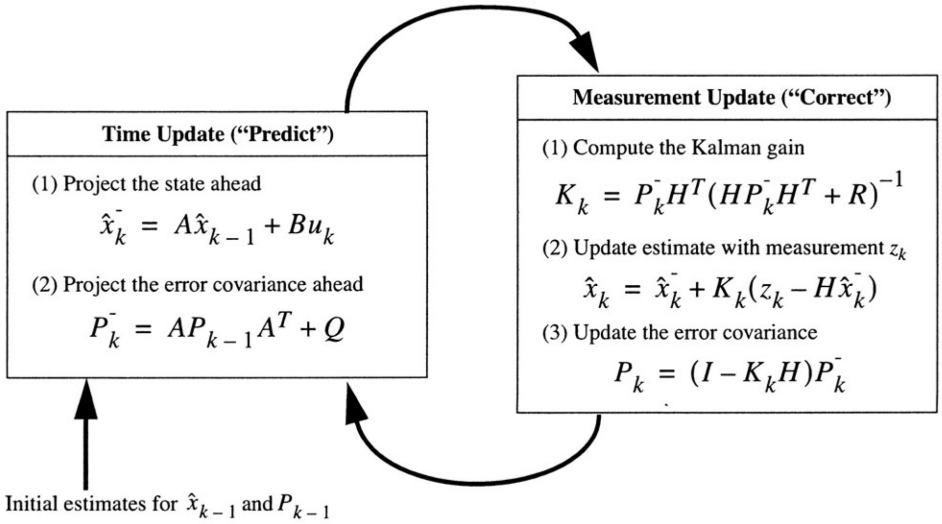

Figure 7: DKF equations (GW)

measurement using the measurement model and the true measurement z_k.

$$(z_k - H\hat{x}_k^-) \tag{21}$$

The residual is weighted by the KG (see the following chapter 3.3) which is the optimal weighting for the residual to minimize the estimation error. In other words the KG is the weighting to optimally correct the a priori estimate using the actual measurement.

3.3 Kalman Gain

The KG is the central factor in the calculation of the KF. It fulfills the optimization. If the error of the measurement R goes to zero the KG weights the residual more heavily. The actual measurement is trusted more and the a priori estimate is corrected more heavily.

$$\lim_{R_k \to 0} K_k = H^{-1} \tag{22}$$

If the error of the a priori estimate P goes to zero the KG goes to zero as well and the residual is weighted less heavily. The a priori estimate and the estimated measurement are trusted more and both are corrected less heavily. If the KG is zero there is no correction and the a priori and the a posteriori estimates and error covariances are the same.

$$\lim_{P_k^- \to 0} K_k = 0 \tag{23}$$

4 Extended Kalman Filter

4.1 Non-Linear State-Space Model

The Extended Kalman Filter (EKF) is the extension of the DKF to the non-linear case. If the process or the measurement model is non-linear the EKF has to be used instead of the formerly described DKF. The non-linear state-space model is as follows.

$$\tilde{x}_k = f(x_{k-1}, u_k, 0) \tag{24}$$

$$\tilde{z}_k = h(x_k, 0) \tag{25}$$

For the tracking problem (ch. 1.4) the measurement model is non-linear and the EKF has to be used.

4.2 Linearization

The difference in the calculation of the EKF and the DKF is small. In the non-linear case the state-space equations are linearized about the current mean and covariance. The linearization is done using partial derivatives and a Taylor expansion of the first order. Because the state-space model is in a matrix and vector form Jacobian matrices are used for the Taylor expansion. The following Jacobians need to be calculated for every run of the EKF algorithm.

$$A = \frac{\partial f}{\partial x}(\hat{x}_{k-1}, u_k, 0) \tag{26}$$

$$H = \frac{\partial h}{\partial x}(\tilde{x}_k, 0) \tag{27}$$

$$W = \frac{\partial f}{\partial w}(\hat{x}_{k-1}, u_k, 0) \tag{28}$$

$$V = \frac{\partial h}{\partial v}(\tilde{x}_k, 0) \tag{29}$$

This gives the following linearized state-space equations.

$$x_k \approx \tilde{x}_k + A(x_{k-1} - \hat{x}_{k-1}) + W w_{k-1} \tag{30}$$

$$z_k \approx \tilde{z}_k + H(x_k - \tilde{x}_k) + V v_k \tag{31}$$

For the non-linear state-space model and with the Jacobians for the linearization the equations for the EKF are as follows. Apart from those two changes the equations are the same as for the DFK.

