LEARNING TO ANTICIPATE GAZE

- Vempati Anurag Sai

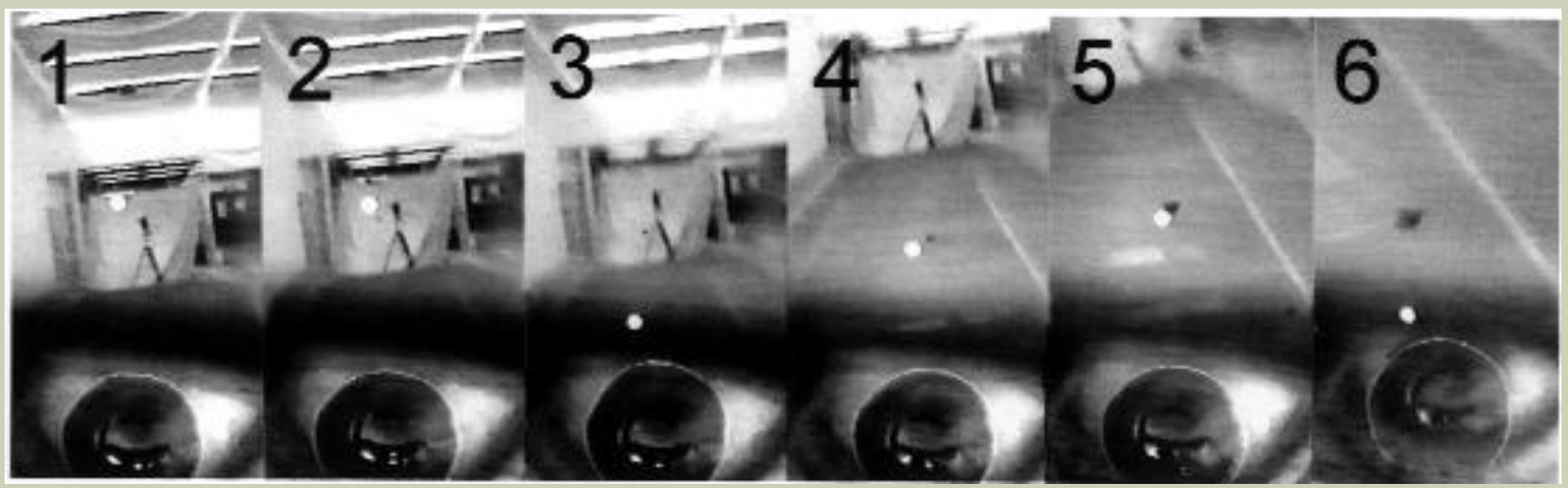
SE367:
Cognitive Science

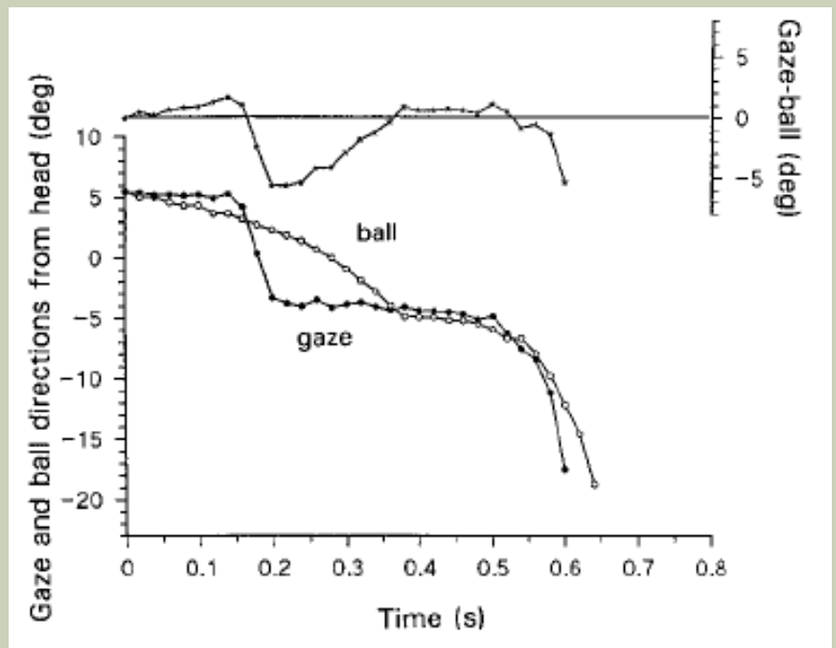
Mentor:

