Tracking Basics

Christopher R. Wren

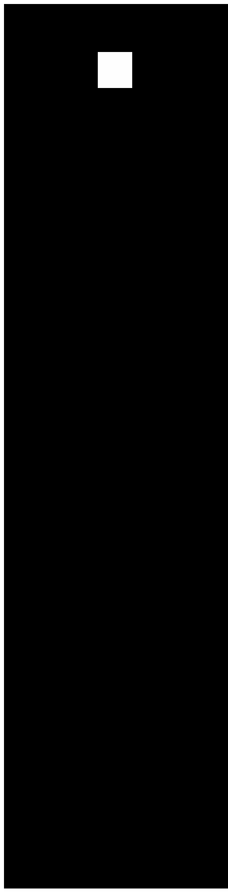
Mitsubishi Electric Research Laboratories
Research Laboratory

Overview

- Sum of Squared Differences Tracker
  - A simple example and a noisy failure
- Motion as a Queue
  - Linear Dynamic Systems and Stochastic Systems
- The Kalman Filter
  - Simple examples and failures
- Multi-mode Trackers (really quick)
- Affecting the Observation Process
  - Examples
  - Research system
- Beyond Tracking
  - Innovations for Behavior Modeling

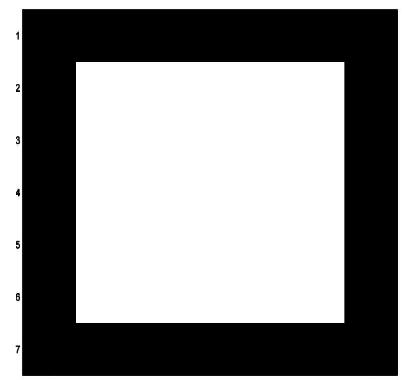
http://www.drwren.com/chris/9.913/clean.avi

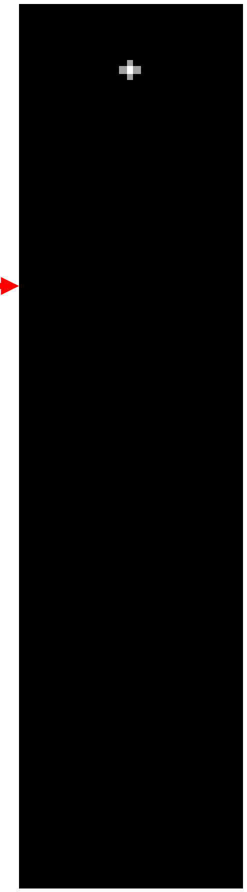
Sum of Squared Differences



http://www.drwren.com/chris/9.913/clean.avi

$$\sum_{i=x}^{x+h} \sum_{j=y}^{y+w} (I(i,j) - T(i,j))^2$$

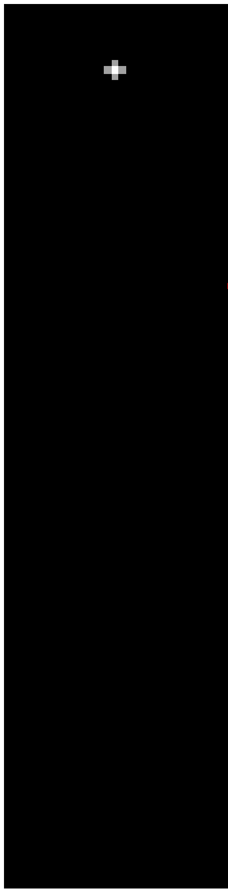




http://www.drwren.com/chris/9.913/clean_ssd.avi

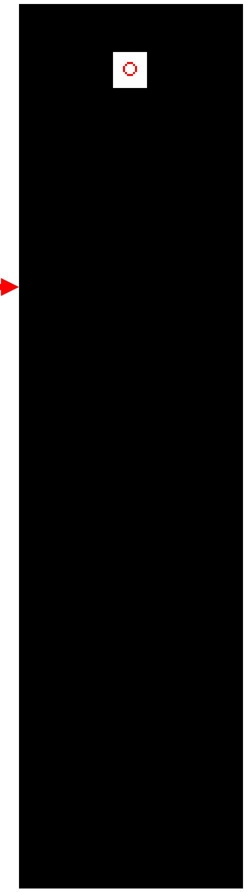
Efficient Region Tracking With Parametric Models of Geometry and Illumination . Belhumeur & Hager, *IEEE PAMI*, 20(10), pp.1125-1139, 1998

The Simplest Tracker



http://www.drwren.com/chris/9.913/clean_ssd.avi

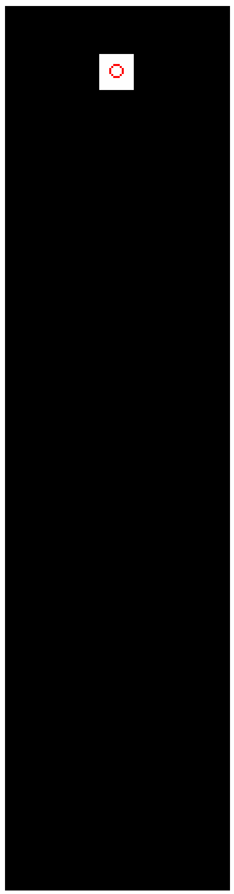
$$\arg\min_{x,y} \mathrm{SSD}(I, T)$$



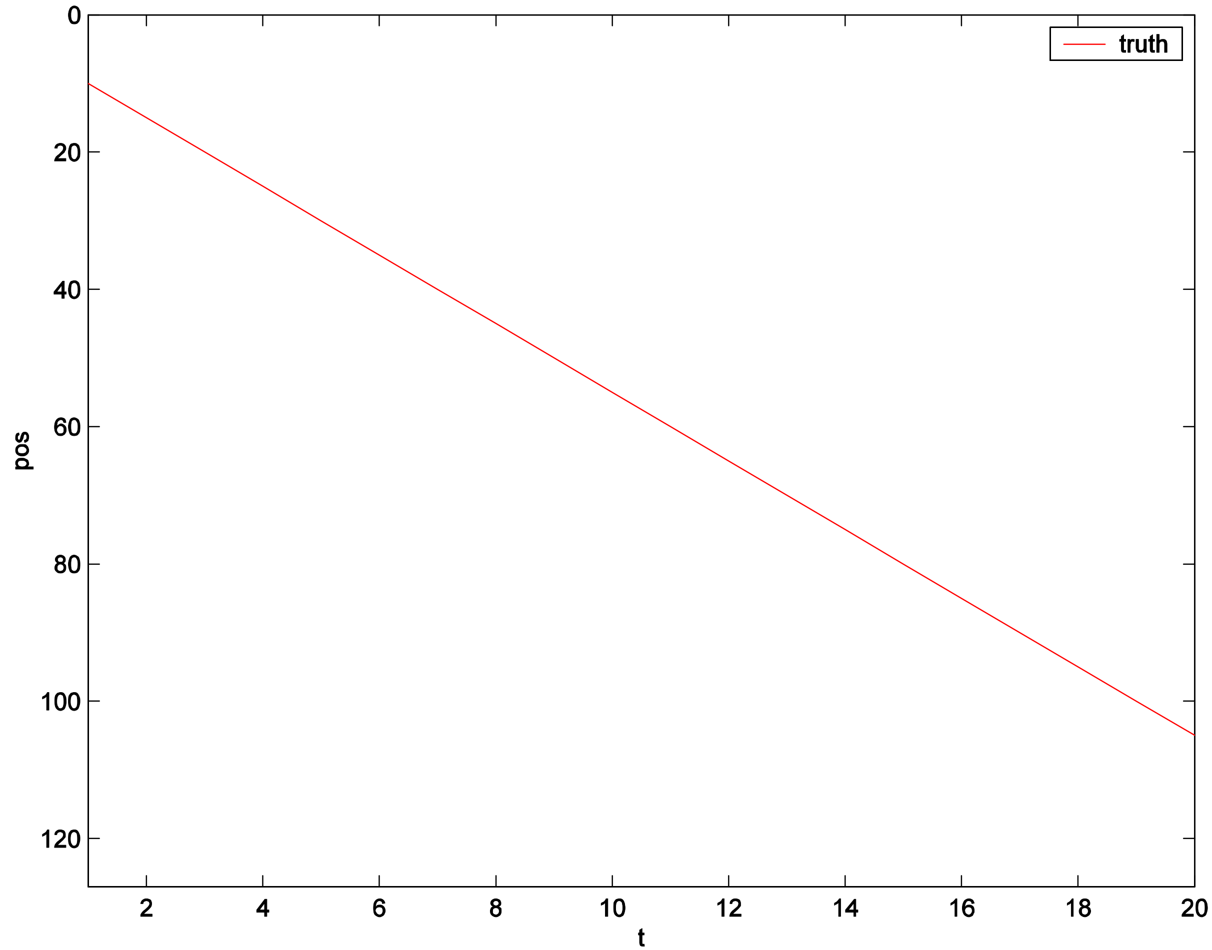
http://www.drwren.com/chris/9.913/clean_track.avi

Tracking as Serial Object Detection

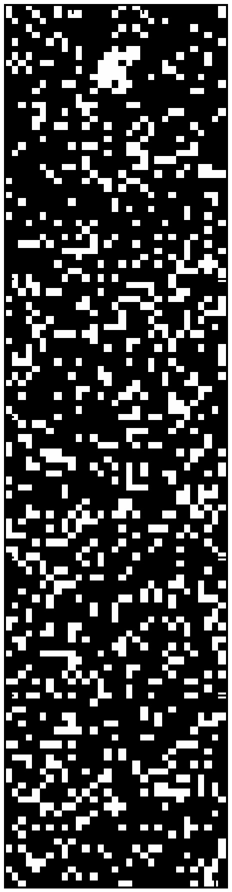
The Simplest Tracking Results



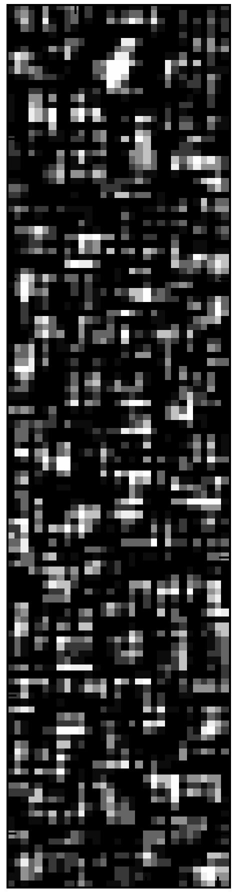
http://www.drwren.com/chris/9.913/clean_track.avi



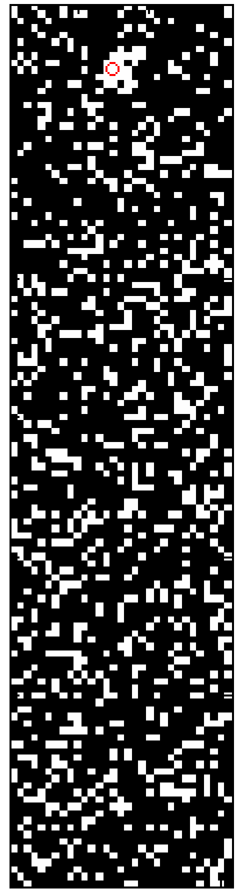
SSD Tracking in Noise



http://www.drwren.com/chris/9.913/noisy.avi



http://www.drwren.com/chris/9.913/noisy_ssd.avi

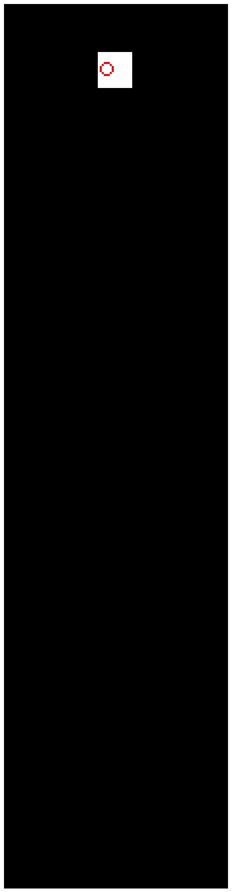


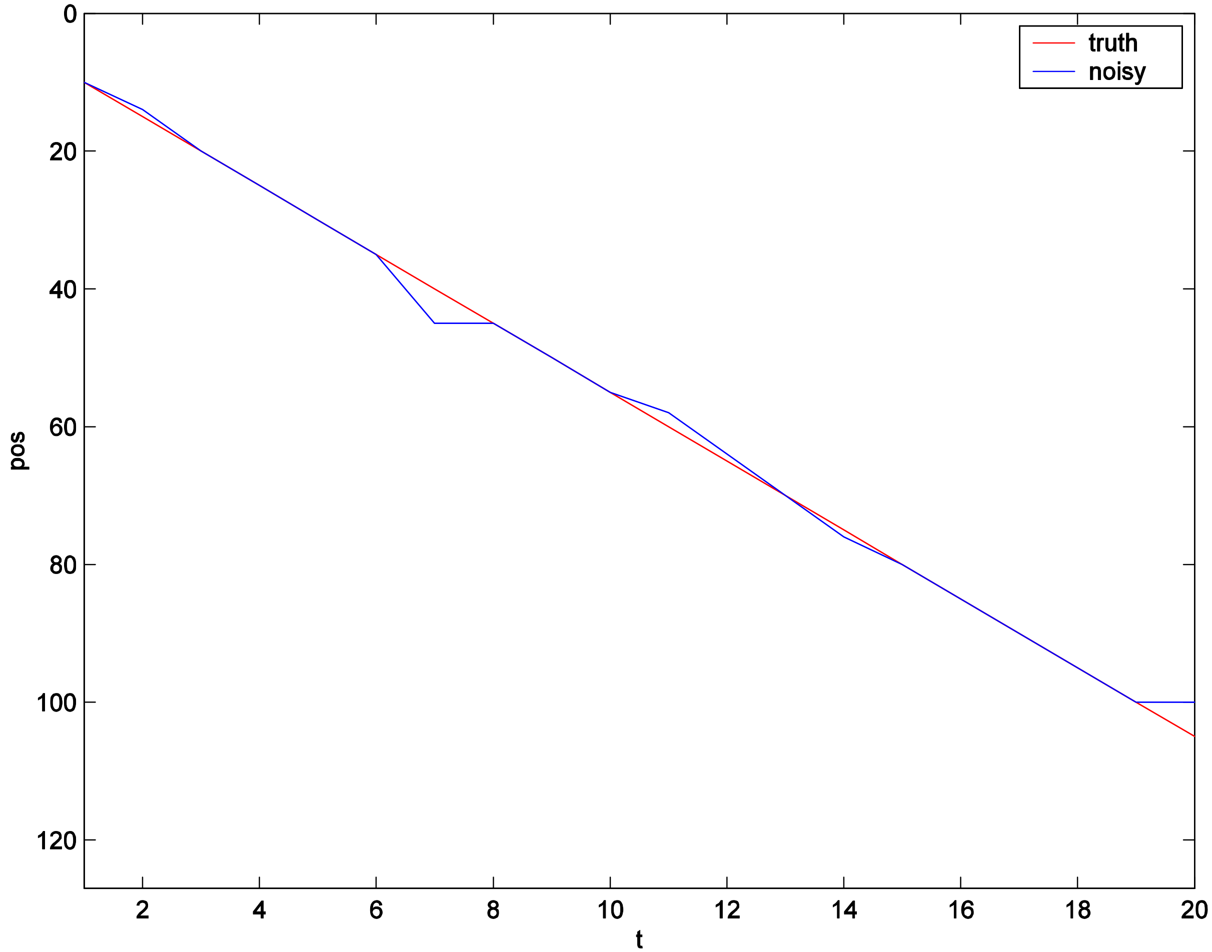
http://www.drwren.com/chris/9.913/noisy_track.avi

http://www.drwren.com/chris/9.913/noisy_cmp.avi

$$J(i,j) = [I(i,j) + N(0,0.5)] > 0.5$$

Noisy Results





http://www.drwren.com/chris/9.913/noisy_cmp.avi

Dynamics: State Vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

State

Everything you need to know about the system to summarize its history:

Dynamics are Markov

- Position
- Orientation
- Velocity

Linear Dynamic Systems

$$\begin{aligned} \mathbf{x}_{t+1} &= \Phi_t \mathbf{x}_t + \mathbf{B}_t \mathbf{u}_t \\ \mathbf{y}_t &= \mathbf{H}_t \mathbf{x}_t \end{aligned}$$

x – state vector

y – observation

u – control signal

Φ – Dynamic Update Matrix

H – Observation Matrix

B – Control Matrix

MIT Course 6.241. **Applied Optimal Estimation.** Ed. Arthur Gelb. MIT Press, 1974.