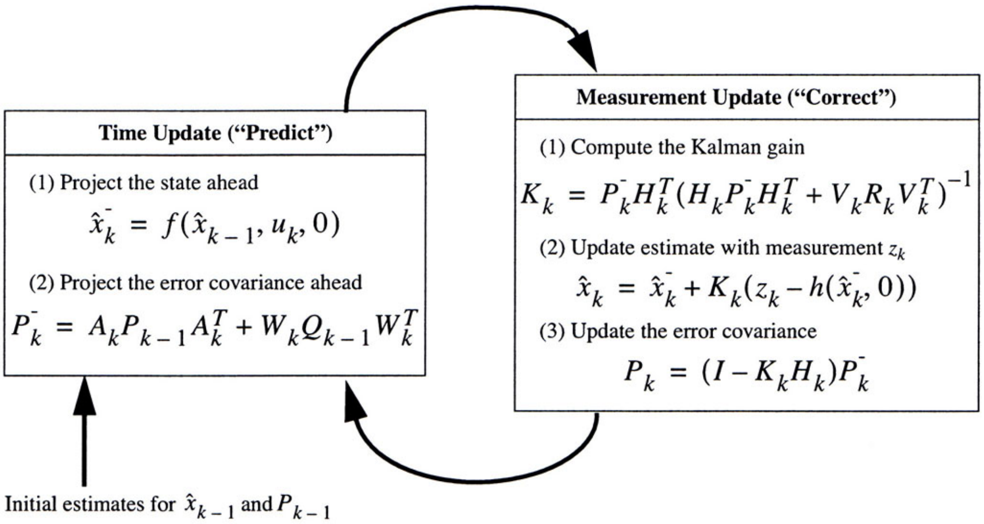

Figure 8: EKF equations (GW)

4.3 Unscented Kalman Filter

The linear approximation for the EKF introduces an error in the calculation of the estimate. What actually is problematic is the assumption of Gaussian distribution for the random variables. This general assumption is problematic in the non-linear case because the first order linearization of the EKF can introduce large errors in the true posterior mean and covariance of the Gaussian random variables. The assumption of a Gaussian distribution does not hold any longer. A solution to this problem has been developed by Julier and Uhlman (SJJ97). It uses a deterministic sampling approach. In this approach sampling points are carefully chosen to capture the posterior mean and covariance with second-order accuracy. Even higher orders can be achieved but second-order accuracy is adequate for the assumption of Gaussian distribution. This solution is called the Unscented Kalman Filter (UKF) and it achieves a higher performance than the EKF at the same complexity. The UKF is a good example for the development of an advanced high performance KF.

5 HiBall Tracking System

The HiBall tracking system is an electro-optical wide-area tracking system. It can cover up to 12 by 12 meters of space and was developed at the University of North Carolina at Chapel Hill in the 90s. It is today marketed by 3rdTech. The tracking system consists of three main components: the HiBall which is the electro-optical sensor, the Ceiling which consists of ceiling panels with integrated LEDs and the Ceiling-HiBall Interface Board (CIB) which provides communication and synchronization between the HiBall, the Ceiling and the host computer.