Dr. Amitabha Mukerjee

INTRODUCTION

- Humans: While moving around, driving…
- Google's self driving car.
- Human Gaze – Tightly connected to motor resonance system. [Sciuttu et al.]
- Sports persons.
  - Batsmen's eye movements monitor the moment when the ball is released, make a predictive saccade to the place where they expect it to hit the ground, wait for it to bounce, and follow its trajectory for 100–200 ms after the bounce. [Land & McLeod]





[Images courtesy of Land & McLeod, 2000]

MECHANISM

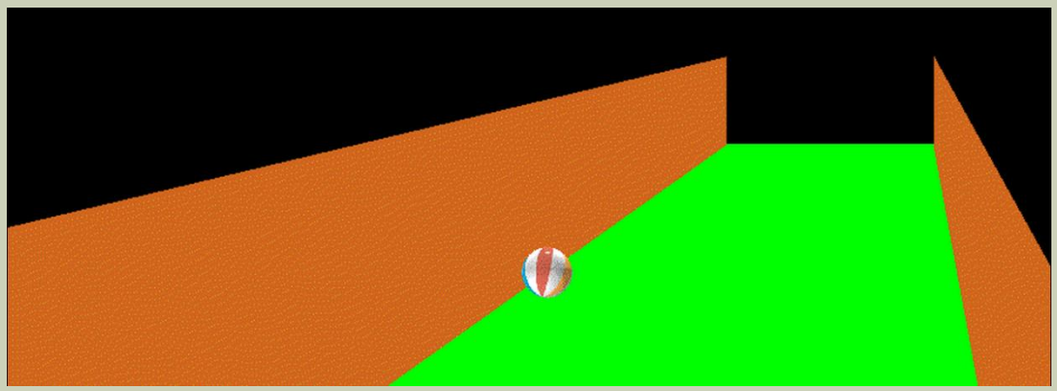
- The model is learnt in unsupervised fashion.
- **DATASET:** Sequences of a ball bouncing off the walls/floor at different velocities as viewed from different viewpoints with a facility to incorporate desired spin/swing.

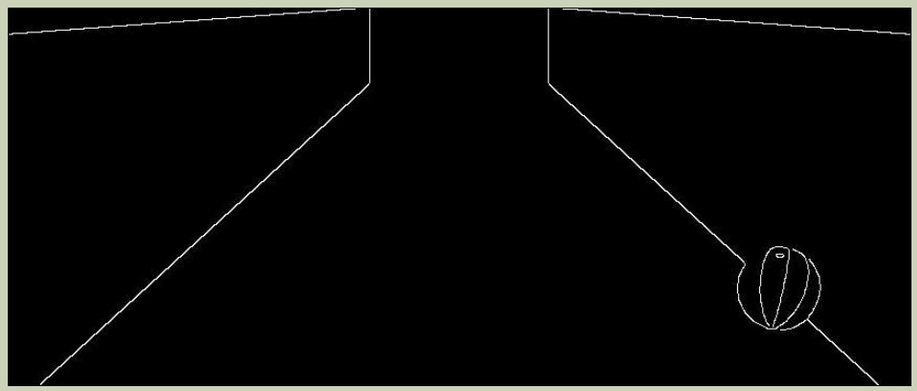
$$\begin{pmatrix} x_G \\ v_G \\ a_G \end{pmatrix} = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_G \\ v_G \\ a_G \end{pmatrix}$$

$$P_{new} = R_G^{new} P_G + O_G^{new}$$

$$X_{VISUAL} = f * X / Z$$

- 'Canny edge detector' + 'Hough Transform'





