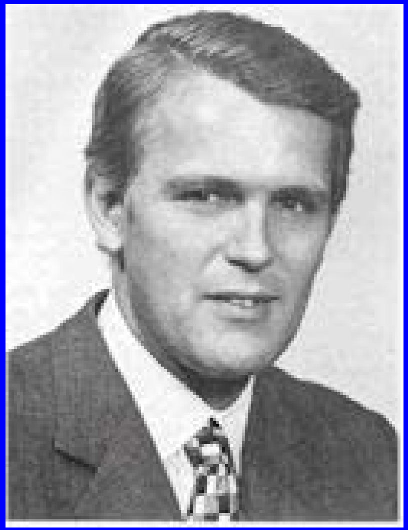
Lecture-14

Kalman Filter

Rudolf Kalman

- Born May 19, 1930, Budapest, Hungary.
- BS (1953) and MS (1954) from MIT.
- Ph.D. in 1957 from Columbia under Professor J. R. Ragzzini.
- 1957-1958 IBM
- 1958 -1964 Research Institute for Advanced Study (RIAS), Baltimore, Maryland
  - RIAS was part of Glenn L. Martin Co, which became Lockheed Martin Corp.
- 1964 Stanford University
- 1971 University of Florida



A New Approach to Linear Filtering and Prediction Problems[1]

R. E. KALMAN
Research Institute for Advanced Study,[2]
Baltimore, Md.

The classical filtering and prediction problem is re-examined using the Bode-Shannon representation of random processes and the "state transition" method of analysis of dynamic systems. New results are:

(1) The formulation and methods of solution of the problem apply without modification to stationary and nonstationary statistics and to growing-memory and infinite-memory filters.

(2) A nonlinear difference (or differential) equation is derived for the covariance matrix of the optimal estimation error. From the solution of this equation the coefficients of the difference (or differential) equation of the optimal linear filter are obtained without further calculations.

(3) The filtering problem is shown to be the dual of the noise-free regulator problem.

The new method developed here is applied to two well-known problems, confirming and extending earlier results.

The discussion is largely self-contained and proceeds from first principles; basic concepts of the theory of random processes are reviewed in the Appendix.

Introduction

An important class of theoretical and practical problems in communication and control is of a statistical nature. Such problems are: (i) Prediction of random signals; (ii) separation of random signals from random noise; (iii) detection of signals of known form (pulses, sinusoids) in the presence of random noise.

In his pioneering work, Wiener [1][3] showed that problems (i) and (ii) lead to the so-called Wiener-Hopf integral equation; he also gave a method (spectral factorization) for the solution of this integral equation in the practically important special case of stationary statistics and rational spectra.

Many extensions and generalizations followed Wiener's basic work. Zadeh and Ragazzini solved the finite-memory case [2]. Concurrently and independently of Bode and Shannon [3], they also gave a simplified method [2] of solution. Booton discussed the nonstationary Wiener-Hopf equation [4]. These results are now in standard texts [5-6]. A somewhat different approach along these main lines has been given recently by Darlington [7]. For extensions to sampled signals, see, e.g., Franklin [8], Lees [9]. Another approach based on the eigenfunctions of the Wiener-Hopf equation (which applies also to nonstationary problems whereas the preceding methods in general don't), has been pioneered by Davis [10] and applied by many others, e.g., Shinbrot [11], Blum [12], Pugachev [13], Solodovnikov [14].

In all these works, the objective is to obtain the specification of a linear dynamic system (Wiener filter) which accomplishes the prediction, separation, or detection of a random signal.[4]

Present methods for solving the Wiener problem are subject to a number of limitations which seriously curtail their practical usefulness:

(1) The optimal filter is specified by its impulse response. It is not a simple task to synthesize the filter from such data.

(2) Numerical determination of the optimal impulse response is often quite involved and poorly suited to machine computation. The situation gets rapidly worse with increasing complexity of the problem.

(3) Important generalizations (e.g., growing-memory filters, nonstationary prediction) require new derivations, frequently of considerable difficulty to the nonspecialist.

(4) The mathematics of the derivations are not transparent. Fundamental assumptions and their consequences tend to be obscured.