LDS: Matrix Forms

$$\mathbf{x}_{t+1} = \mathbf{\Phi}_t \mathbf{x}_t + \mathbf{B}_t \mathbf{u}_t$$
$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t$$

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\mathbf{\Phi}_t = \begin{bmatrix} 10 \\ 01 \end{bmatrix} \qquad \mathbf{H}_t = \begin{bmatrix} 10 \\ 01 \end{bmatrix}$$

LDS: Matrix Forms

$$\begin{aligned}\mathbf{x}_{t+1} &= \boldsymbol{\Phi}_t \mathbf{x}_t + \mathbf{B}_t \mathbf{u}_t \\ \mathbf{y}_t &= \mathbf{H}_t \mathbf{x}_t\end{aligned}$$

$$\mathbf{x} = \begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix}$$

$$\boldsymbol{\Phi}_t = \begin{bmatrix} 1100 \\ 0100 \\ 0011 \\ 0001 \end{bmatrix}$$

$$\mathbf{H}_t = \begin{bmatrix} 1000 \\ 0010 \end{bmatrix}$$

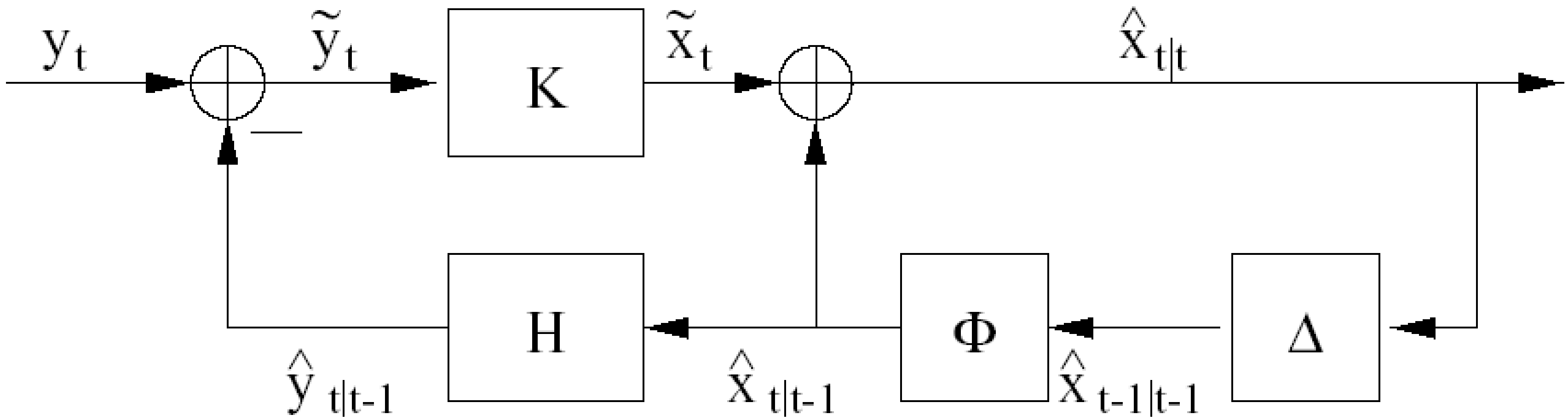
LDS in Noise

$$\mathbf{x}_{t+1} = \mathbf{\Phi}_t \mathbf{x}_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{L}_t \boldsymbol{\xi}_t$$
$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \boldsymbol{\theta}_t$$

$$E[\boldsymbol{\xi}_t] = 0$$
$$E[\boldsymbol{\theta}_t] = 0$$
$$E[\boldsymbol{\xi}_t \boldsymbol{\xi}_\tau^T] = \mathbf{\Xi}_t \delta_{t,\tau}$$
$$E[\boldsymbol{\theta}_t \boldsymbol{\theta}_\tau^T] = \mathbf{\Theta}_t \delta_{t,\tau}$$
$$E[\boldsymbol{\xi}_t \boldsymbol{\theta}_\tau^T] = 0$$

MIT Course 6.433. **Applied Optimal Estimation.** Ed. Arthur Gelb. MIT Press, 1974.

The Kalman Filter



- The Kalman Filter is the Minimum Mean Squared Error Estimator of the state of a LDS in zero-mean, white, Gaussian noise.
- The current state of the system is estimated in a MMSE sense with respect to all the observations up to the current time.
- The difference between the predicted observation and the actual observation is weighted by the Kalman Gain Matrix to become the innovation.

MIT Course 6.433. **Applied Optimal Estimation.** Ed. Arthur Gelb. MIT Press, 1974.

Kalman Filter: Predict

$$\mathbf{\Sigma}_{0|0} \stackrel{\triangle}{=} \mathbf{\Lambda}_{\tilde{\mathbf{x}}_{0|0}}$$

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{\Phi}_t \hat{\mathbf{x}}_t + \mathbf{B}_t \mathbf{u}_t$$

$$\hat{\mathbf{y}}_{t+1|t} = \mathbf{H}_{t+1} \hat{\mathbf{x}}_{t+1|t}$$

$$\mathbf{\Sigma}_{t+1|t} = \mathbf{\Phi}_t \mathbf{\Sigma}_{t|t} \mathbf{\Phi}_t^T + \mathbf{L}_t \mathbf{\Xi}_t \mathbf{L}_t^T$$

MIT Course 6.433. **Applied Optimal Estimation.** Ed. Arthur Gelb. MIT Press, 1974.

Kalman Filter: Update

$$\hat{\mathbf{x}}_{t+1|t+1} = \mathbf{x}_{t+1|t} + \mathbf{K}_{t+1}[\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t}]$$

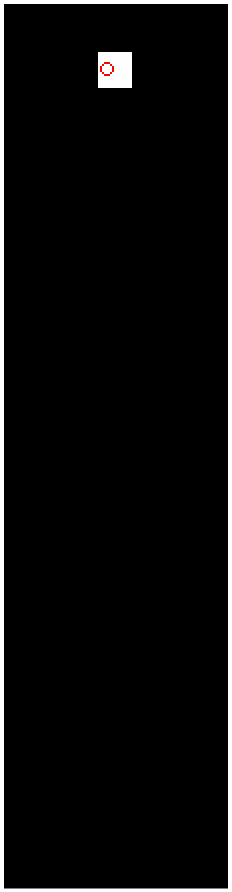
$$\boldsymbol{\Sigma}_{t+1|t+1} = [\mathbf{I} - \mathbf{K}_{t+1}\mathbf{H}_{t+1}]\,\boldsymbol{\Sigma}_{t+1|t}$$

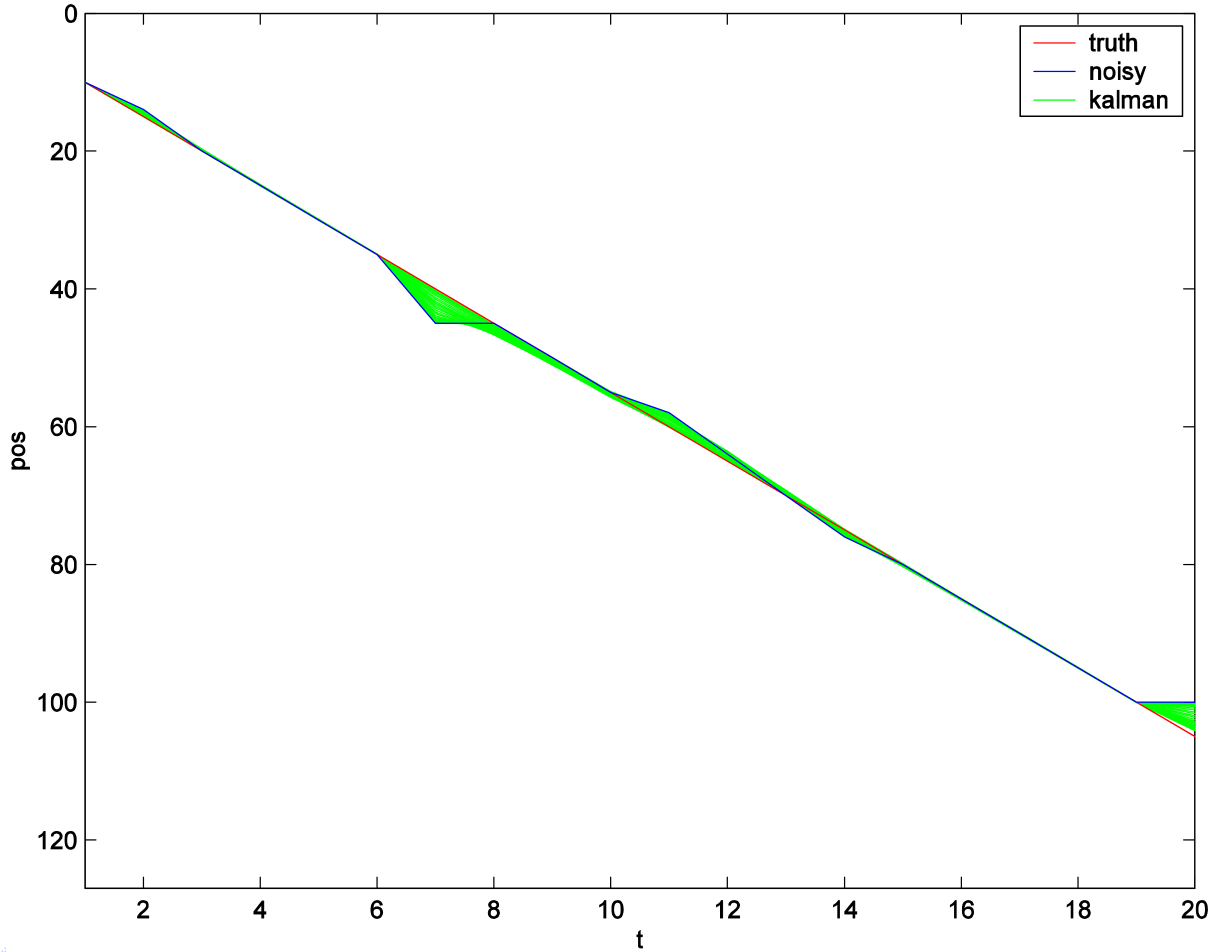
Kalman Gain Matrix

$$\mathbf{K}_{t+1} = \boldsymbol{\Sigma}_{t+1|t}\mathbf{H}_t^T\left[\mathbf{H}_t\boldsymbol{\Sigma}_{t+1|t}\mathbf{H}_t^T + \boldsymbol{\Theta}_{t+1}\right]^{-1}$$

MIT Course 6.433. **Applied Optimal Estimation.** Ed. Arthur Gelb. MIT Press, 1974.

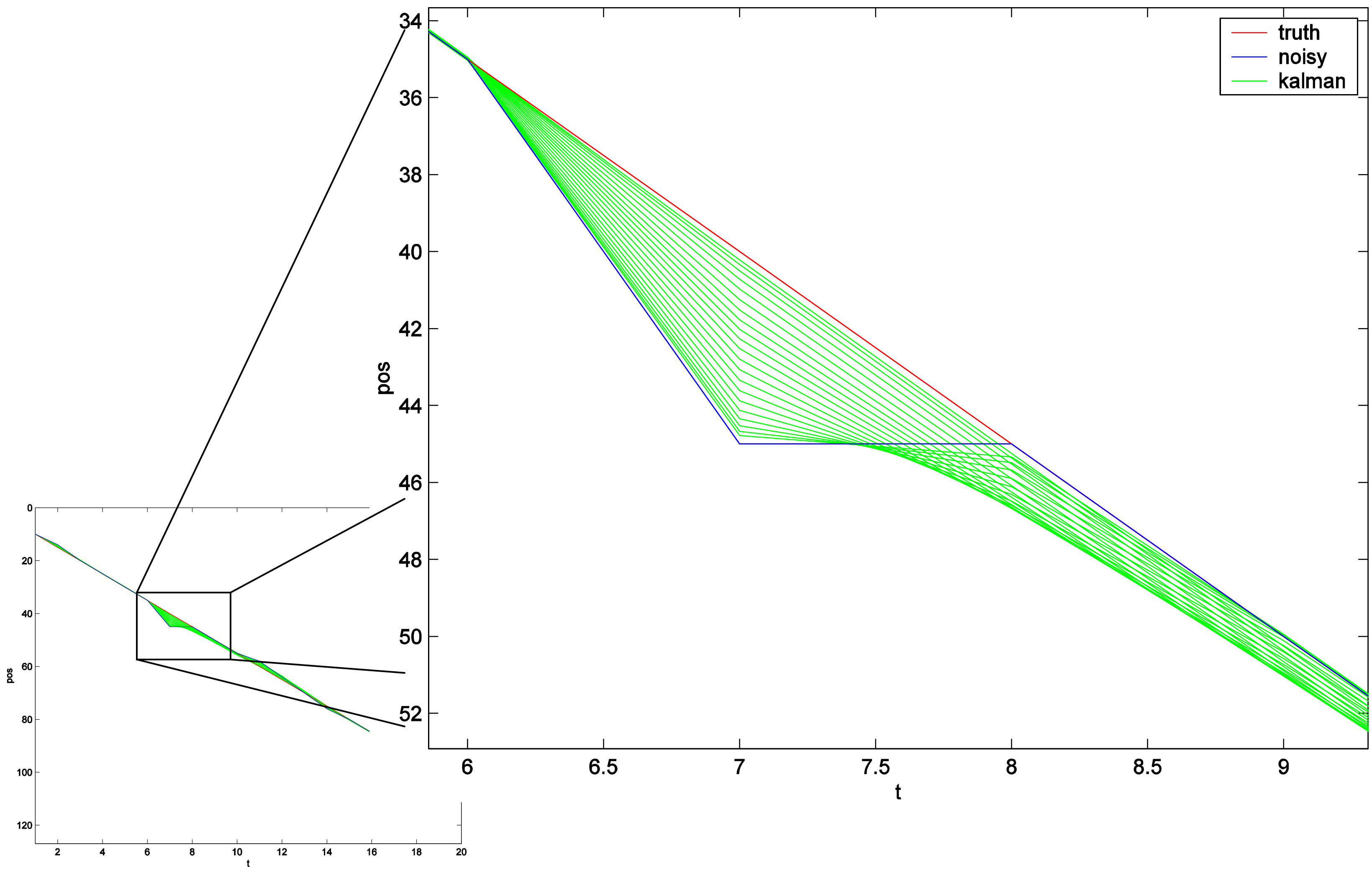
Kalman Smoothing Noisy Data





http://www.drwren.com/chris/9.913/noisy_cmp.avi

Kalman Smoothing Noisy Data: Detail



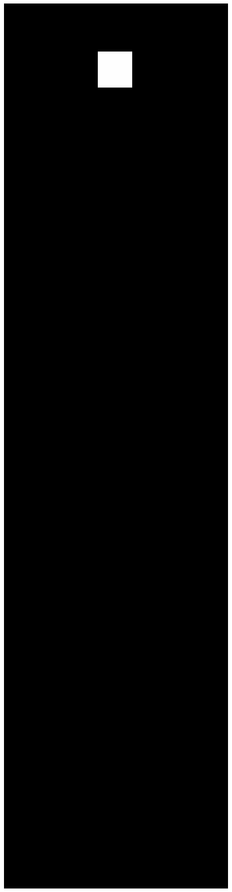
Kalman Insights

Error Covariance is mostly influenced by system noise and prior error.

Kalman Gain determines the mixing between prediction and observation based on the Error Covariance and the Observation Noise Covariance.

L, $\Sigma 0$, Θ, and Ξ together control how much the Kalman Filter will trust the observations versus the dynamic model.

What happens when the Model is Wrong?

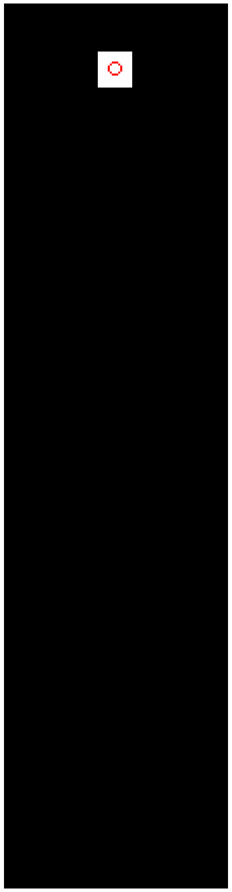


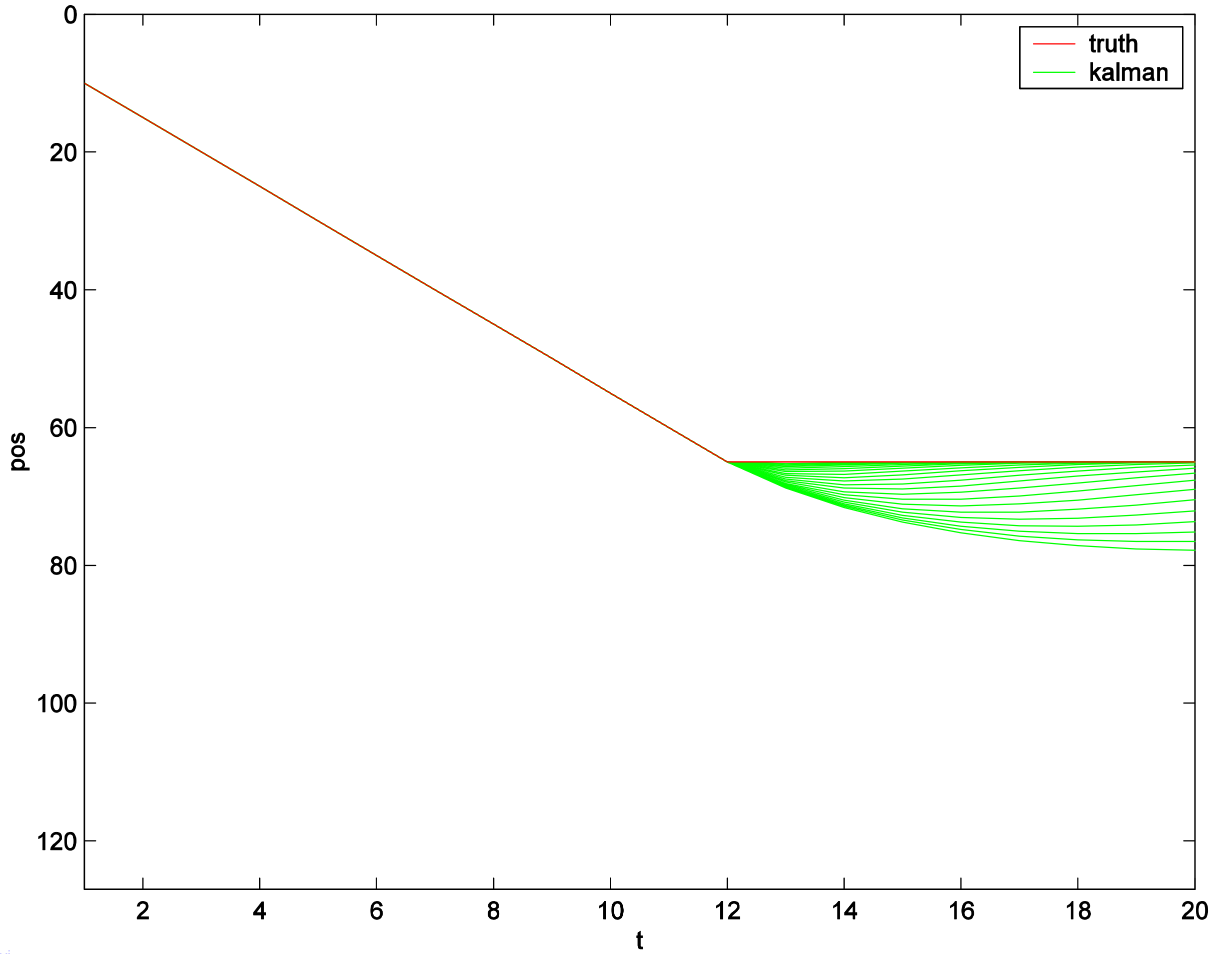
http://www.drwren.com/chris/9.913/stop.avi

→ Kalman Filter → ?

