Learning to Anticipate Gaze: Top-Down Approach

Mentor:
Dr. Amitabha Mukerjee

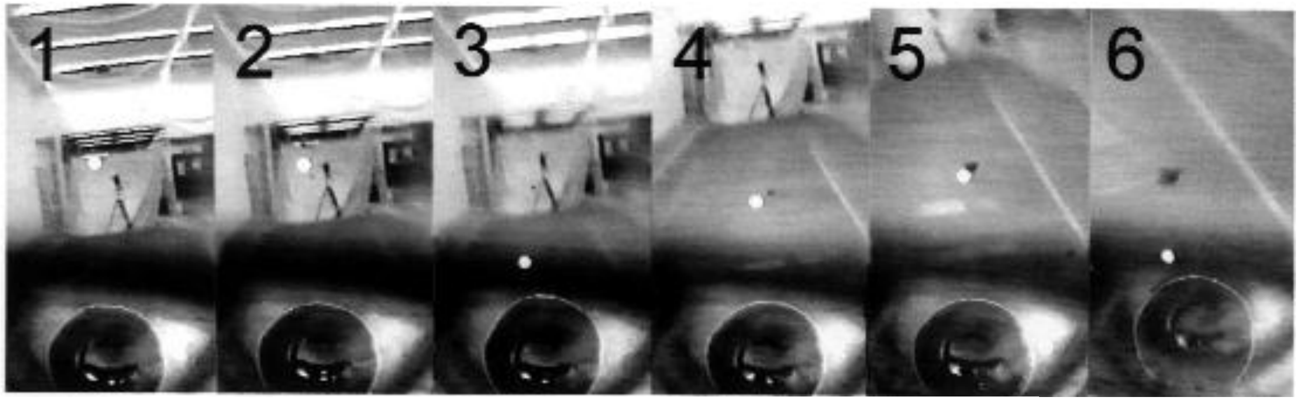
Presented by
Vempati Anurag Sai

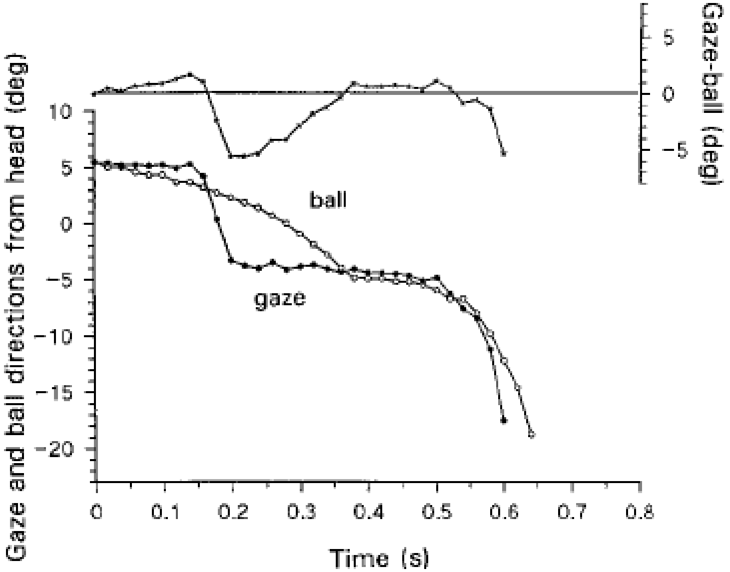
SE367 – Cognitive Science

Introduction

- Humans deploy anticipatory gaze in many situations. While moving around, driving…
- Google's self driving car has a Kalman Filter that tracks each and every vehicle in its sight and anticipates their future positions so that it doesn't run into them.
- 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]

