# Learning video saliency from human gaze using candidate selection [ Rudoy et al, CVPR 2013]

Saliency map: probability of where viewer looks

**Claims:**

* Video saliency tends to be tighter and concentrated on single object while image saliency covers several interesting locations.
* Video saliency is sparse and computing it at every pixel is redundant.
* Video saliency is conditional to the previous frames, while image saliency is independent for every image

**Selecting Candidate locations**

1. **Static candidates**

Graph based visual saliency, find peaks using mean shift and fit Gaussian of in neighborhood size of h/5

1. **Motion candidates**

Optical flow with DoG filtering and mean-shift clustering and Gaussian fitting

1. **Semantic candidate**

- Center

- Poselet based detector : head, 2 shoulders, torso, 2 eyes, nose and mouth

**Modeling gaze dynamics**

1. **Features**
   1. **Static**

Local contrast in a neighborhood around candidate points

* 1. **Motion**

DoG of horizontal and vertical components of optical flow around candidate points

* 1. **Semantic**

Face and person detection scores

Labels : motion, saliency, face, body, center

1. Gaze transitions for training

Scene cuts to 15 frames after that (it takes 5-10 frames to fixate)

1. Learning transition probability

Random forest classifier

**Validation**

* Comparision with fixation data smoothed with Gaussian
* AUC is used for comparison

**Results**

* Candidate selection performs better than dense estimation
* Outperforms state-of-the-art methods on DIEM dataset