Given: Random Walk in Velocity

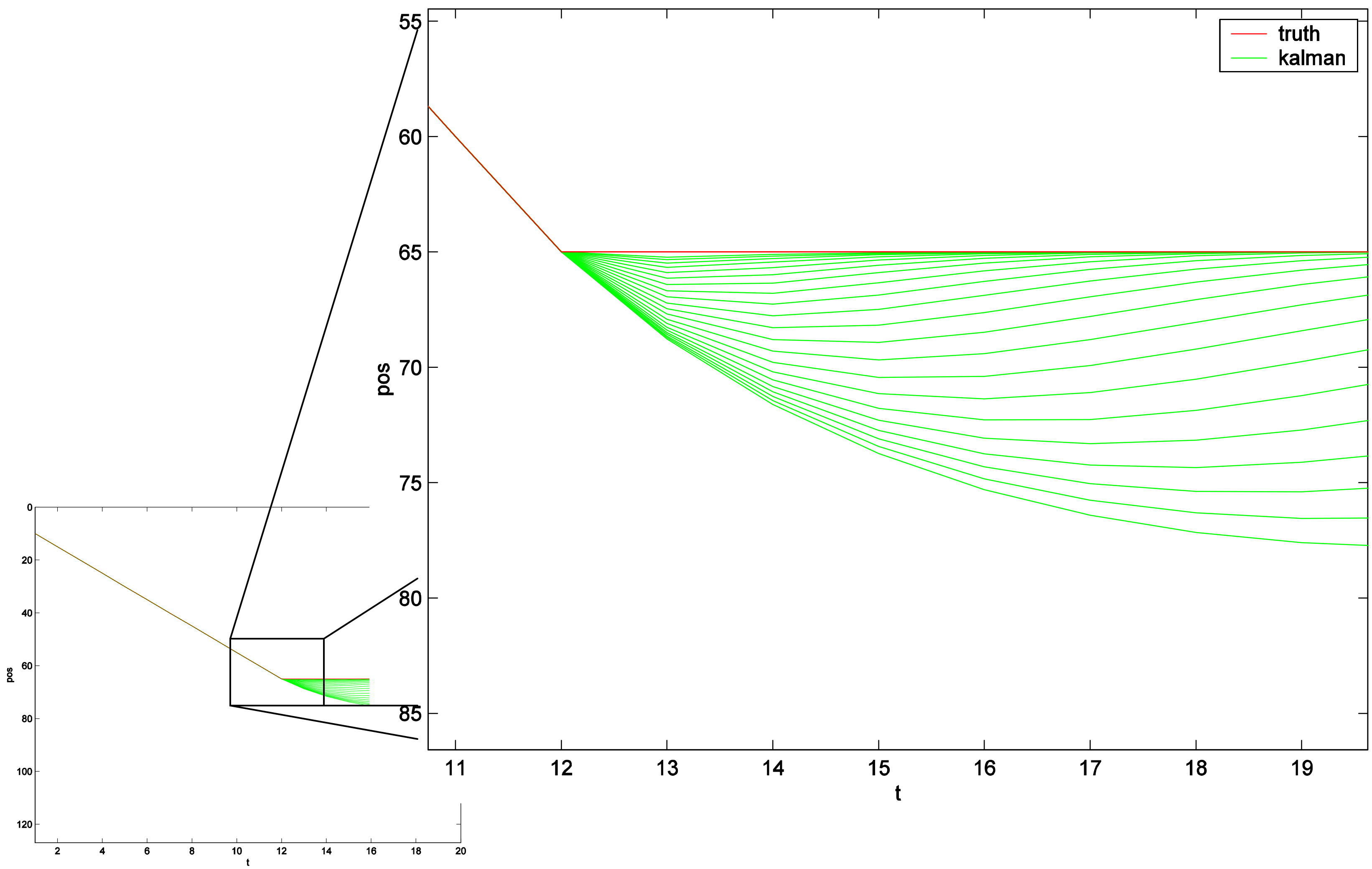
Overshoot, Lag, etc.





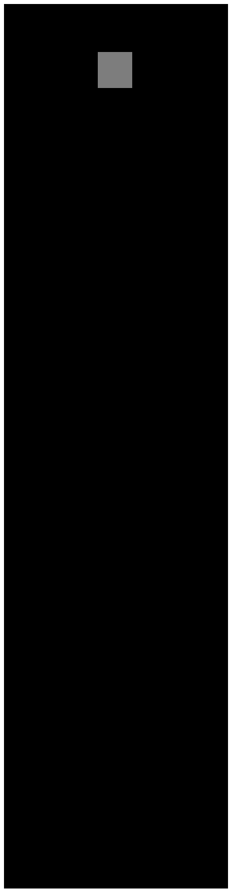
http://www.drwren.com/chris/9.913/stop_track.avi

Overshoot: Detail

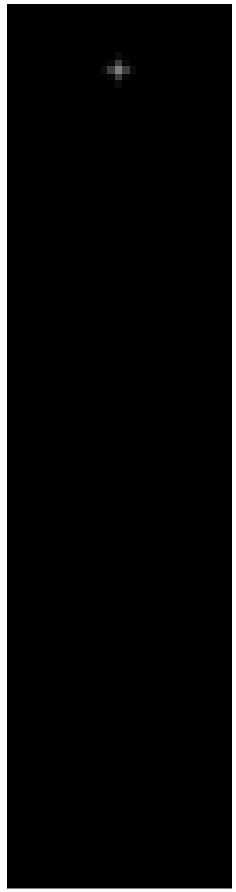


Occlusion

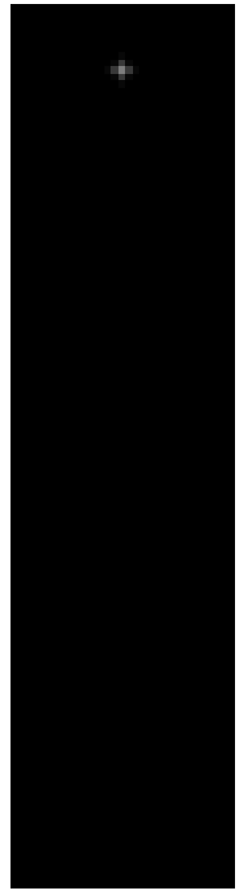
Windowed tracker in the presence of occlusion.



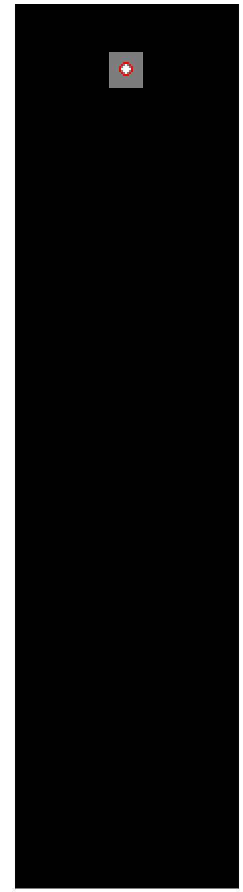
http://www.drwren.com/chris/9.913/occlude.avi



http://www.drwren.com/chris/9.913/occlude_ssd.avi



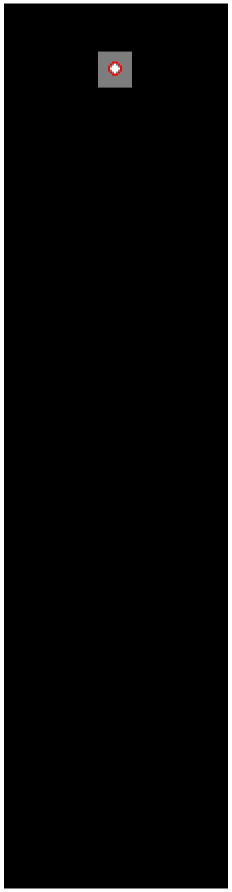
http://www.drwren.com/chris/9.913/occlude_map.avi

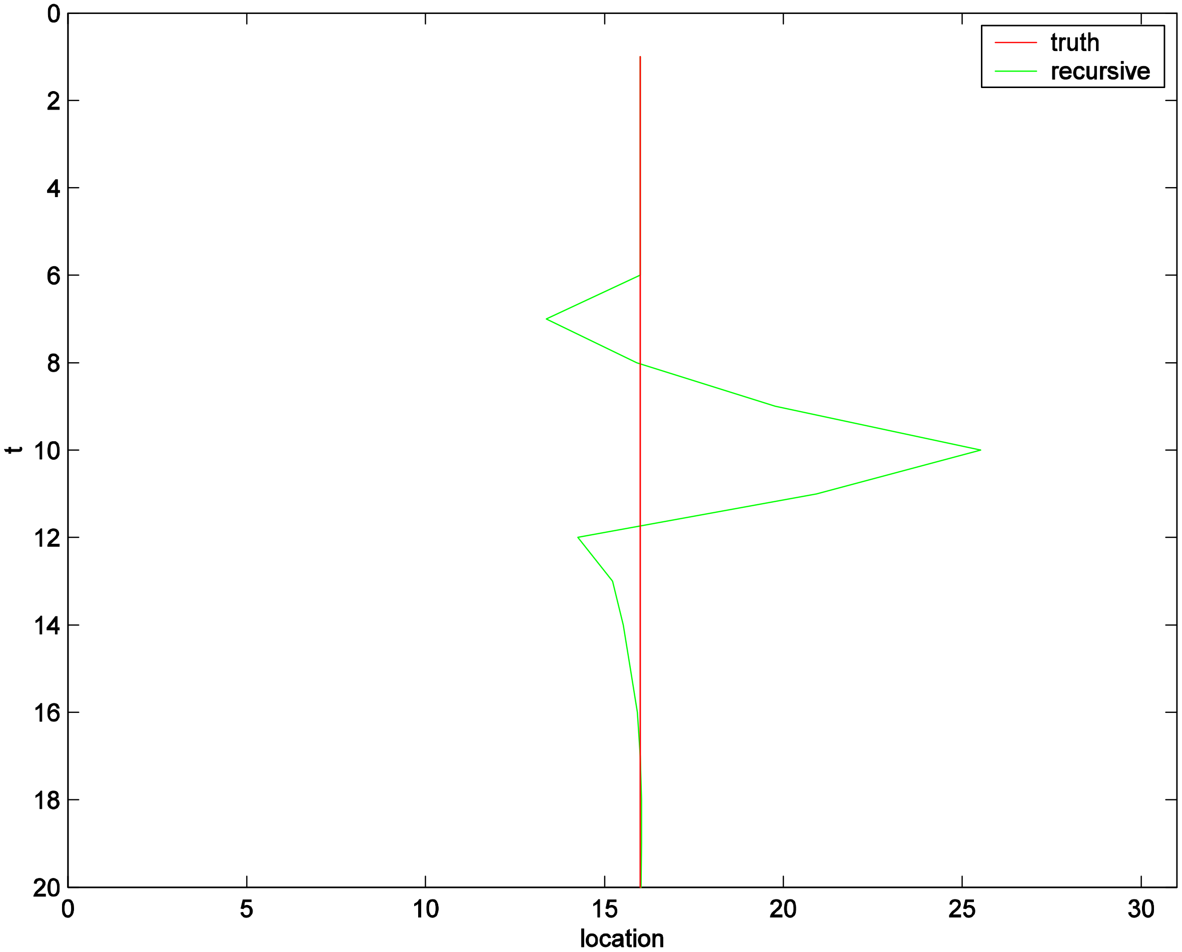


http://www.drwren.com/chris/9.913/occlude_cmp.avi

Occlusion Results

- Why doesn't the second occlusion distract the tracker?
- Why did it recover from the first distraction?





http://www.drwren.com/chris/9.913/occlude_cmp.avi

Multi-Mode Trackers

- Kalman Filter
  - Uni-modal, parametric density
- Multiple Hypothesis Testing
  - Multi-modal, parametric density
  - Many dynamic models, Φ
  - Many Innovations
  - Select the smallest Innovation as the best one and reset other filters
- CONDENSATION
  - Drop the Error Covariance – just state estimates – "particle filtering"
  - Multi-modal, non-parametric density
  - Many dynamic models, often simply unity with large process noise
  - Resample distribution weighted by observations
  - MANY MANY samples - does it scale high dimensional spaces?

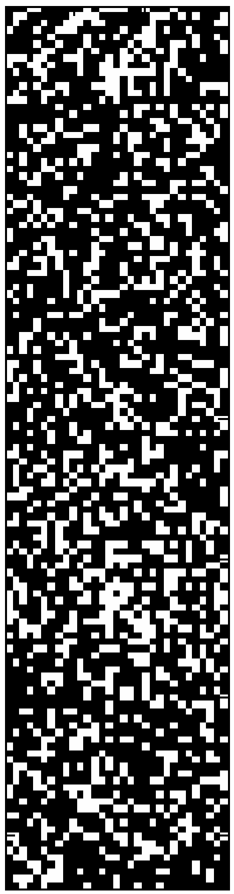
MIT Course 6.433

"A dynamic bayesian network approach to figure tracking using learned dynamic models " Pavlovic, Rehg, Cham, Murphy. ICCV, 1999.
"Contour tracking by stochastic propagation of conditional density". Isard, Blake. ECCV, 1996.
"Adaptive estimation and parameter identification using multiple model estimation algorithm." Athans, Chang. MIT Lincoln Lab TR1976-28. 1976.