Introduction



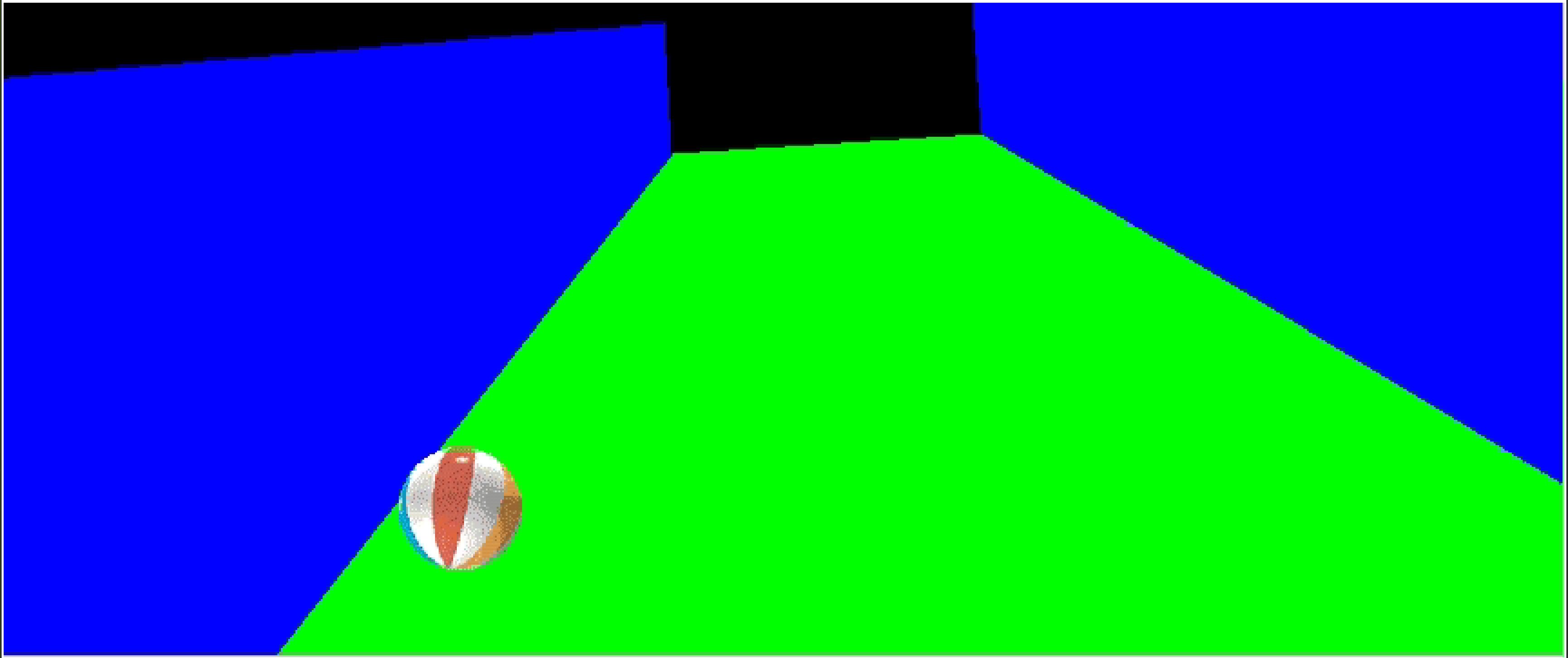


Mechanism

- Basically, hoping to achieve the degree of anticipation as in a professional cricketer
- The model is learnt in unsupervised fashion.
- Various sequences of a ball bouncing off the walls/floor viewed from different viewpoints is created for the training phase.

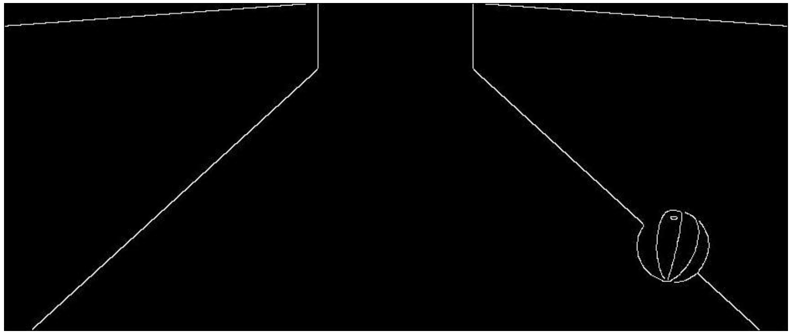
$$\begin{pmatrix} x_G \\ v_G \\ a_G \end{pmatrix} = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_G \\ v_G \\ a_G \end{pmatrix}$$