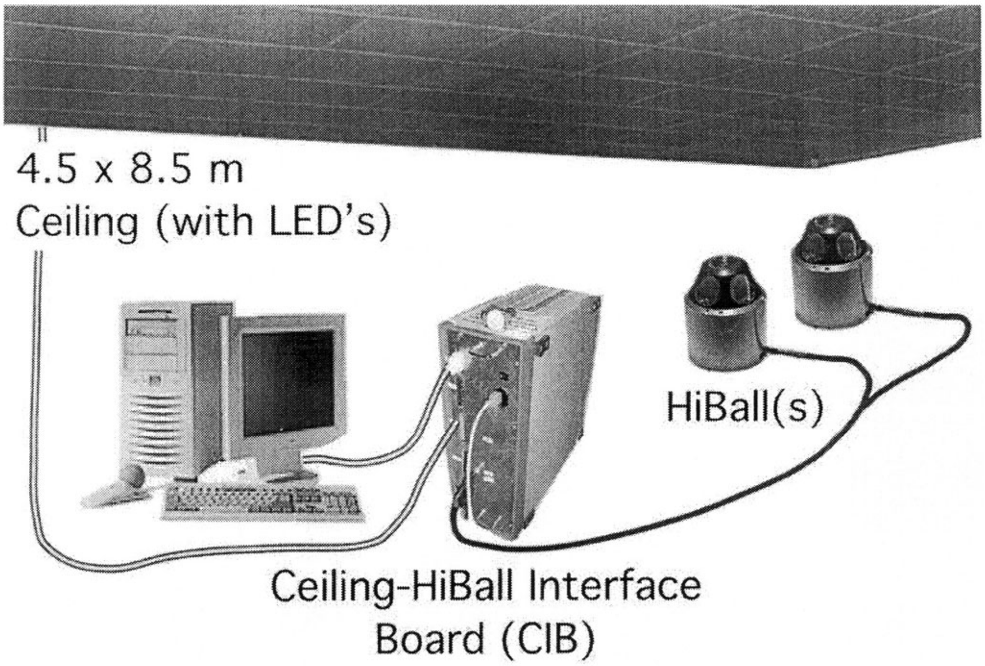

Figure 9: Components of the HiBall tracking system (GW01)

5.1 What is the HiBall?

The HiBall is an electro-optical sensor which is tracked by the system. For example in virtual reality a person carrying the sensor can be tracked and move in a virtual environment. The HiBall has six lenses and six electro-optical sensors which provide 26 fields of view in different directions. It measures the light that is emitted from the LEDs in the Ceiling.

5.2 SCAAT

The HiBall tracking system uses the SCAAT (single constraint at a time) algorithm. This algorithm contains an EKF for tracking. To understand the new idea behind the

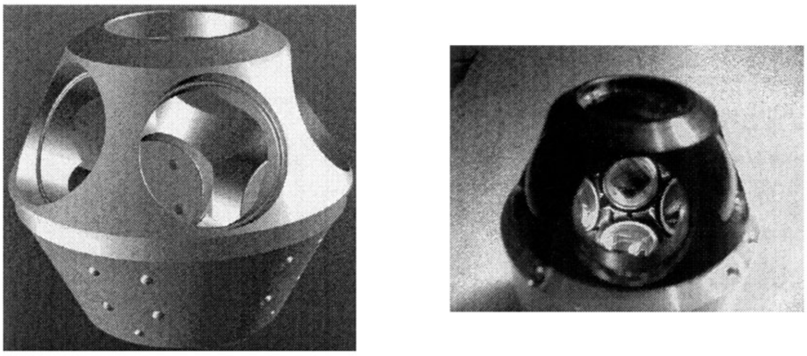

Figure 10: HiBall electro-optical sensor (GW01)

SCAAT algorithm, the use of the EKF and the resulting advantages the traditional tracking problem is reviewed. The tracking problem implies a system of equations for the unknown variables. For a tracking we have at least six unknowns, three for the position in space and three for the orientation. To solve a system of equations it has to be determined and this means that in tracking at least six measurements are needed. Usually those measurements are collected over a period of time and not taken simultaneously. The collection of measurements can introduce significant error in the solution because they are not taken at one point due to the movement of the sensor. This violates the simultaneity assumptions for solving the system of equations.