This paper introduces a new look at this whole assemblage of problems, sidestepping the difficulties just mentioned. The following are the highlights of the paper:

(5) *Optimal Estimates and Orthogonal Projections.* The Wiener problem is approached from the point of view of conditional distributions and expectations. In this way, basic facts of the Wiener theory are quickly obtained; the scope of the results and the fundamental assumptions appear clearly. It is seen that all statistical calculations and results are based on first and second order averages; no other statistical data are needed. Thus difficulty (4) is eliminated. This method is well known in probability theory (see pp. 75–78 and 148–155 of Doob [15] and pp. 455–464 of Loève [16]) but has not yet been used extensively in engineering.

(6) *Models for Random Processes.* Following, in particular, Bode and Shannon [3], arbitrary random signals are represented (up to second order average statistical properties) as the output of a linear dynamic system excited by independent or uncorrelated random signals ("white noise"). This is a standard trick in the engineering applications of the Wiener theory [2–7]. The approach taken here differs from the conventional one only in the way in which linear dynamic systems are described. We shall emphasize the concepts of *state* and *state transition;* in other words, linear systems will be specified by systems of first-order difference (or differential) equations. This point of view is

[1] This research was supported in part by the U. S. Air Force Office of Scientific Research under Contract AF 49 (638)-382.

[2] 7212 Bellona Ave.

[3] Numbers in brackets designate References at end of paper.

[4] Of course, in general these tasks may be done better by nonlinear filters. At present, however, little or nothing is known about how to obtain (both theoretically and practically) these nonlinear filters.

Contributed by the Instruments and Regulators Division and presented at the Instruments and Regulators Conference, March 29– April 2, 1959, of The American Society of Mechanical Engineers.

Note: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, February 24, 1959. Paper No. 59—IRD-11.

Main Points

- Very useful tool.
- It produces an optimal estimate of the state vector based on the noisy measurements (observations).
- For the state vector it also provides confidence (certainty) measure in terms of a covariance matrix.
- It integrates an estimate of state over time.
- It is a sequential state estimator.

State-Space Model

State model error
With covariance
Q(k)

State-transition equation

$$\mathbf{z}(k) = \Phi(k, k-1)\mathbf{z}(k-1) + \mathbf{w}(k)$$

State Vector

Measurement (observation) equation

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{z}(k) + \mathbf{v}(k)$$

Observation
Noise with covariance
R(k)

Measurement Vector

Kalman Filter Equations

State Prediction

$$\hat{\mathbf{z}}_b(k) = \Phi(k, k-1)\hat{\mathbf{z}}_a(k-1)$$

Covariance Prediction

$$\mathbf{P}_b(k) = \Phi(k, k-1)\mathbf{P}_a(k-1)\Phi^T(k, k-1) + \mathbf{Q}(k)$$

Kalman Gain

$$\mathbf{K}(k) = \mathbf{P}_b(k)\mathbf{H}^T(k)(\mathbf{H}(k)\mathbf{P}_b(k)\mathbf{H}^T(k) + \mathbf{R}(k))^{-1}$$

State-update

$$\hat{\mathbf{z}}_a(k) = \hat{\mathbf{z}}_b(k) + \mathbf{K}(k)[\mathbf{y}(k) - \mathbf{H}(k)\hat{\mathbf{z}}_b(k)]$$

Covariance-update

$$\mathbf{P}_a(k) = \mathbf{P}_b(k) - \mathbf{K}(k)\mathbf{H}(k)\mathbf{P}_b(k)$$

Two Special Cases

- Steady State

$$\Phi(k,k-1)=\Phi$$

$$\mathbf{Q}(k)=\mathbf{Q}$$

$$\mathbf{H}(k)=\mathbf{H}$$

$$\mathbf{R}(k)=\mathbf{R}$$

- Recursive least squares

$$\Phi(k,k-1)=\mathbf{I}$$

$$\mathbf{Q}(k)=0$$

Comments

- In some cases, state transition equation and the observation equation both may be non-linear.
- We need to linearize these equation using Taylor series.