MECHANISM

- Using (x, y, z) pairs in the dataset, learn the state transition matrix **F**.
- Regression problem.

$$\begin{pmatrix} x \\ v_x \\ a_x \\ y \\ v_y \\ a_y \\ z \\ v_z \\ a_z \end{pmatrix} = \begin{bmatrix} A_{3\times3} & O_{3\times3} & O_{3\times3} \\ O_{3\times3} & A_{3\times3} & O_{3\times3} \\ O_{3\times3} & O_{3\times3} & A_{3\times3} \end{bmatrix} \begin{pmatrix} x \\ v_x \\ a_x \\ y \\ v_y \\ a_y \\ z \\ v_z \\ a_z \end{pmatrix}$$

$$A_{3\times3} = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}$$

State Transition Matrix

State vector

LEAST SQUARES ESTIMATE

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

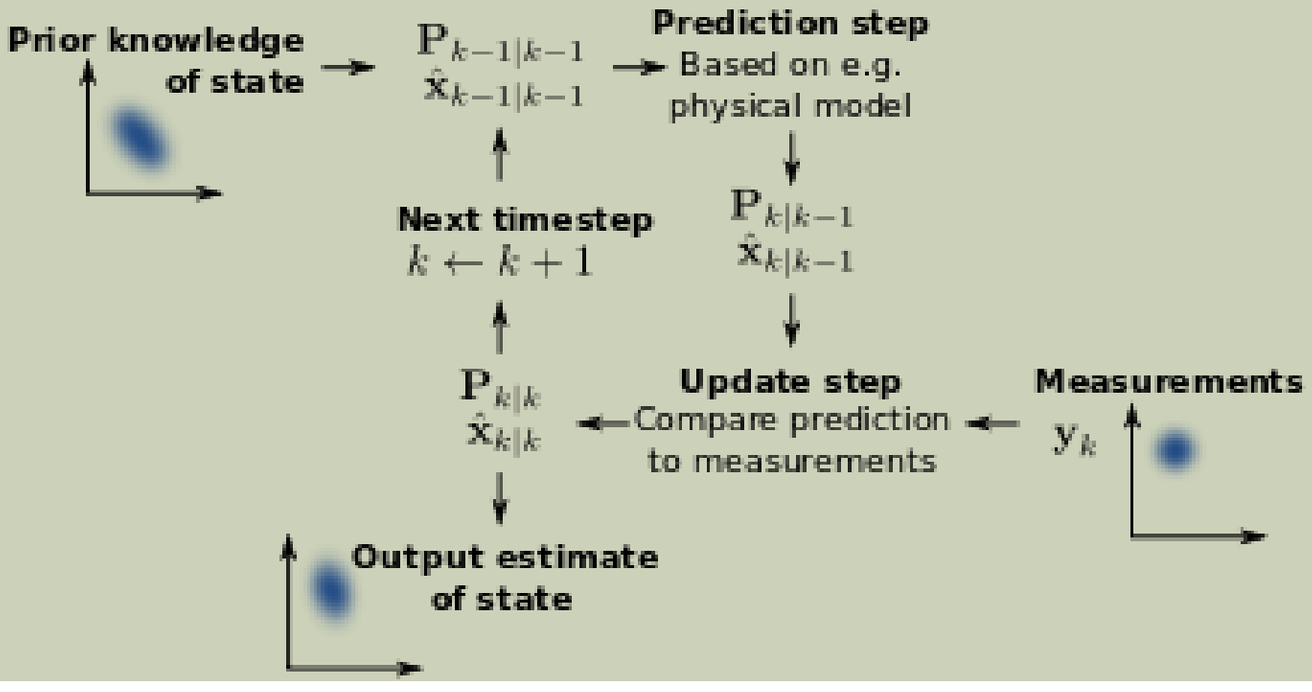
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \left(\tfrac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\tfrac{1}{n}\sum \mathbf{x}_i y_i\right).$$