The idea of the SCAAT algorithm is to use every single measurement at a time because every measurement contains information about the position of the sensor. This way the mathematical solution of the system of equations is underdetermined and can not be solved. The idea is to use the EKF to improve the estimate incrementally step by step. The future position of the sensor is estimated with the EKF and then compared with a measurement of the position. The HiBall system updates the position of the estimate 2000 times a second, i.e. it takes 2000 measurements a second. Using the EKF the HiBall tracking system adopts the general advantages of the KF. The algorithm is of low complexity and therefore runs faster. This improves the latency and the accuracy of the tracking system. Because the algorithm runs faster measurements are taken more frequently and thus there is less motion of the sensor between different measurement and estimation cycles which further improves the accuracy. The SCAAT algorithm contains more features than just the EKF, e.g. it includes an auto-calibration (Wel96).

The algorithm runs through with the following steps which are related to the KF. First the position of the sensor is estimated using the EKF. The process model of the EKF is a linear position-velocity model like we have seen it in the example for vehicle tracking (ch. 1.4).

$$x(t) = A(\Delta t)x(t - \Delta t) + w(t) \tag{32}$$

For the estimated position one field of view is chosen randomly. For the chosen field of view a region on the Ceiling is located and in this region one LED is flashed. A measurement of the light emitted by the LED is taken with the sensor and the taken measurement is compared with the estimated measurement. The measurement model of the system is non-linear and therefore the EKF is used.

$$z_k = h_k(x(t), \alpha(t), b(t), c(t)) + v_k(t) \tag{33}$$

In the measurement model $b(t)$ and $c(t)$ are the LED and the sensor parameters. α is the external orientation quaternion which gives the orientation of the sensor. The models are described in detail in (Wel96). Finally the error covariance and the estimate are updated for the next run of the EKF.

5.3 Advantages & Applications

The HiBall tracking system has several advantages which are based on the use of the EKF (3rd05). It has a low latency of 1 msec which allows fast tracking. The resolution of the system is 0.2 mm in position and 0.01 degrees in orientation and its accuracy is 0.4 mm in position and 0.02 degrees in orientation. This high resolution and accuracy allow its use in applications like industrial tool tracking. The system is robust, i.e. it seldom looses track and seldom needs to be reinitialized. It is flexible in its application because it consists of different modules and can be expanded for a wider range later. The HiBall tracking system is especially applicable in virtual reality, e.g. in the demonstration of architectural models, or in augmented reality, e.g. in medicine.

6 Conclusions

The KF is a very practical tool. Its success is based on its practical and easy implementation in digital computing because of its recursive nature and little complexity. As computing and digital data processing are used in many fields the KF is a good solution to many problems. It needs little memory which makes it fast and applicable in real-time applications like tracking. Over 40 years the KF has proven in many applications that it is a robust and reliable estimator. Prediction and tracking are major fields of the KF but not its only applications and functions. Many advanced KF have been developed for special applications and high performance. The large knowledge of the KF, its application and function gives a reliable basis for its use.

References

[3rd05] 3rdTech: HiBall-3000 Wide-Area Tracker and 3D Digitizer. (2005). – http://www.3rdtech.com/HiBall.htm

[AHS94] A. H. Sayed, T.Kailath: A state-space approach to adaptive RLS filtering. In: *IEEE Signal Processing Magazine, Vol. 11* (1994)

[GW] Greg Welch, Gary B.: An Introduction to the Kalman Filter. . – http://www.cs.unc.edu/˜welch/kalman/

[GW01] Greg Welch, Gary B.: High-Performance Wide-Area Optical Tracking. (2001). – http://www.cs.unc.edu/˜tracker/

[Hay02] Haykin, Simon: *Adaptive Filter Theory*. Prentice Hall, 2002

[Kal60] Kalman, Rudolph E.: A New Approach to Linear Filtering and Prediction Problems. (1960)

[Kay93] Kay, Steven M.: *Fundamentals of Statistical Signal Processing, Estimation Theory*. Prentice Hall, 1993

[SJJ97] Simon J. Julier, Jeffrey K. U.: A New Extension of the Kalman Filter to Nonlinear Systems. (1997)

[TKM00] Todd K. Moon, Wynn C. S.: *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000

[Wel96] Welch, Gregory F.: *SCAAT: Incremental Tracking with Incomplete Information*, UNC-Chapel Hill Dept. of Computer Science, Diss., 1996