Extended Kalman Filter

$$\mathbf{z}(k) = \mathbf{f}(\mathbf{z}(k-1)) + \mathbf{w}(k)$$

$$\mathbf{y}(k) = \mathbf{h}(\mathbf{z}(k)) + \mathbf{v}(k)$$

$$\mathbf{f}(\mathbf{z}(k-1)) \approx \mathbf{f}(\hat{\mathbf{z}}_a(k-1)) + \frac{\partial \mathbf{f}(\mathbf{z}(k-1))}{\partial \mathbf{z}(k-1)}(\mathbf{z}(\mathrm{k}-1) - \hat{\mathbf{z}}_a(k-1))$$

Taylor series

$$\mathbf{h}(\mathbf{z}(k)) \approx \mathbf{h}(\hat{\mathbf{z}}_b(k)) + \frac{\partial \mathbf{h}(\mathbf{z}(k))}{\partial \mathbf{z}(k)}(\mathbf{z}(k) - \hat{\mathbf{z}}_b(k-1))$$

Extended Kalman Filter

$$\mathbf{z}(k) = \mathbf{f}(\mathbf{z}(k-1)) + \mathbf{w}(k)$$

$$\mathbf{z}(k) = \mathbf{f}(\hat{\mathbf{z}}_a(k-1)) + \frac{\partial \mathbf{f}(\mathbf{z}(k-1))}{\partial \mathbf{z}(k-1)}(\mathbf{z}(\mathrm{k}-1) - \hat{\mathbf{z}}_a(k-1)) + \mathbf{w}(k)$$

$$\mathbf{z}(k) \approx \Phi(k, k-1)\mathbf{z}(k-1) + \mathbf{u}(k) + \mathbf{w}(k)$$

$$\mathbf{u}(k) = \mathbf{f}(\hat{\mathbf{z}}_a(k-1)) - \Phi(k, k-1)\hat{\mathbf{z}}_a(k-1)$$

$$\Phi(k, k-1) = \frac{\partial \mathbf{f}(\mathbf{z}(k-1))}{\partial \mathbf{z}(k-1)}$$

Extended Kalman Filter

$$\mathbf{y}(k) = \mathbf{h}(\mathbf{z}(k)) + \mathbf{v}(k)$$

$$\mathbf{y}(k) = \mathbf{h}(\hat{\mathbf{z}}_b(k)) + \frac{\partial \mathbf{h}(\mathbf{z}(k))}{\partial \mathbf{z}(k)}(\mathbf{z}(k) - \hat{\mathbf{z}}_b(k-1)) + \mathbf{v}(k)$$

$$\tilde{\mathbf{y}}(k) \approx \mathbf{H}(k)\mathbf{z}(k) + \mathbf{v}(k)$$

$$\tilde{\mathbf{y}}(k) = \mathbf{y}(k) - \mathbf{h}(\hat{\mathbf{z}}_b(k)) + \mathbf{H}(k)\hat{\mathbf{z}}_b(k)$$

$$\mathbf{H}(k) = \frac{\partial \mathbf{h}(\mathbf{z}(k))}{\partial \mathbf{z}(k)}$$

Multi-Frame Feature Tracking

Application of Kalman Filter

- Assume feature points have been detected in each frame.
- We want to track features in multiple frames.
- Kalman filter can estimate the position and uncertainty of feature in the next frame.
  - Where to look for a feature
  - how large a region should be searched

$$\mathbf{p}_k = [x_k, y_k]^T$$ Location

$$\mathbf{v}_k = [u_k, v_k]^T$$ Velocity

$$\mathbf{Z} = [x_k, y_k, u_k, v_k]^T$$ State Vector

System Model

$$\mathbf{p}_k = [x_k, y_k]^T \qquad \mathbf{p}_k = \mathbf{p}_{k-1} + \mathbf{v}_{k-1} + \xi_{k-1}$$

noise

$$\mathbf{v}_k = [u_k, v_k]^T \qquad \mathbf{v}_k = \mathbf{v}_{k-1} + \eta_{k-1}$$

$$\mathbf{Z} = [x_k, y_k, u_k, v_k]^T \qquad \mathbf{Z}_k = \Phi_{k-1}\mathbf{Z}_{k-1} + \mathbf{w}_{k-1}$$

$$\Phi_{k-1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{w}_{k-1} = \begin{bmatrix} \xi_{k-1} \\ \eta_{k-1} \end{bmatrix}$$

Measurement Model

$$\mathbf{p}_k = [x_k, y_k]^T$$

$$\mathbf{v}_k = [u_k, v_k]^T$$

$$\mathbf{z} = [x_k, y_k, u_k, v_k]^T$$

$$\mathbf{y}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_k \\ \mathbf{v}_k \end{bmatrix} + \mu_k$$

$$\mathbf{y}_k = \mathbf{H} \begin{bmatrix} \mathbf{p}_k \\ \mathbf{v}_k \end{bmatrix} + \mu_k$$