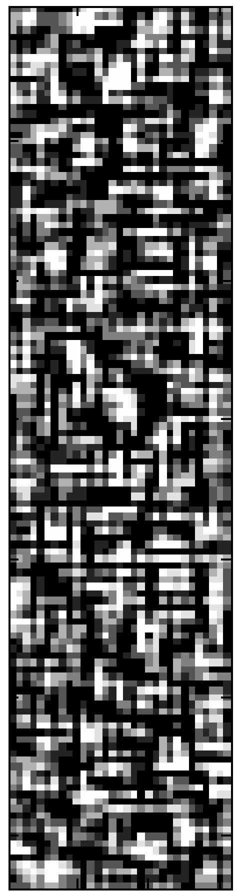
CONDENSATION: the Movie

http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/ISARD1/condensation.html

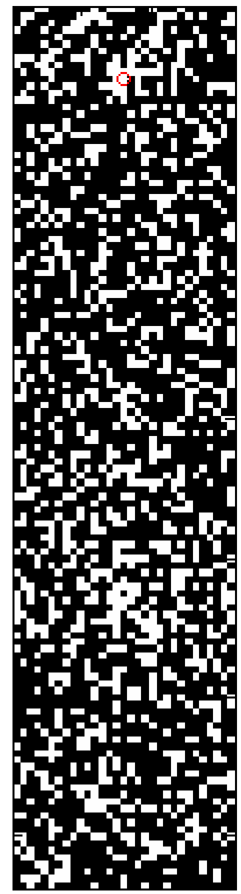
SSD Tracking in Extreme Noise



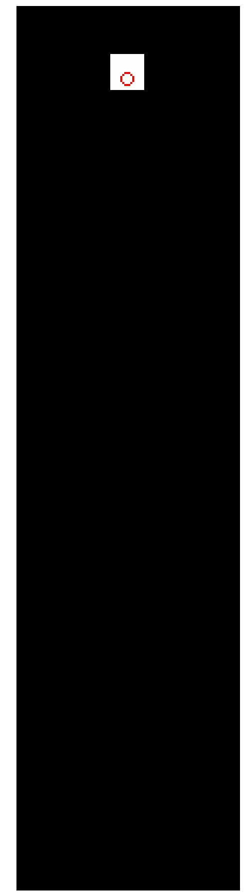
http://www.drwren.com/chris/9.913/fail.avi



http://www.drwren.com/chris/9.913/fail_ssd.avi



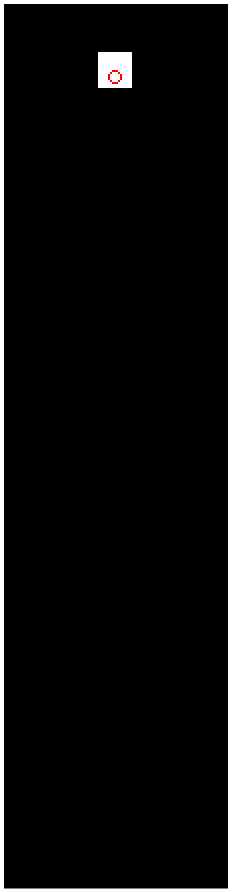
http://www.drwren.com/chris/9.913/fail_track.avi



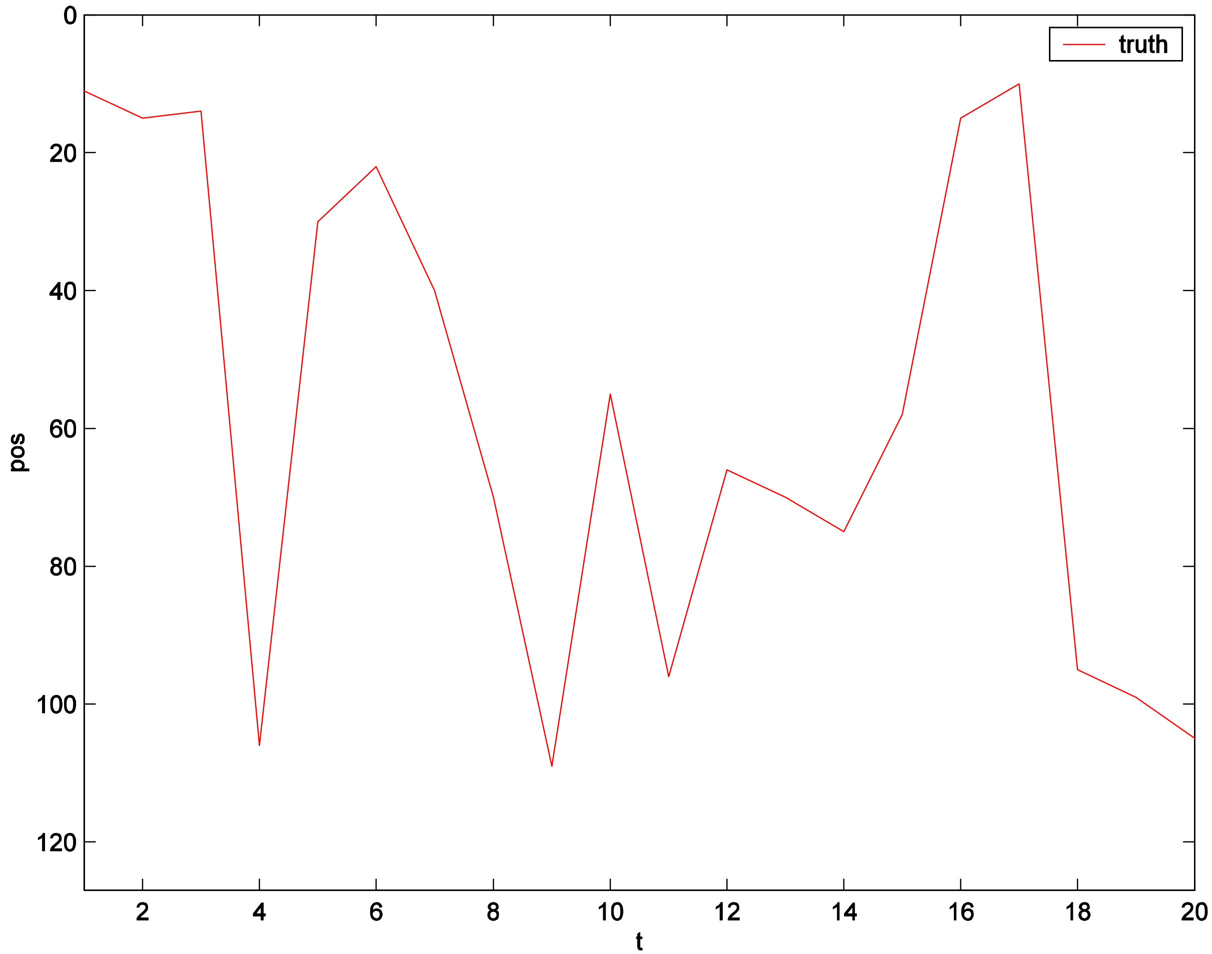
http://www.drwren.com/chris/9.913/fail_cmp.avi

$$J(i,j) = [I(i,j) + N(0,1)] > 0.5$$

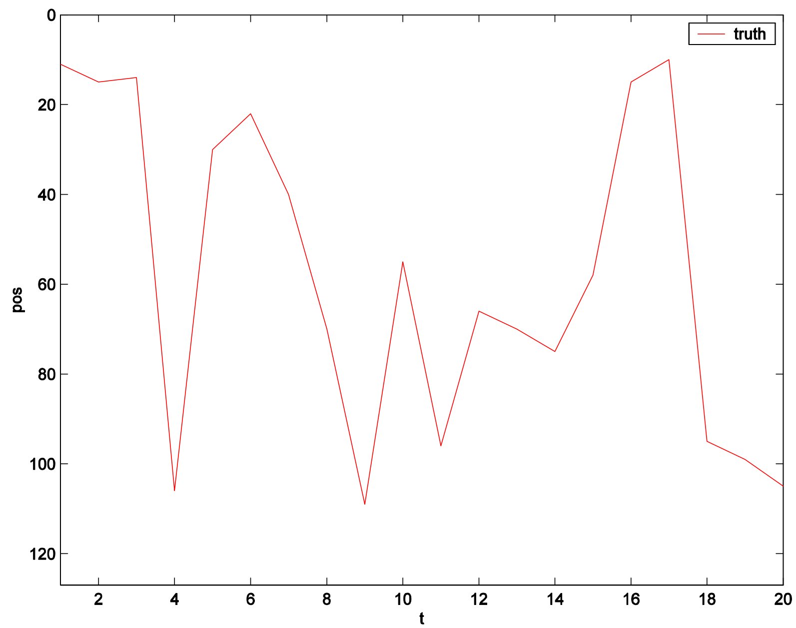
Extreme Noise Results



http://www.drwren.com/chris/9.913/fail_cmp.avi



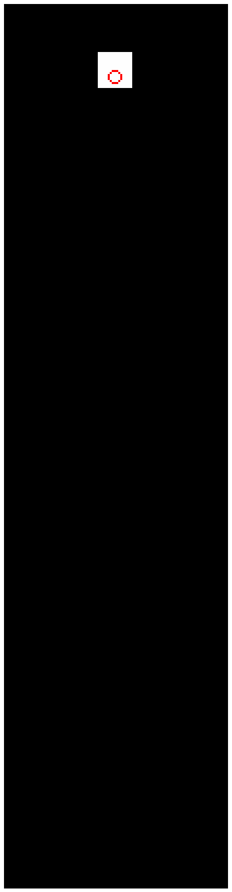
Kalman to the Rescue?



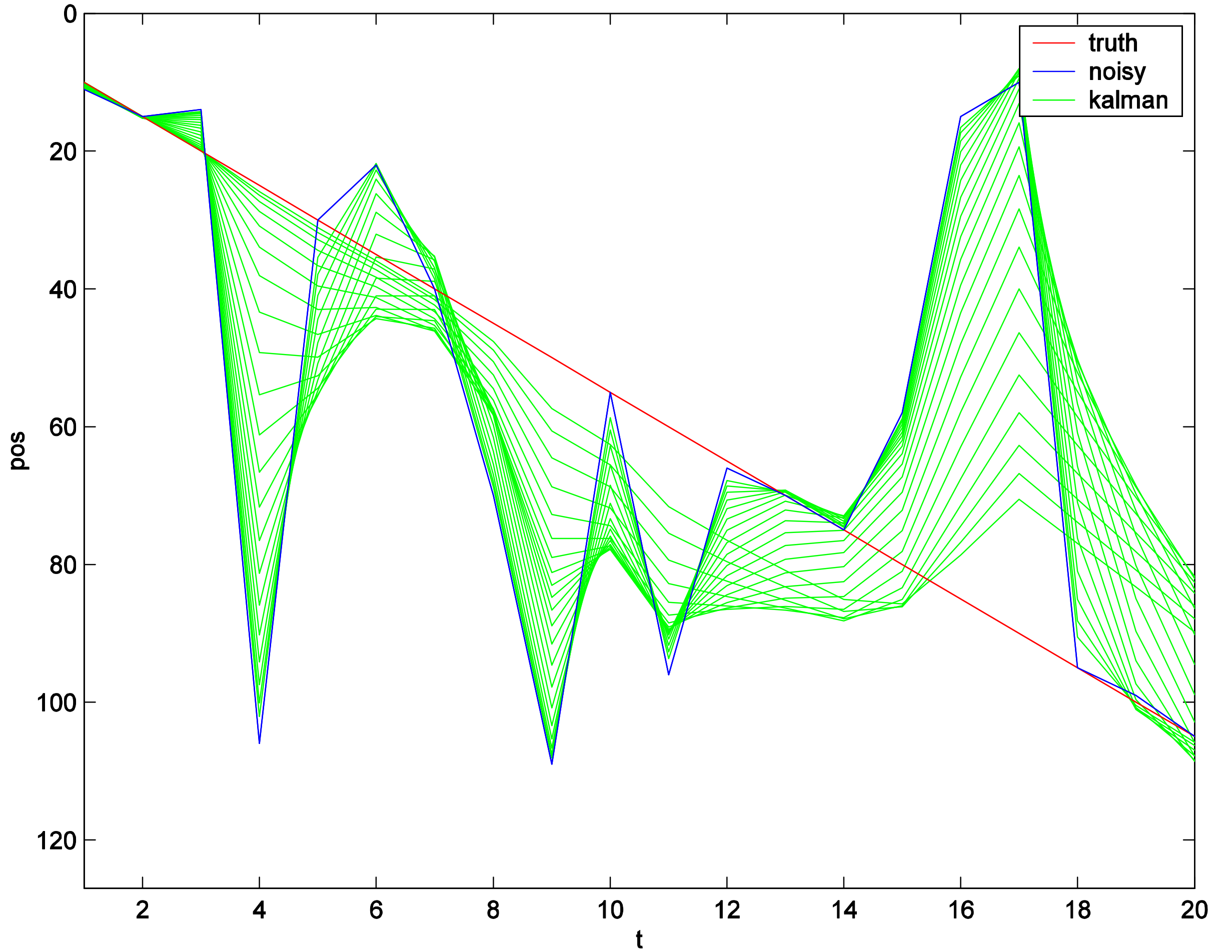
Can a Kalman Filter fix these tracking results?

Is this zero-mean, white, Gaussian noise?

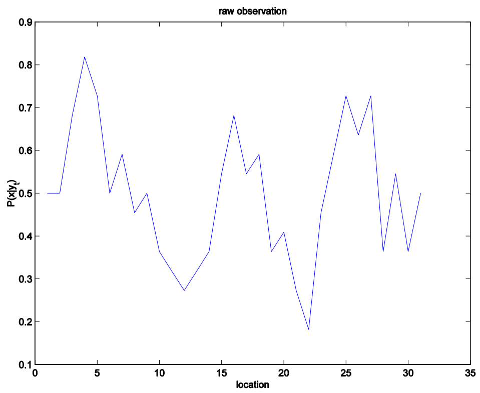
No, Not Really.



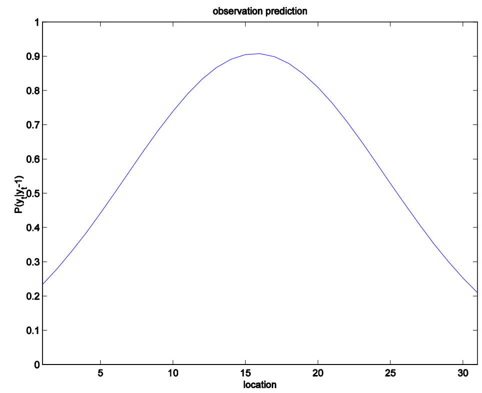
http://www.drwren.com/chris/9.913/fail_cmp.avi



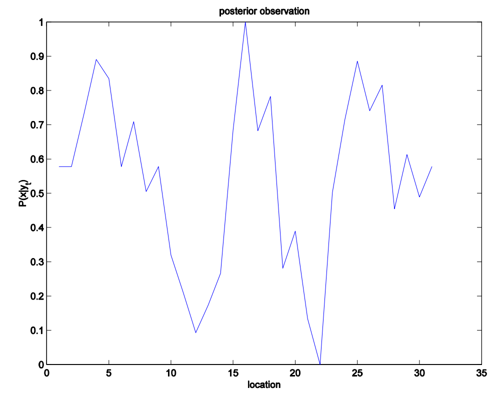
Affecting the Observation Process



X



=

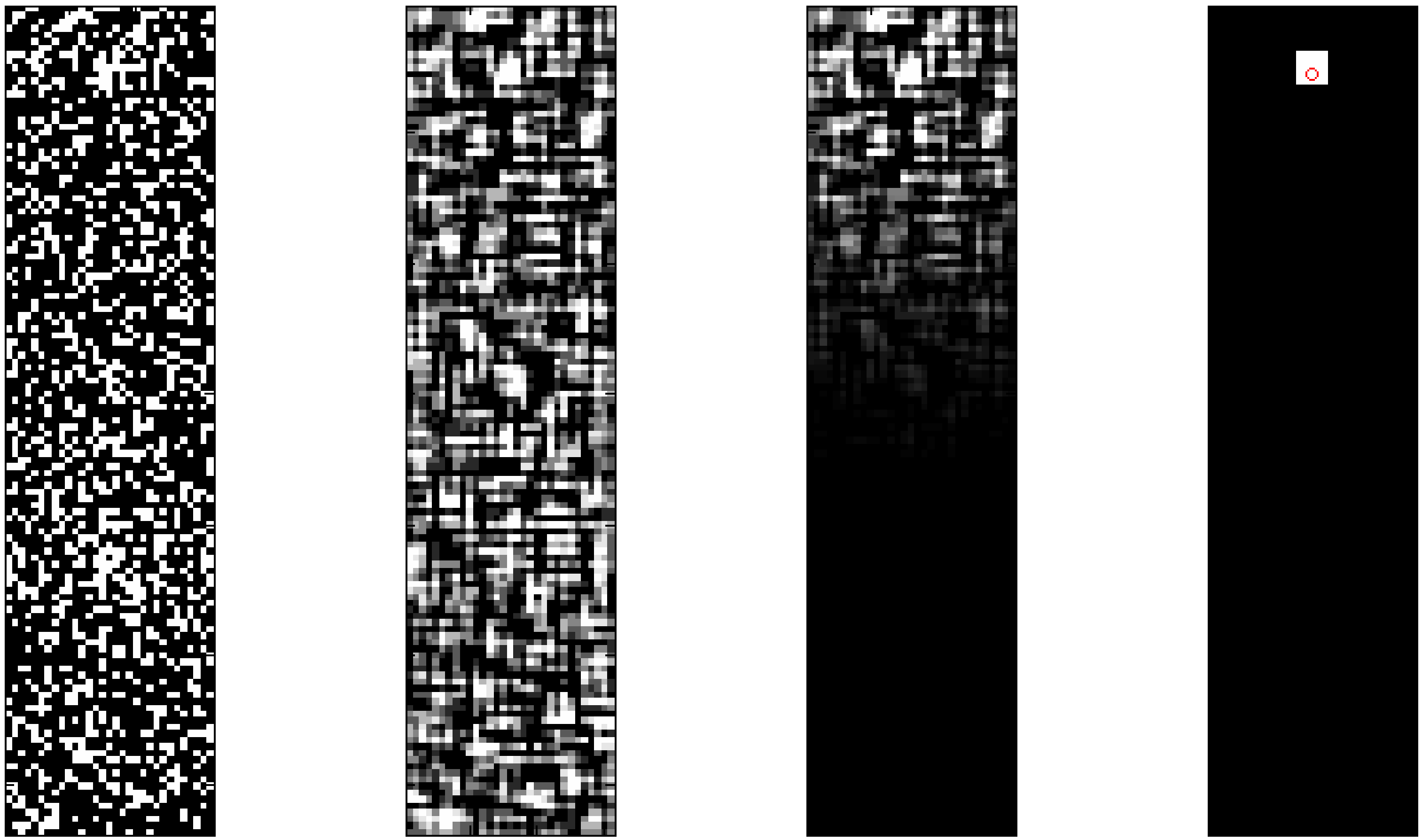


Raw Obs. Prob. Error Covariance Posterior Prob.

