

A Sketch-Based Approach To Video Retrieval Using Qualitative Features

Koustav Ghosal and Anoop Namboodiri

Center for Visual Information Technology

IIIT Hyderabad, India

Content Based Video Retrieval

Different Form Of Queries :

Text Example Sketch

Text Based Query :

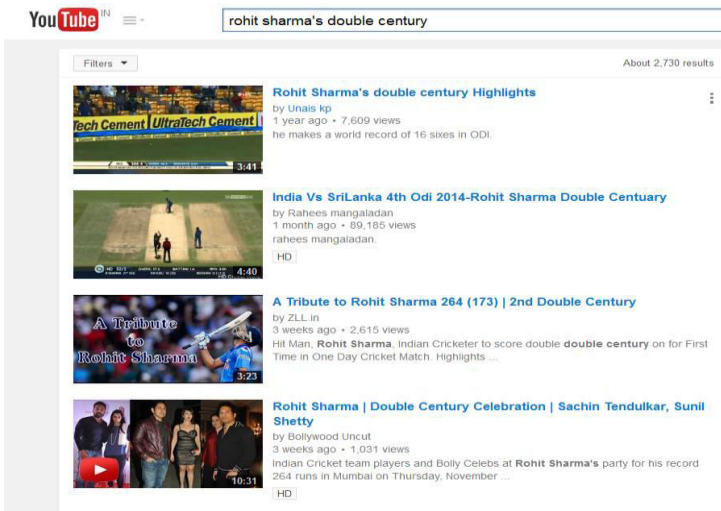
Video Search from Youtube

“ Coke Studio Pakistan ”

“ Markov random Field Daphne Koller ”

“ Rohit Sharma's Double Century ”

Metadata : **Tags**, **Comments** etc.



Content Based Video Retrieval

Different Form Of Queries :

Text **Example** Sketch

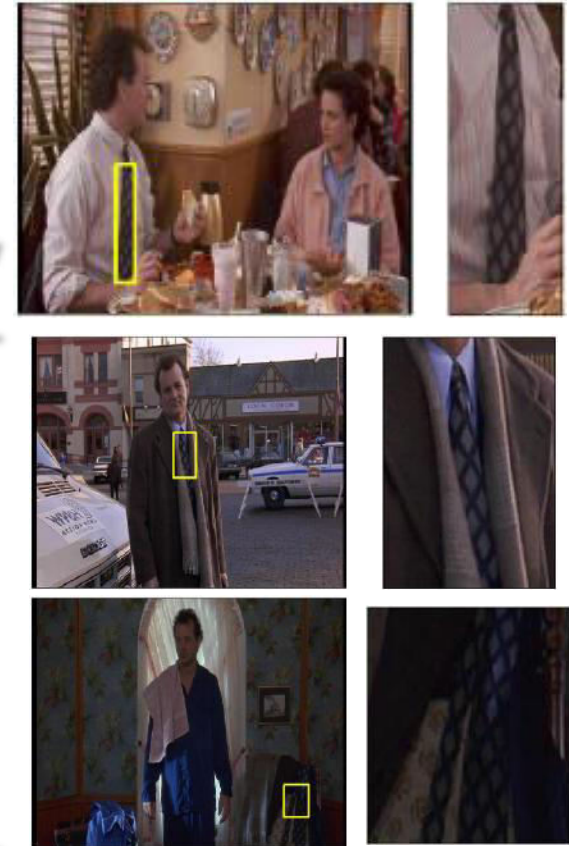
Example Based Query :
Video Google¹

Input :

Image or Objects of interest

Output :