Measurement matrix

Kalman Filter Equations

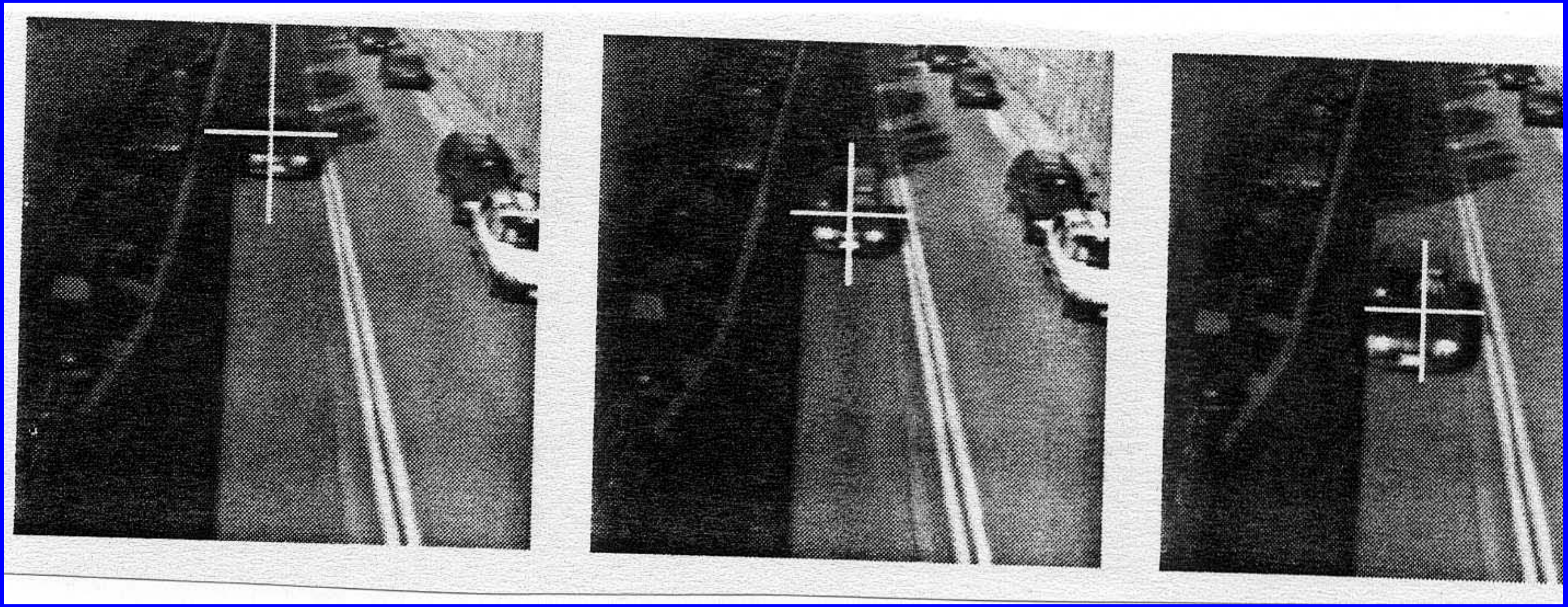
State Prediction $$\hat{\mathbf{z}}_b(k) = \Phi(k, k-1)\hat{\mathbf{z}}_a(k-1)$$

Covariance Prediction $$\mathbf{P}_b(k) = \Phi(k, k-1)\mathbf{P}_a(k-1)\Phi^T(k, k-1) + \mathbf{Q}(k)$$

Kalman Gain $$\mathbf{K}(k) = \mathbf{P}_b(k)\mathbf{H}^T(k)(\mathbf{H}(k)\mathbf{P}_b(k)\mathbf{H}^T(k) + \mathbf{R}(k))^{-1}$$

State-update $$\hat{\mathbf{z}}_a(k) = \hat{\mathbf{z}}_b(k) + \mathbf{K}(k)[\mathbf{y}(k) - \mathbf{H}(k)\hat{\mathbf{z}}_b(k)]$$