- Is it still a Kalman Filter?
- Is it still the MMSEE?
  - The observation noise is not independent of the state estimate.
  - What does it mean for an image to be an obs. In a Kalman Filter anyway?
- The model just became potentially very powerful

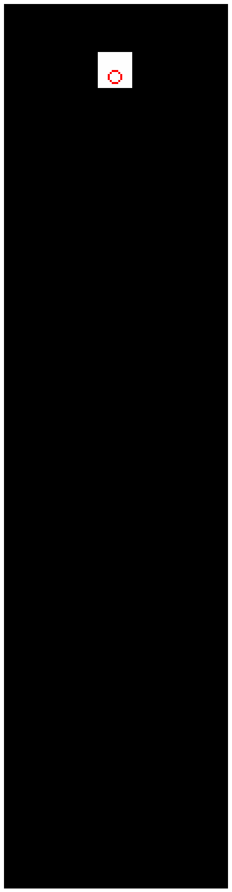
Predicted State and Error Covariance as Prior Information

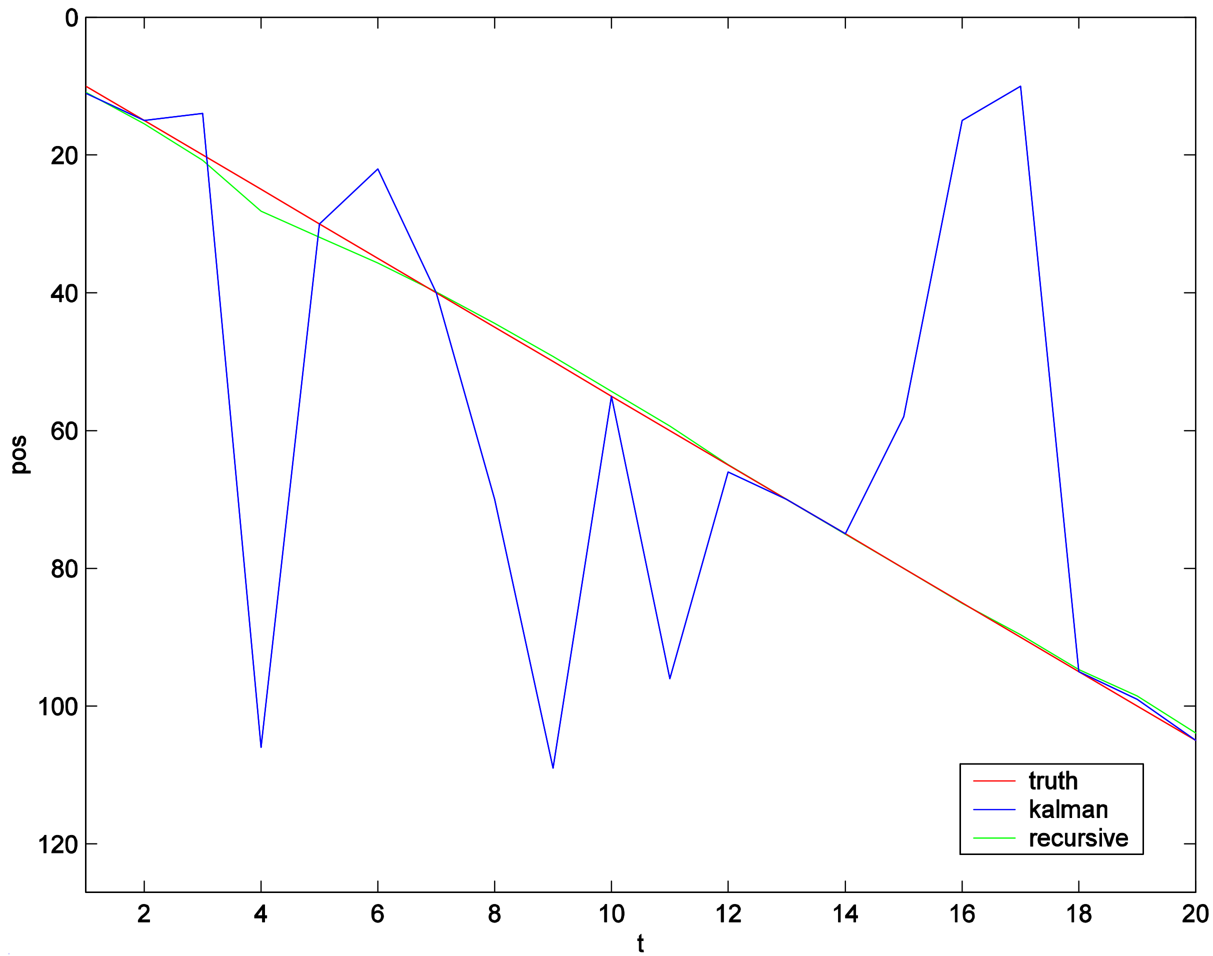
Weight SSD output by Predicted State/Error Gaussian before peak detection



http://www.drwren.com/chris/9.913/recover.avi http://www.drwren.com/chris/9.913/recover_ssd.avi http://www.drwren.com/chris/9.913/recover_win.avi http://www.drwren.com/chris/9.913/recover_cmp.avi

A Solution?





http://www.drwren.com/chris/9.913/recover_cmp.avi

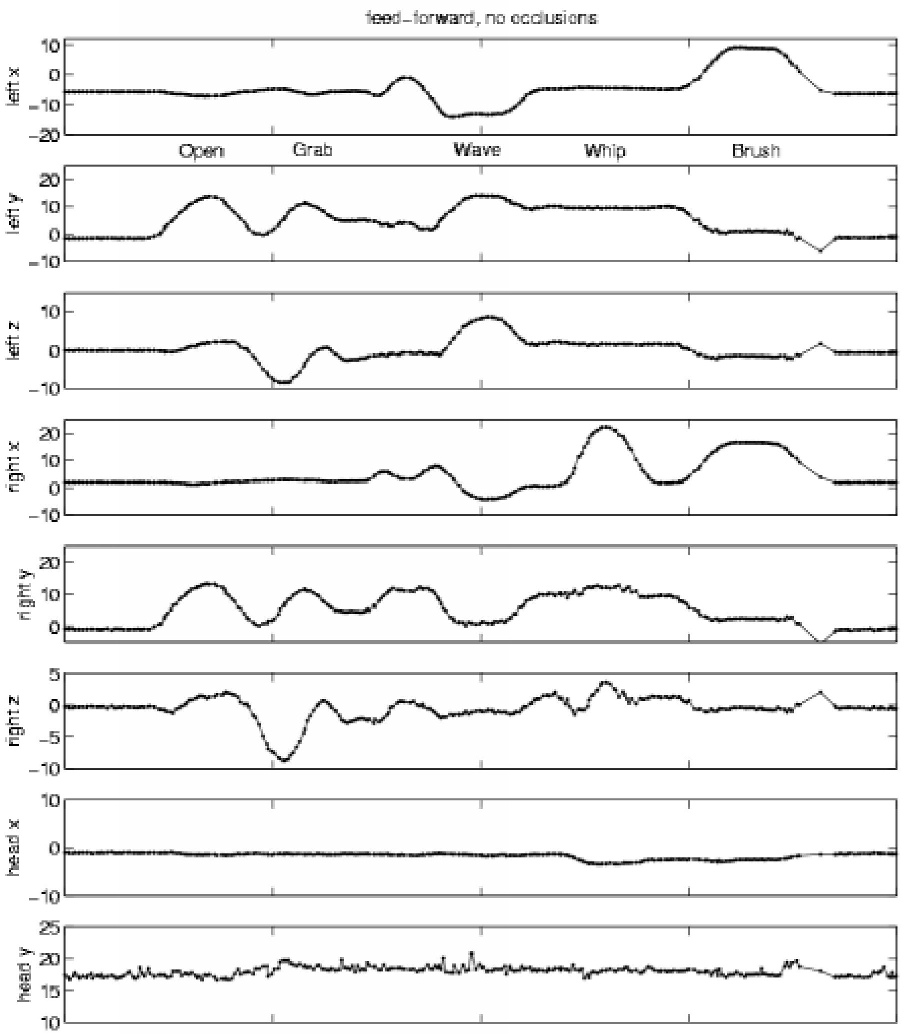
T'ai Chi Teacher

http://www.drwren.com/chris/9.913/taichi5mm.mpg

- Stereo 2D blob trackers
- 3D blob estimator
- HMM Gesture Recognition
  - Recognize gesture attempt
  - Threshold Probability to judge
  - Analyze lattice to find specific mistakes in the gesture
- State machine:
  - Teach
  - Watch
  - Critique
- Virtual Sensei character
  - Play ideal gestures
  - Mirror student
  - Replay student's last gesture
  - Indicate time and hand in error

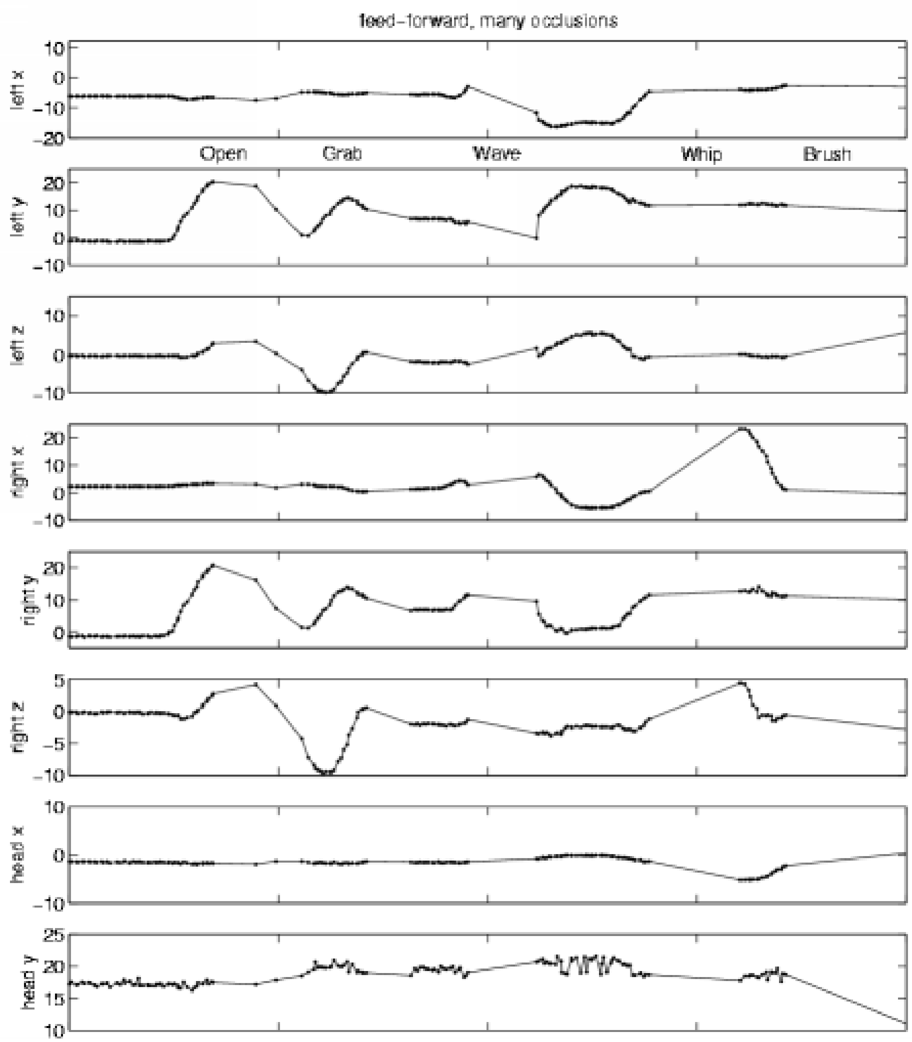
"Sensei: A Real-Time Recognition, Feedback, and Training System for T'ai Chi Gestures" **David** A. **Becker.** M.S. Thesis, MIT, 1997

Sensei: Normal Operation



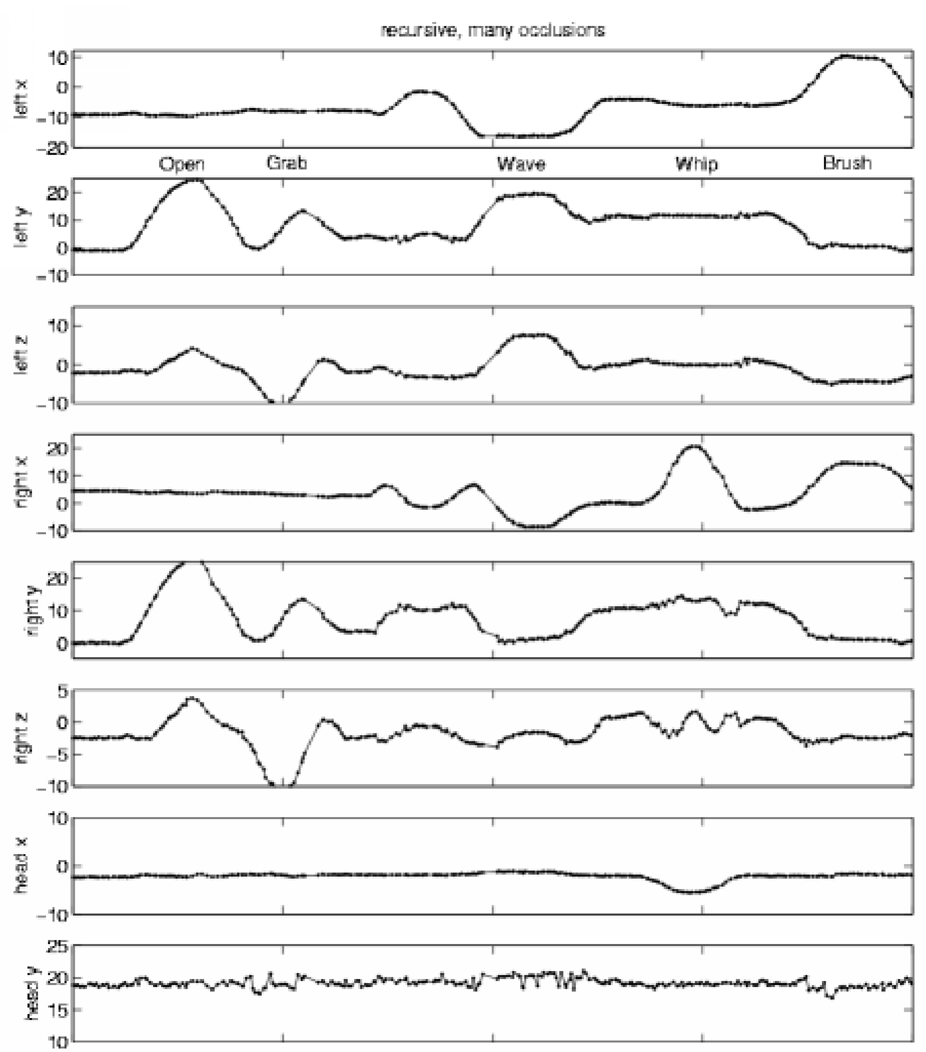
- Student who knows the gestures, and knows the limits of the system.
- Clean Tracks
- Evenly-spaced Samples

Sensei: Tracker Failure



- Pessimal Experience (ie. a fake session where everything that can go wrong did)
- Opening Form tracking fails because the hands go too high and occlude the face – a common novice mistake.
- The other gestures fail mostly because the hands occlude each other.
- Recovery often takes a long time.
- Uneven sampling wreaks havoc on the HMM.
- Those parts of the gesture causing the errors are invisible to the Sensei, so there's no way to give helpful feedback.

Sensei: Repaired with Recursive Tracker



- Occlusions can be predicted, and therefore handled easily.
- Tracking holes are filled in with state estimates to create even sampling.
- Clean data gives the HMM something to analyze: so feedback will be more appropriate.