KALMAN FILTER

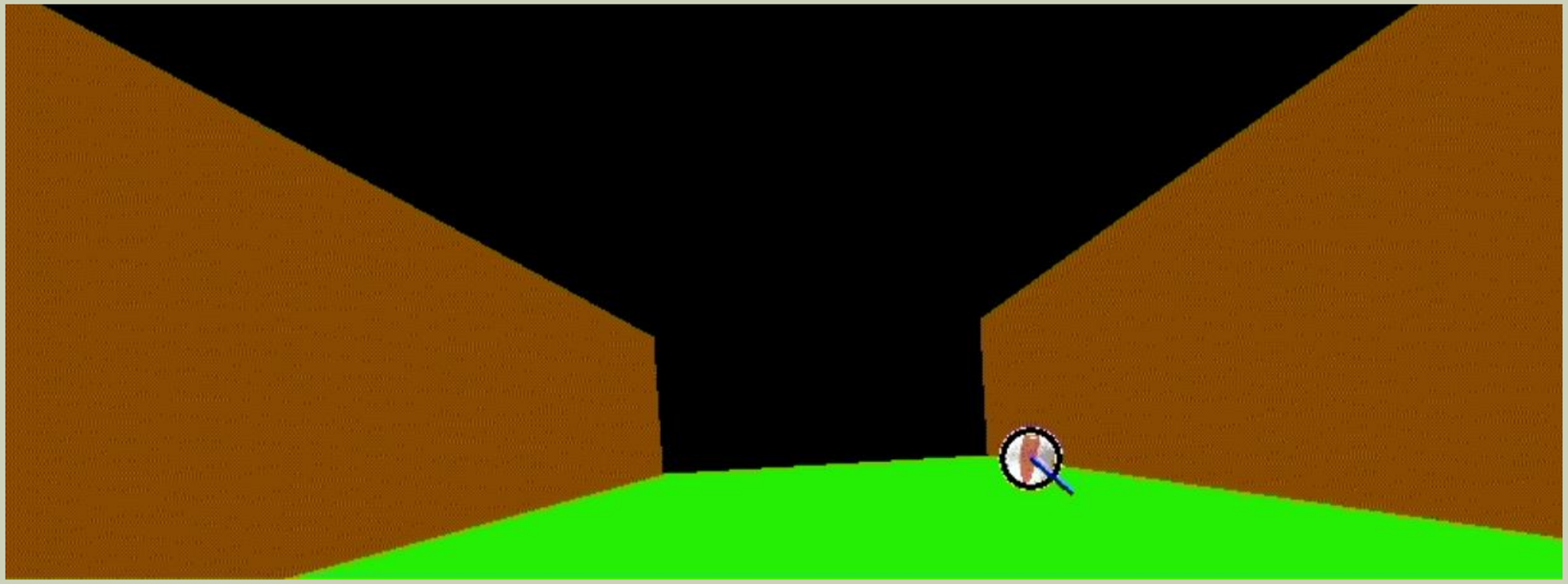
- Why?
  - Takes care of Noisy Measurements (Rounding effects/Edge detector…)
  - Just the measurement of position will do
  - Several cycles of prediction can be done before next measurement update

$$\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{u}_k + \mathbf{w}_k$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$



RESULTS & REFERENCES



- Land, Michael F., and Peter McLeod. "From eye movements to actions: how batsmen hit the ball." Nature neuroscience 3.12 (2000): 1340-1345.
- Sciutti, Alessandra, et al. "Anticipatory gaze in human-robot interactions."*Gaze in HRI from modeling to communication" workshop at the 7th ACM/IEEE international conference on human-robot interaction, Boston, Massachusetts, USA*. 2012.
- Perse, Matej, et al. "Physics-based modelling of human motion using kalman filter and collision avoidance algorithm." International Symposium on Image and Signal Processing and Analysis, ISPA05, Zagreb, Croatia. 2005.
- http://en.wikipedia.org/wiki/Kalman_filter