Covariance-update $$\mathbf{P}_a(k) = \mathbf{P}_b(k) - \mathbf{K}(k)\mathbf{H}(k)\mathbf{P}_b(k)$$



Kalman Filter: Relation to Least Squares

Estimate state such that the following is minimized:

-first term: initial estimate weighted by corresponding covariance

-second term: other measurements weighted by corresponding covariances

$$C = (\hat{\mathbf{Z}}_0 - \mathbf{Z})^T P_0^{-1} (\hat{\mathbf{Z}}_0 - \mathbf{Z}) + \sum_{i=1}^{k} (\mathbf{Y}_i - H_i \mathbf{Z})^T W^{-1}{}_i (\mathbf{Y}_i - H_i \mathbf{Z})$$

minimize

$$\hat{\mathbf{Z}} = [P_0^{-1} + \sum_{i=1}^{k} H_i{}^T W_i^{-1} H_i]^{-1} [P_0^{-1} \hat{\mathbf{Z}}_0 + \sum_{i=1}^{k} H_i{}^T W_i^{-1} \mathbf{Y}_i]$$

Batch Mode

Kalman Filter: Relation to Least Squares

$$\hat{\mathbf{Z}}_k = [P_0^{-1} + \sum_{i=1}^{k} H_i^T W_i^{-1} H_i]^{-1}[P_0^{-1}\hat{\mathbf{Z}}_0 + \sum_{i=1}^{k} H_i^T W_i^{-1} \mathbf{Y}_i]$$

$$\hat{\mathbf{Z}}_{k-1} = [P_0^{-1} + \sum_{i=1}^{k-1} H_i^T W_i^{-1} H_i]^{-1}[P_0^{-1}\hat{\mathbf{Z}}_0 + \sum_{i=1}^{k-1} H_i^T W_i^{-1} \mathbf{Y}_i]$$

Recursive Mode

Kalman Filter: Relation to Least Squares

$$\mathbf{Z}_k = \mathbf{Z}_{k-1} + K_k(Y_k - H_k\mathbf{Z}_{k-1})$$

$$K_k = P_{k-1}H^T{}_k\ (W_k + H_kP_{k-1}H_k{}^T)^{-1}$$

$$P_k = (I - K_kH_k)P_{k-1}$$

$$\Phi(k,k-1) = \mathbf{I}$$

$$\mathbf{Q}(k) = 0$$

Kalman Filter (Least Squares)

State Prediction

$$\hat{\mathbf{z}}_b(k) = \Phi(k, k-1)\hat{\mathbf{z}}_a(k-1)$$

$$\hat{\mathbf{z}}_b(k) = \hat{\mathbf{z}}_a(k-1)$$

Covariance Prediction

$$\mathbf{P}_b(k) = \Phi(k, k-1)\mathbf{P}_a(k-1)\Phi^T(k, k-1) + \mathbf{Q}(k)$$

$$\mathbf{P}_b(k) = \mathbf{P}_a(k-1)$$

Kalman Gain

$$\mathbf{K}(k) = \mathbf{P}_b(k)\mathbf{H}^T(k)(\mathbf{H}(k)\mathbf{P}_b(k)\mathbf{H}^T(k) + \mathbf{R}(k))^{-1}$$

$$\mathbf{K}(k) = \mathbf{P}_b(k)\mathbf{H}^T(k)(\mathbf{H}(k)\mathbf{P}_b(k)\mathbf{H}^T(k) + \mathbf{W}(k))^{-1}$$

Kalman Filter (Least Squares)

State-update

$$\hat{\mathbf{z}}_a(k) = \hat{\mathbf{z}}_b(k) + \mathbf{K}(k)[\mathbf{y}(k) - \mathbf{H}(k)\hat{\mathbf{z}}_b(k)]$$

$$\hat{\mathbf{z}}\ (k) = \hat{\mathbf{z}}\ (k-1) + \mathbf{K}(k)[\mathbf{y}(k) - \mathbf{H}(k)\hat{\mathbf{z}}\ (k-1)]$$

Covariance-update

$$\mathbf{P}_a(k) = \mathbf{P}_b(k) - \mathbf{K}(k)\mathbf{H}(k)\mathbf{P}_b(k)$$

$$\mathbf{P}\ (k) = \mathbf{P}\ (k-1) - \mathbf{K}(k)\mathbf{H}(k)\mathbf{P}\ (k-1)$$

Computing Motion Trajectories

Algorithm For Computing Motion Trajectories

- Compute tokens using Moravec’s interest operator (intensity constraint).
- Remove tokens which are not interesting with respect to motion (optical flow constraint).
  - Optical flow of a token should differ from the mean optical flow around a small neighborhood.