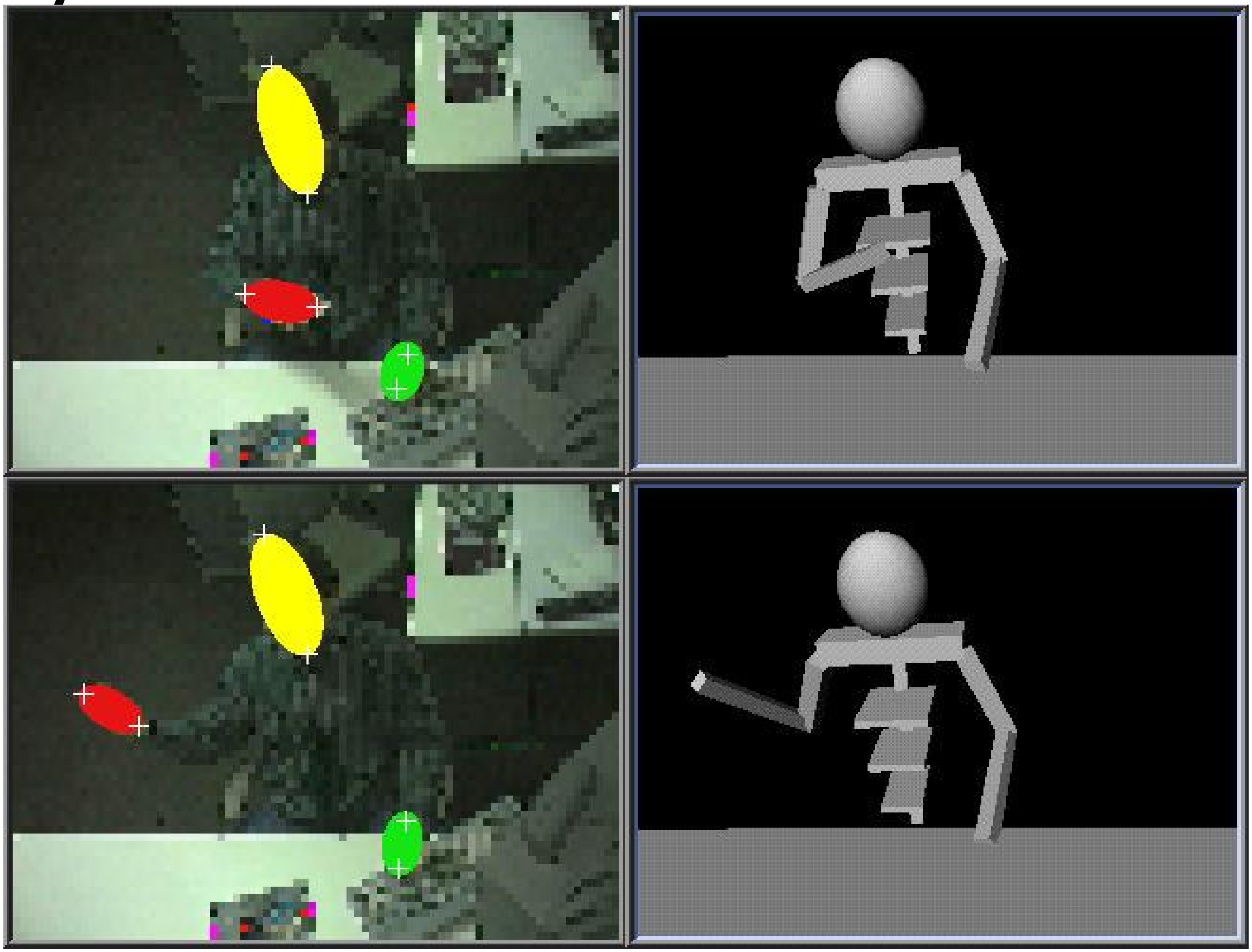
“Understanding Expressive Action”. Christopher R. Wren, Ph.D. Thesis, MIT EECS, 2000.

Non-linear Kalman Filtering

- Extended Kalman Filter
  - Linearize around current estimate
  - User standard Kalman Filter framework in tangent space
- Unscented Kalman Filter
  - Sample distribution
  - Push samples though non-linearity
- Dyna
  - Not exactly a Kalman filter
  - Predictions modify observation process
  - Full 6DOF physics model for body links
  - Modular constraint system

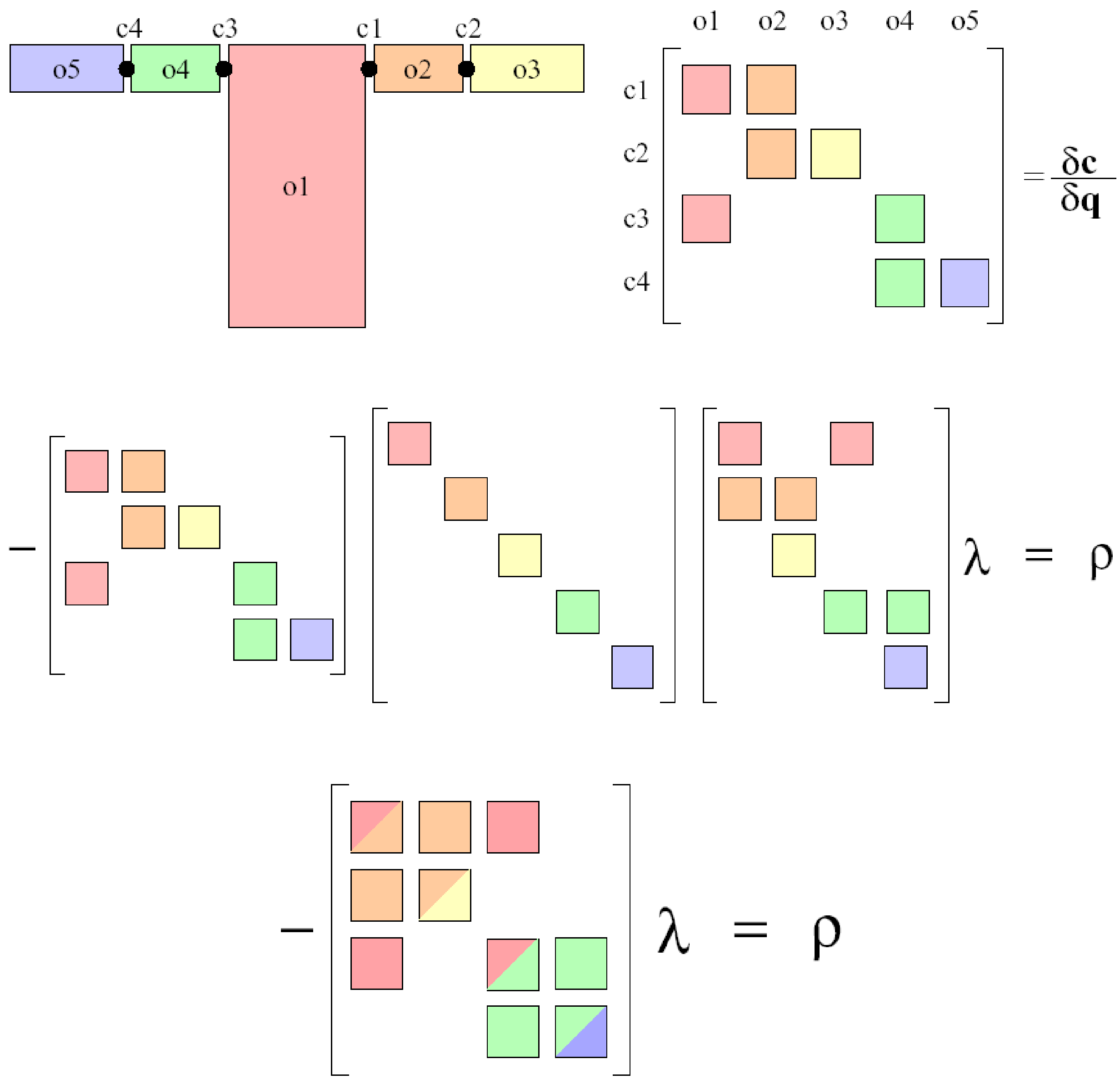
MIT Course 6.433. **Applied Optimal Estimation.** Ed. Arthur Gelb. MIT Press, 1974.
"The Unscented Kalman Filter for Nonlinear Estimation", Wan, van der Merwe IEEE *AS-SPCC*, 2000

Dyna: Observations and State



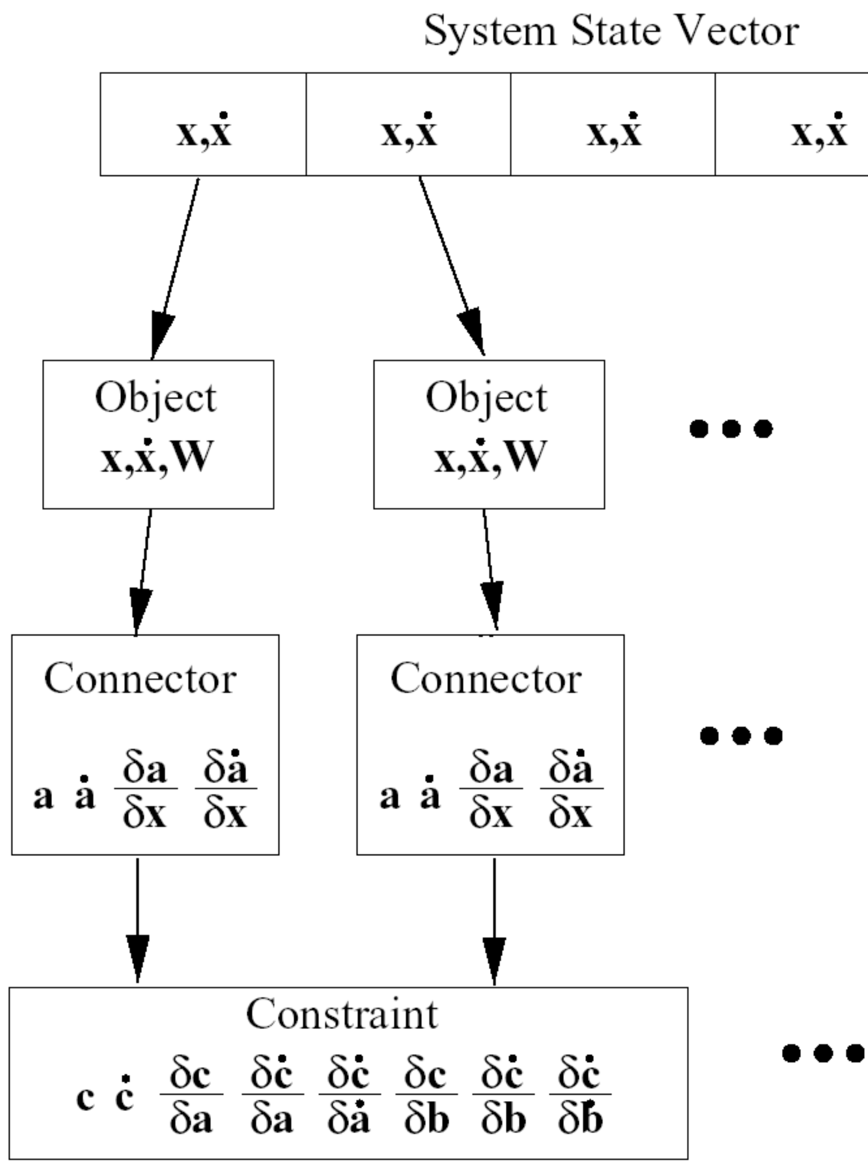
“Understanding Expressive Action”. Christopher R. Wren, Ph.D. Thesis, MIT EECS, 2000.

Dyna: Constraints



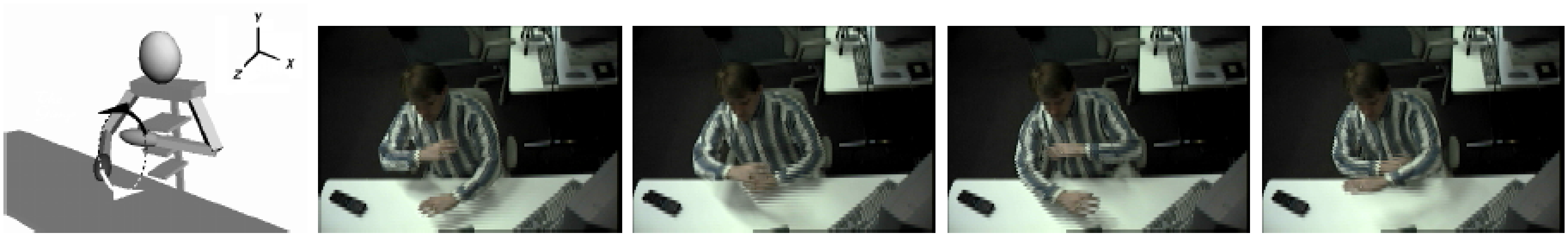
"Understanding Expressive Action". Christopher R. Wren, Ph.D. Thesis, MIT EECS, 2000.

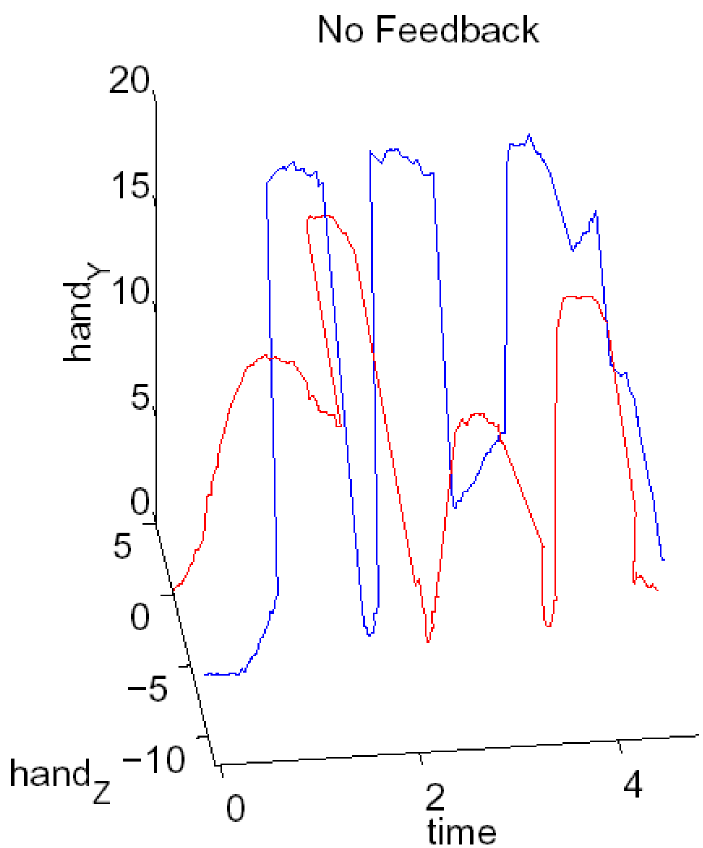
Dyna: Virtual Work Constraint System

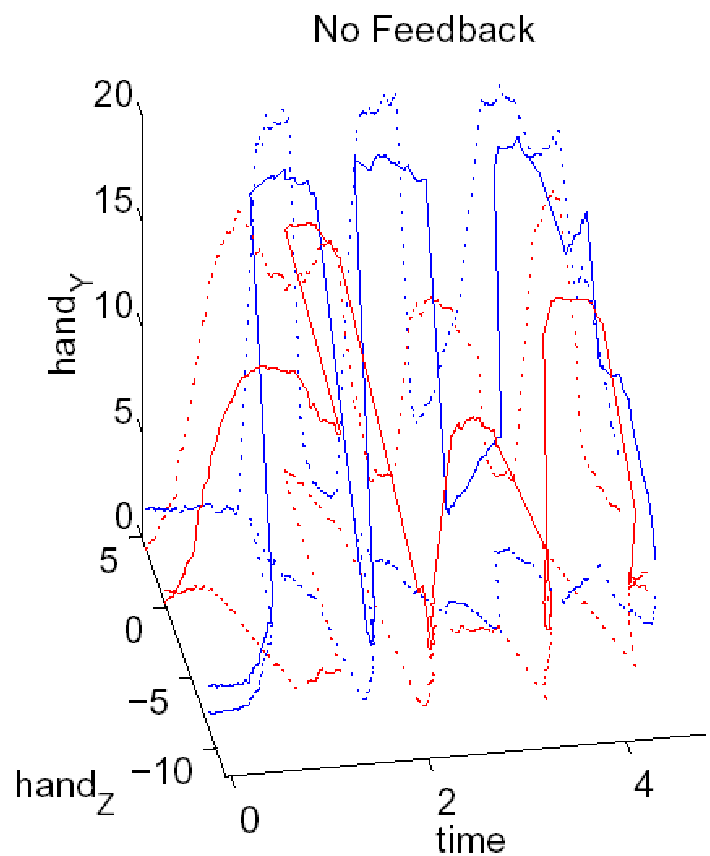


"Interactive Dynamics" Witkin, Gleicher, Welch.. *Computer Graphics*, 24(2), March1990.