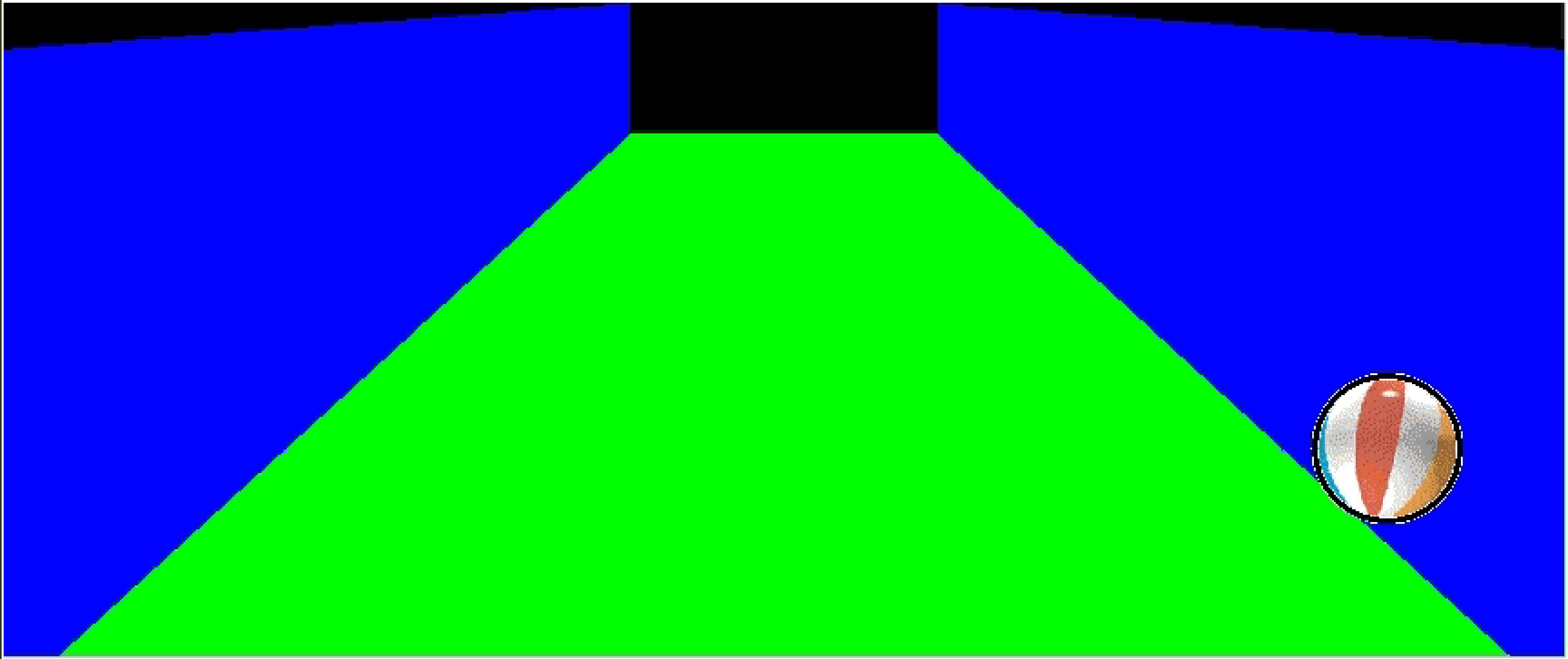
$$P_{new} = R_G^{new} P_G + O_G^{new}$$



Mechanism

- Then we search for any moving round objects. The pixel coordinates and size of the ball are stored to get a dataset for training phase.
- Segmentation/ Optical flow will be a better choice in general. But, since we know the shape of object, better options are available.
- ‘Canny edge detector’ + ‘Hough Transform’





Mechanism

- Size of the ball gives 'z' component.
- Using (x, y, z) pairs in the dataset, learn the state transition matrix **F**.
- Regression problem.

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

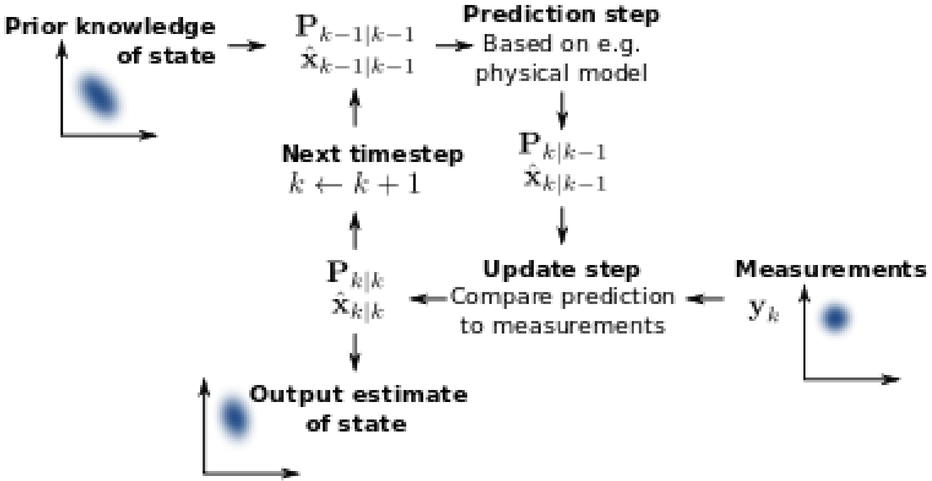
$$A_{3\times3} = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}$$

State Transition Matrix

State vector

Mechanism

- Kalman Filter is then used to predict the trajectory in advance.
- Why Kalman Filter?
  - Takes care of Noisy Measurements
  - Just the measurement of position will do
  - Several cycles of prediction can be done before next measurement update



Kalman Filter

- Assumes the true state at time k is evolved from the state at (k-1) according to:

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

  - $\mathbf{F}_k$ is the state transition model which is applied to the previous state $\mathbf{x}_{k-1}$
  - $\mathbf{B}_k$ is the control-input model which is applied to the control vector $\mathbf{u}_k$
  - $\mathbf{w}_k$ is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance $\mathbf{Q}_k$.
- At time k an observation (or measurement) $\mathbf{z}_k$ of the true state $\mathbf{x}_k$ is made according to

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

  - where $\mathbf{H}_k$ is the observation model which maps the true state space into the observed space and $\mathbf{v}_k$ is the observation noise which is assumed to be zero mean Gaussian noise with covariance $\mathbf{R}_k$

What next?

- Evaluate performance on real videos
- Answer the bigger question!
- Better Learning Paradigm
- Compare human gaze anticipation with the developed model

REFERENCES

I. Land, Michael F., and Peter McLeod. "From eye movements to actions: how batsmen hit the ball." Nature neuroscience 3.12 (2000): 1340-1345.
II. 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.
III. 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.
IV. http://en.wikipedia.org/wiki/Kalman_filter

QUESTIONS??