Algorithm For Computing Motion Trajectories

- Link optical flows of a token in different frames to obtain motion trajectories.
  - Use optical flow at a token to predict its location in the next frame.
  - Search in a small neighborhood around the predicted location in the next frame for a token.
- Smooth motion trajectories using Kalman filter.

Kalman Filter (Ballistic Model)

$$x(t) = .5a_x t^2 + v_x t + x_0 \qquad \mathbf{Z} = (a_x, a_y, v_x, v_y)$$

$$y(t) = .5a_y t^2 + v_y t + y_0 \qquad \mathbf{y} = (x(t), y(t))$$

$$f(\mathbf{Z}, \mathbf{y}) = (x(t) - .5a_x t^2 - v_x t - x_0, y(t) - .5a_y t^2 - v_y t - y_0)$$

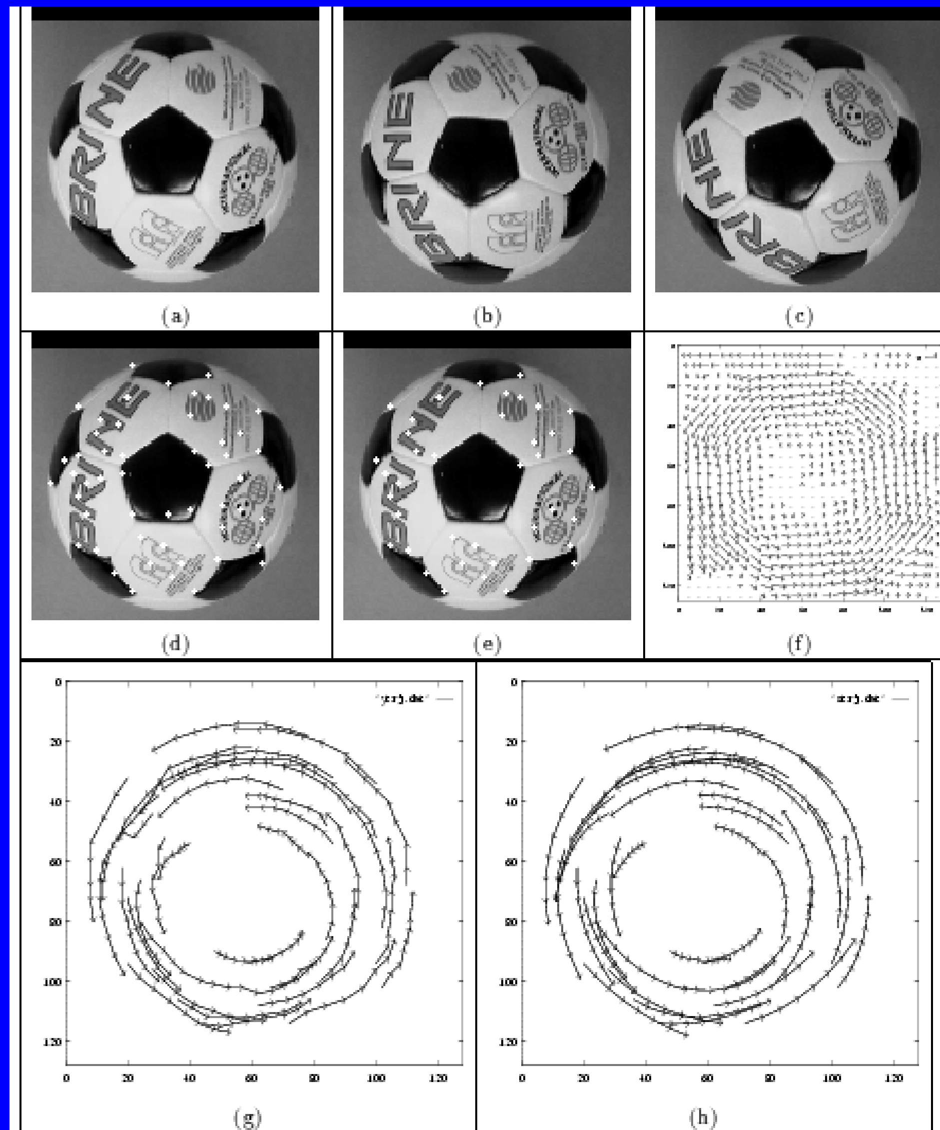
$$\frac{\partial f}{\partial \mathbf{Z}} = \begin{bmatrix} -.5t^2 & 0 & -t & 0 \\ 0 & -.5t^2 & 0 & -t \end{bmatrix}$$