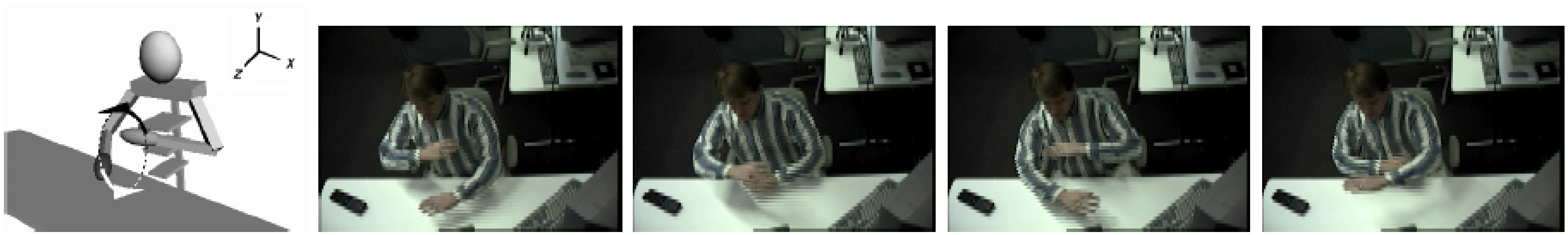
Repeated Occlusions: without recursion

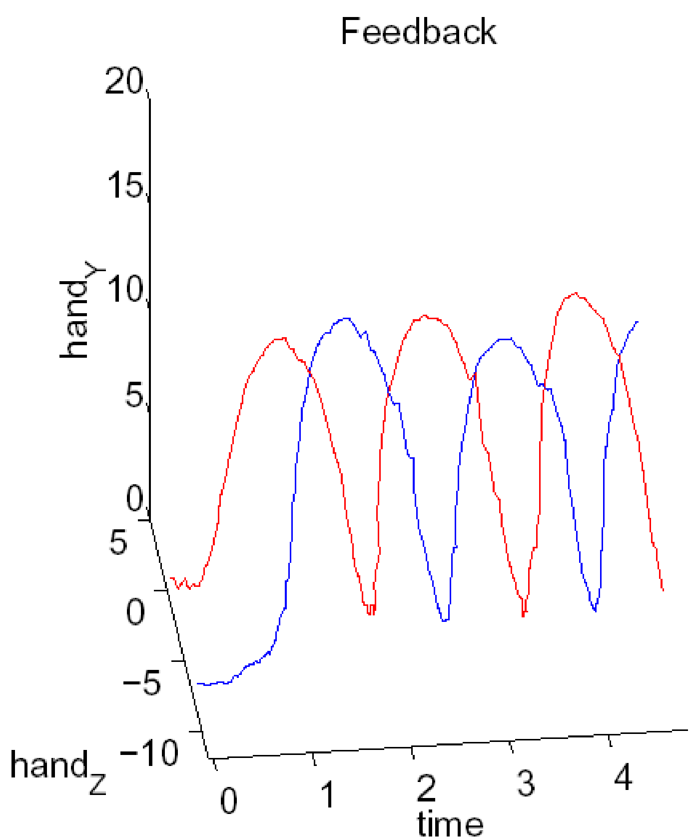


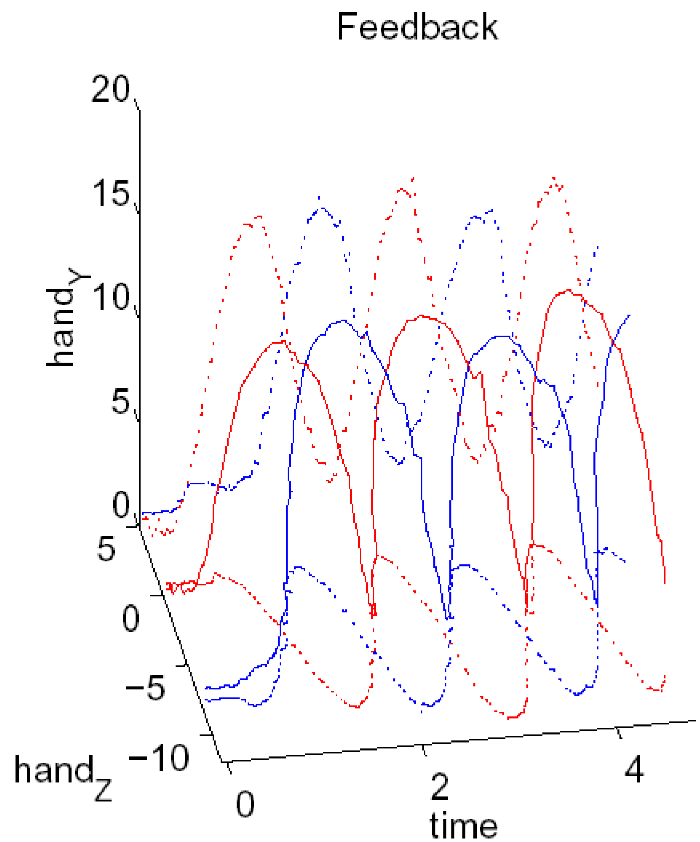




Repeated Occlusions: with recursion







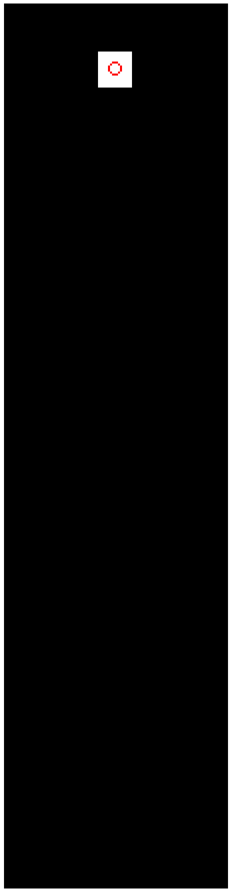
Dyna: the Movie

http://www.drwren.com/chris/9.913/gesture.mpg

"Understanding Expressive Action". Christopher R. Wren, Ph.D. Thesis, MIT EECS, 2000.

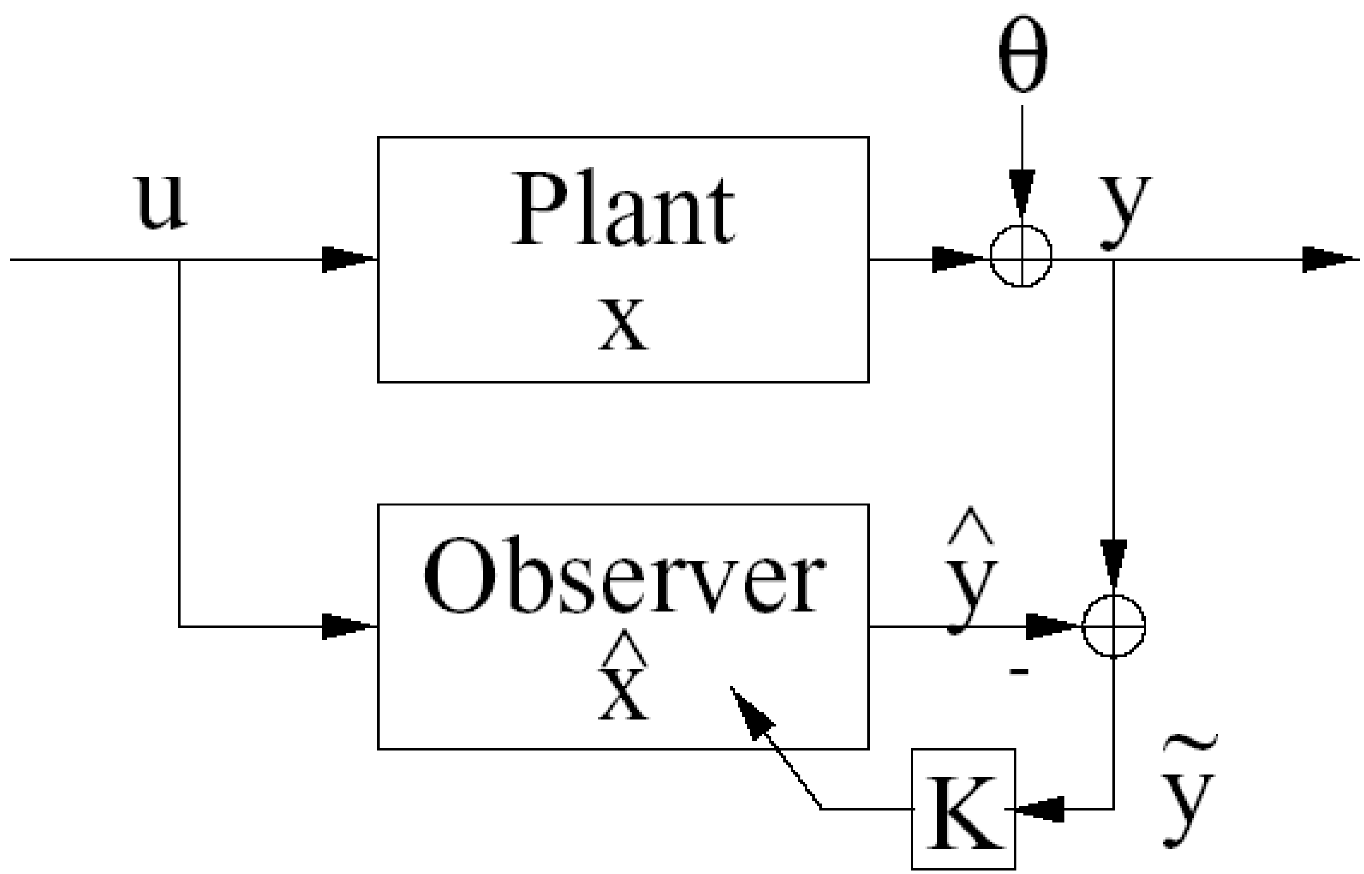
Beyond Tracking

- Dyna models the physics of the human body
- What else is there?
  - Muscles?
  - Nerves?
  - Brain?
- Longer timescales: more complex phenomenon

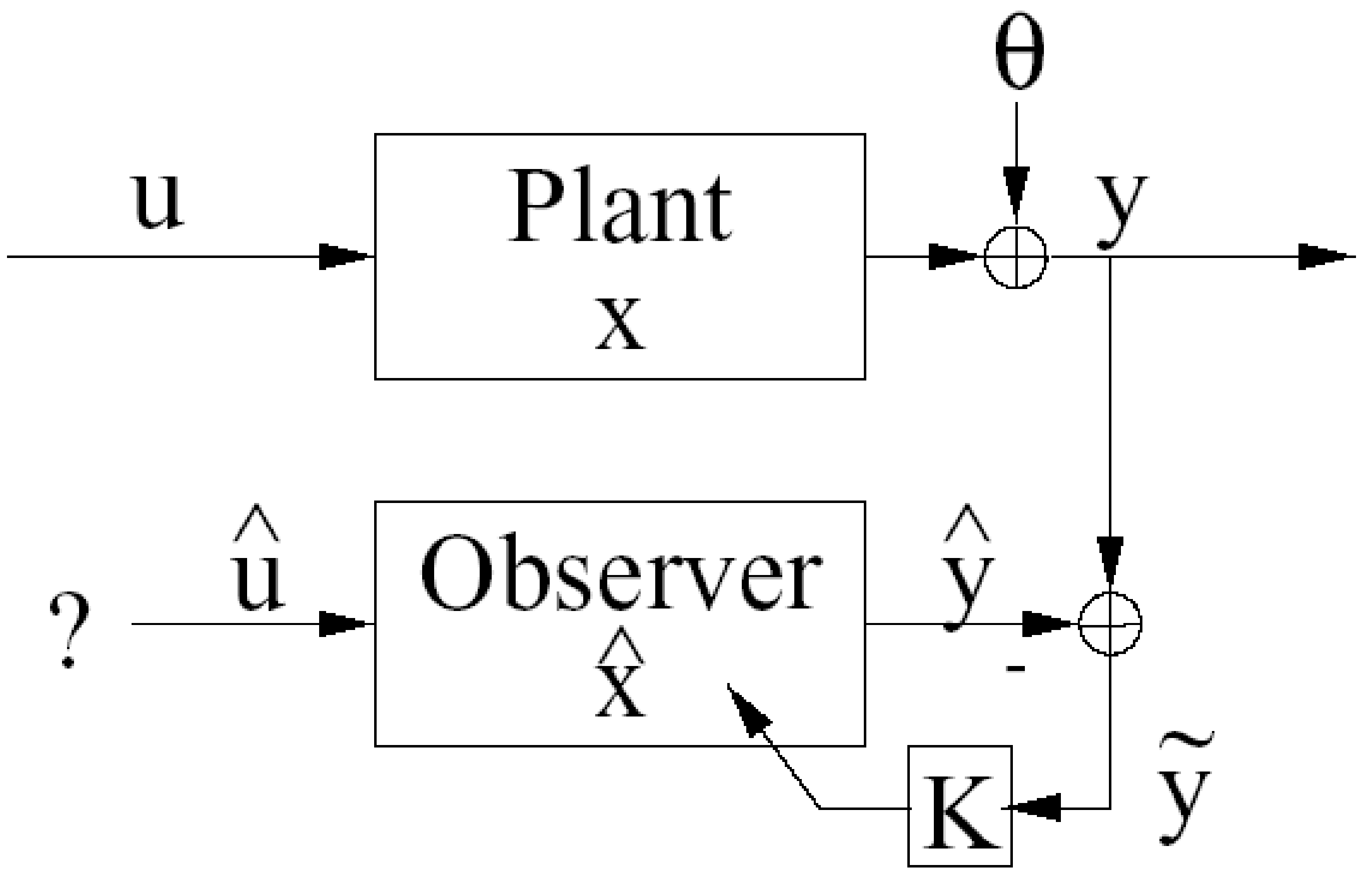


http://www.drwren.com/chris/9.913/stop_track.avi

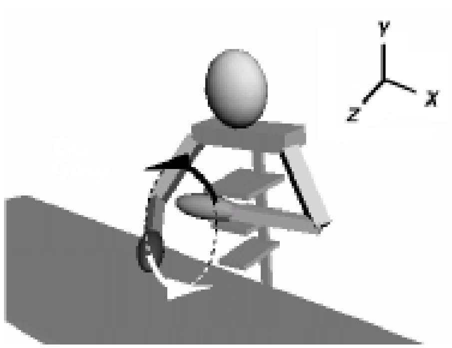
Classic Observer

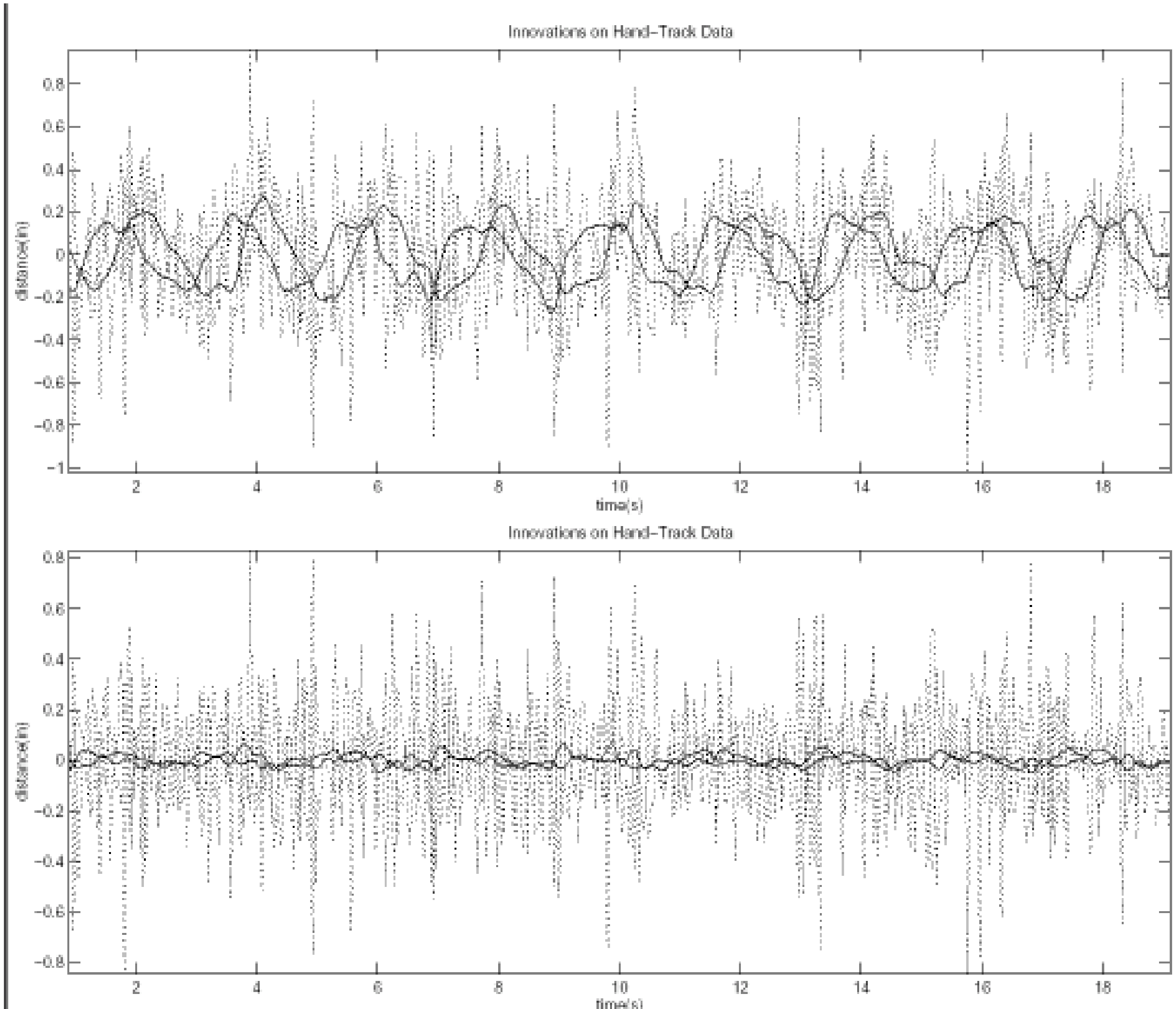


Control?

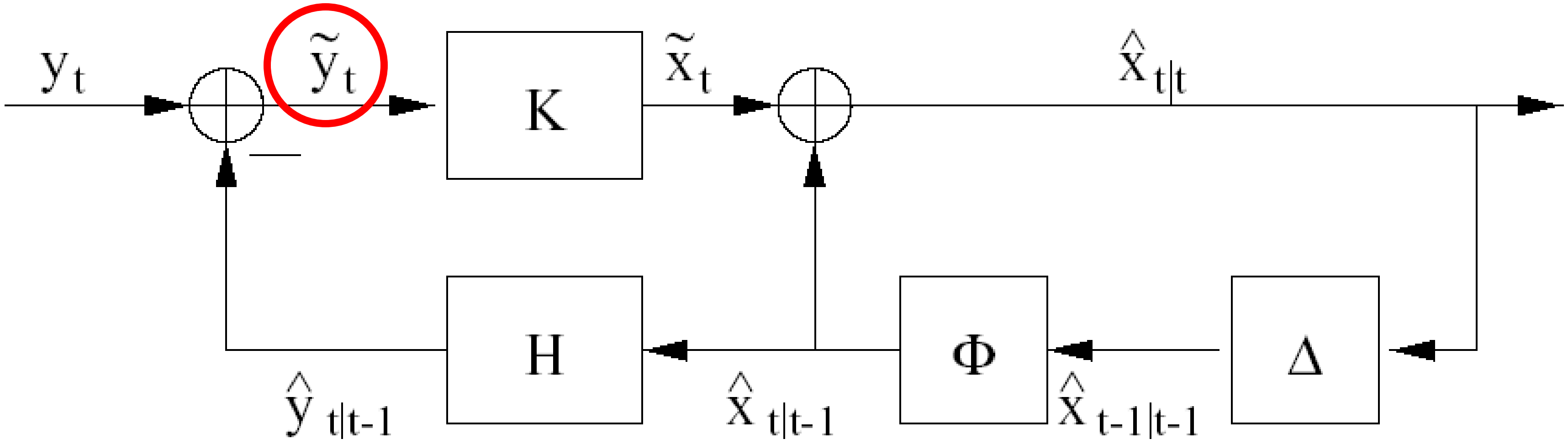


So what happens to the Kalman Filter?



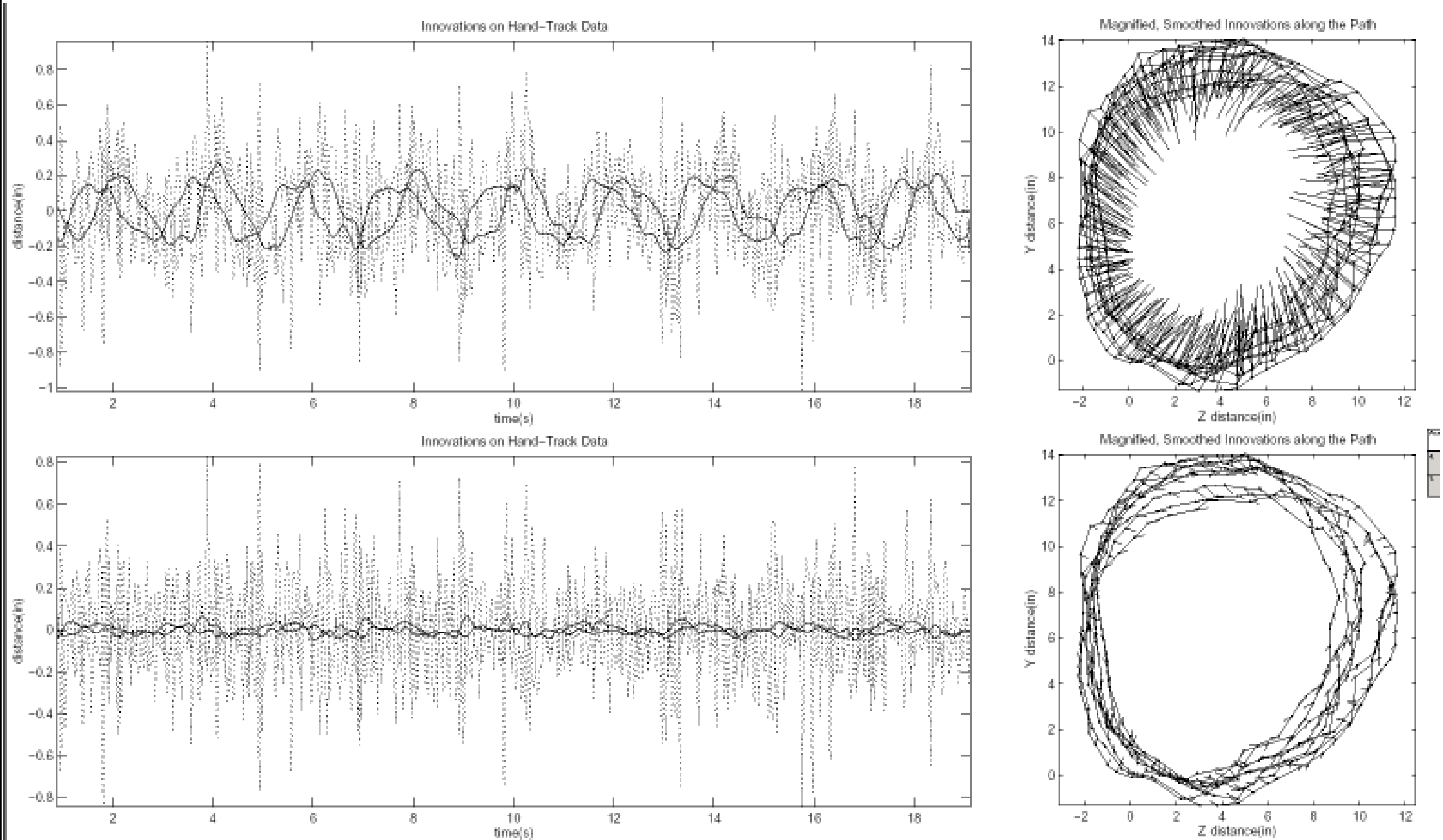


Innovations?



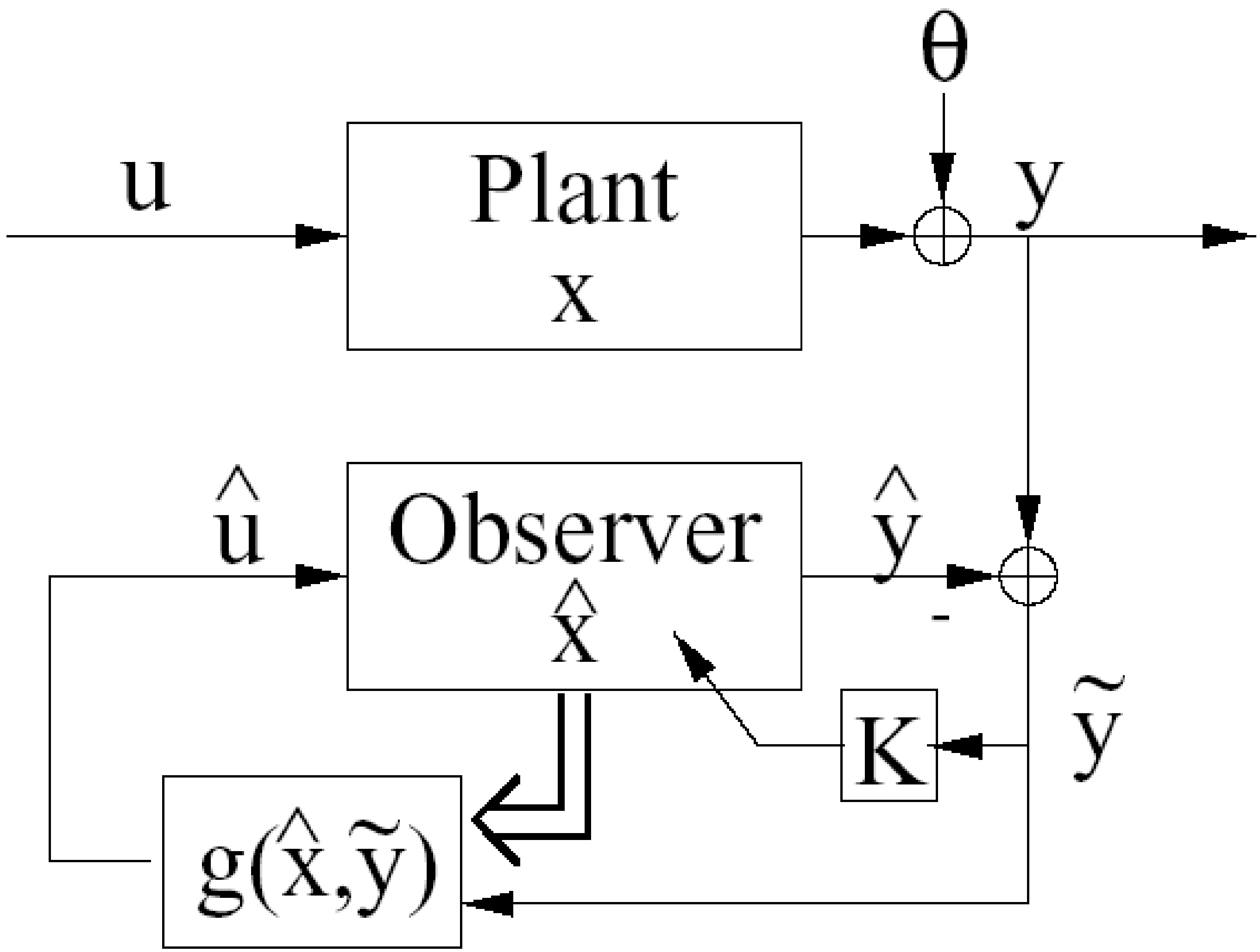
- Innovations
  - Zero-mean, white Gaussian noise
  - Structure in the Innovations means errors in the model
  - Errors: like omitting behavior, habits., etc.

Structured Innovations



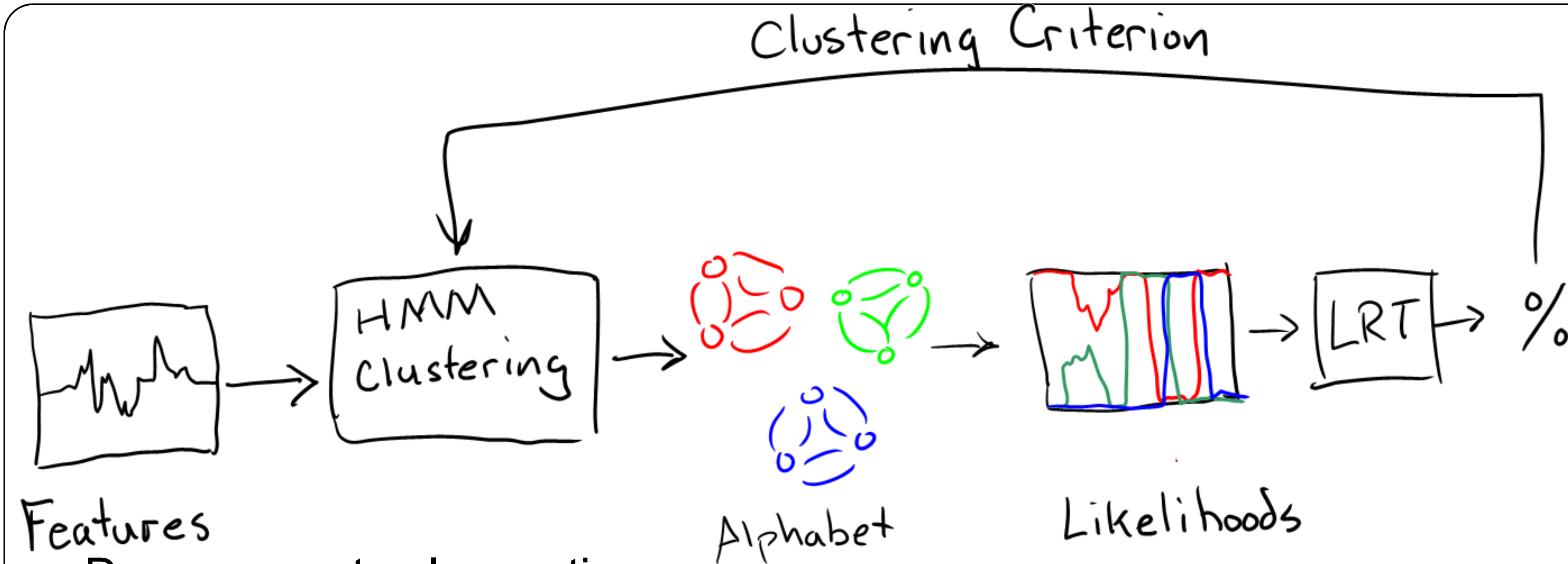
- In this case, the structure is dominated by the active, conscious control .

Learn Control Signal



What's the form of g(x,y)?

Cogno



- Dyna generates Innovations
  - Everything that the model couldn't predict
- Cogno clusters the innovations stream into letters of an alphabet
- The alphabet is used for gesture recognition (like phonemes for speech)
- The alphabet encodes motor programs
  - Generate control signals? (HMMs aren't very good at generating signals)

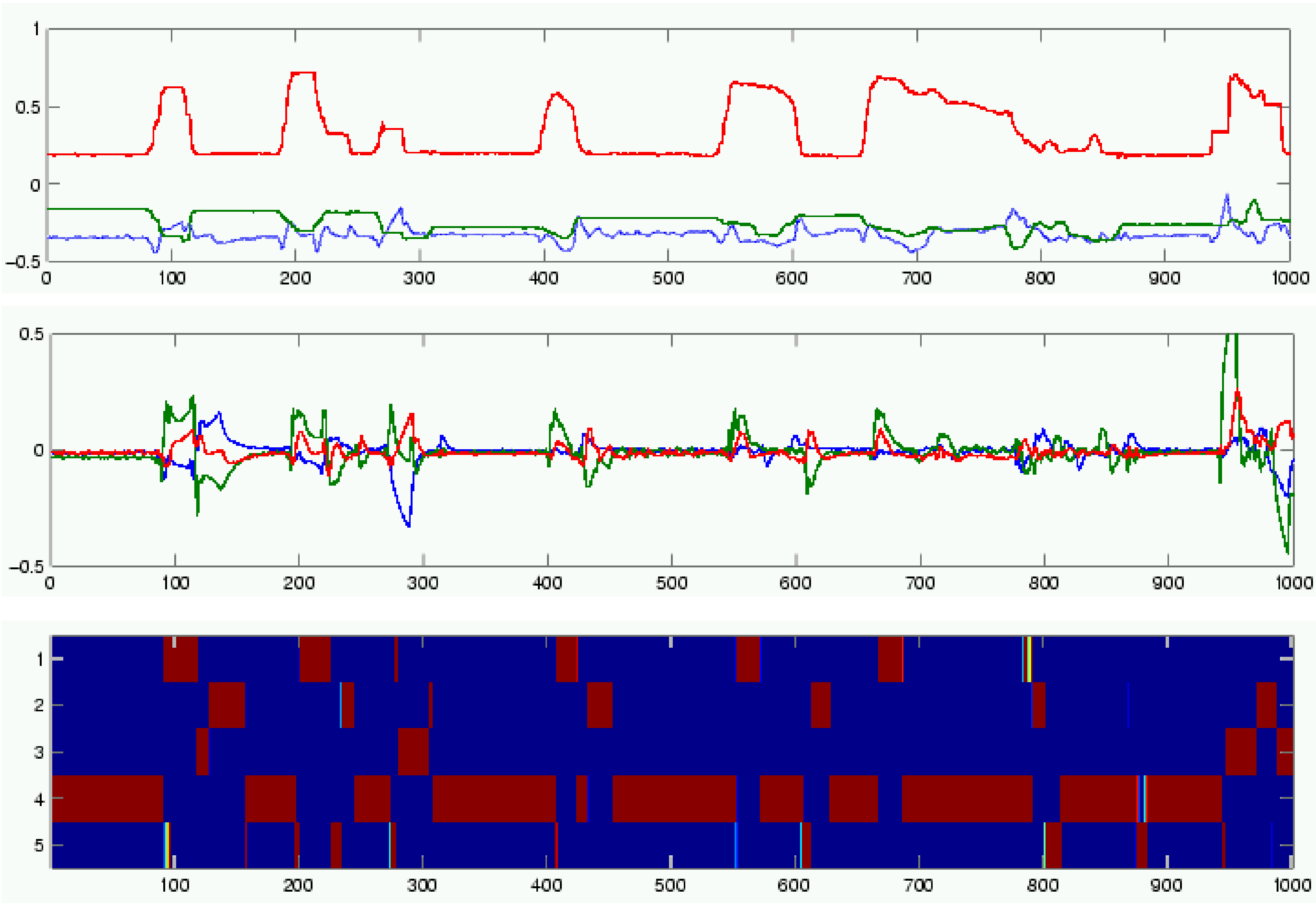
"Unsupervised Clustering of Ambulatory Audio and Video". Clarkson, Pentland, ICAASP 1999..

Task Domain

- Whack-A-Wuggle
- No instruction at all for the subjects
- Functional labels
  - Based on what happened in game
  - Not what the user was actually doing
- Informal experiment (obviously)

http://www.drwren.com/chris/9.913/whacka.mpg

Whacka Data


1
0.5
0
−0.5
0 100 200 300 400 500 600 700 800 900 1000
0.5
0
−0.5
0 100 200 300 400 500 600 700 800 900 1000
1
2
3
4
5
100 200 300 400 500 600 700 800 900 1000

Whacka Results

- 75% overall correct classification rate:

	Whack	Pop	Fail
Whack	87%	10%	03%
Pop	25%	71%	03%
Fail	50%	25%	25%

"Understanding Purposeful Human Motion". Wren, Clarkson, Pentland, Face and Gesture 2000.

Summary

- Tracking is about dynamics
- Appearance models combined with models of dynamics
- Kalman Filters
  – Optimal estimator, given linear models and ZMWG noise
  – Combines observations with a dynamic model
  – A particular recursive filter
- Recursive Filters for Vision
  – Non-linear systems
  – Images are tough observations to deal with
  – Recursive filters generate more than just state estimates
    - Prediction
    - Innovations