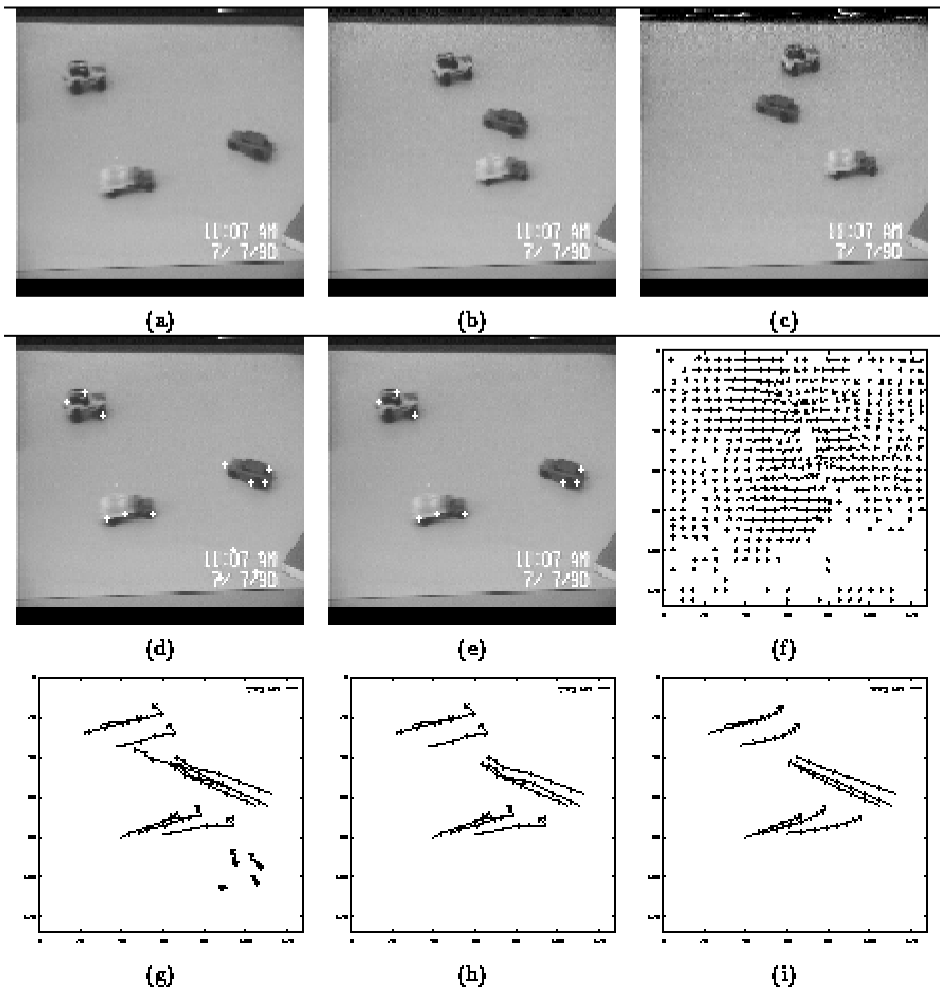
Kalman Filter (Ballistic Model)

$$\mathbf{Z}(k) = \mathbf{Z}(k-1) + K(k)(Y(k) - H(k)\mathbf{Z}(k-1))$$

$$K(k) = P(k-1)H^T(k)\ \ (W(k) + H^T P(k-1)H^T(k))^{-1}$$

$$P(k) = (I - K(k)H(k))P(k-1)$$


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)



(a) (b) (c) (d) (e) (f) (g) (h) (i)

2008 Charles Stark Draper Prize

For the development and dissemination of the optimal digital technique (known as the Kalman Filter) that is pervasively used to control a vast array of consumer, health, commercial and defense products.