Relevant Shots or Key Frames

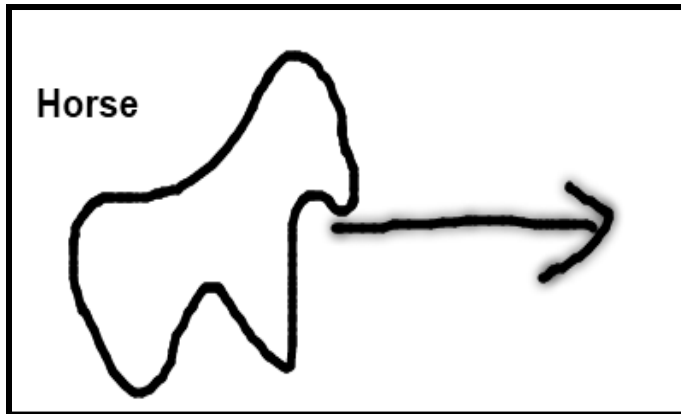


1. Sivic et al, ICCV 2003

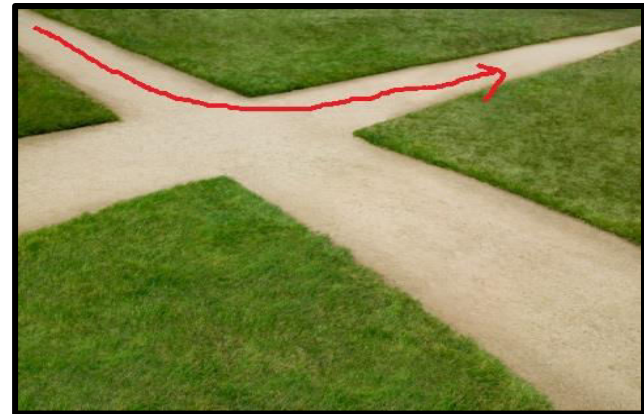
Content Based Video Retrieval

Different Form Of Queries :

Text Example **Sketch**



Video Q ¹



Surveillance ²

1. Chen and Chang : Fifth ACM international conference on Multimedia. ACM, 1997.
2. Hu et al : PAMI 2006

Sketch vs Text/Example based



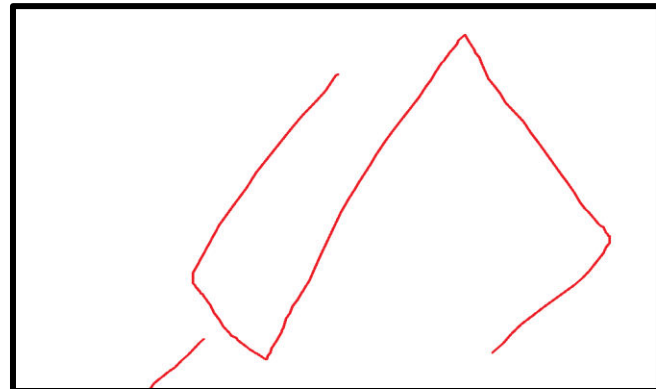
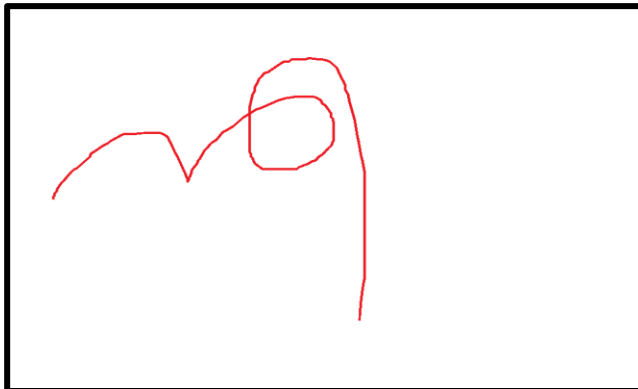
“ A diving video where the diver does a somersault before going down ... ”

Example is not always available.



“A pool shot where the player strikes from below and the white ball moves north-east, then north west then south-west and finally strikes another ball. The second ball drops inside the pocket ... ”

Sketch vs Text/Example based

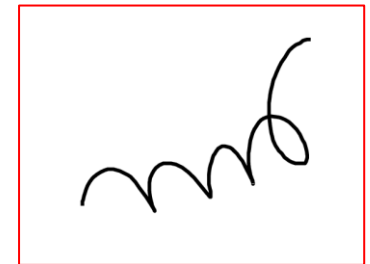
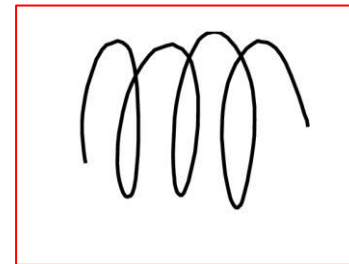
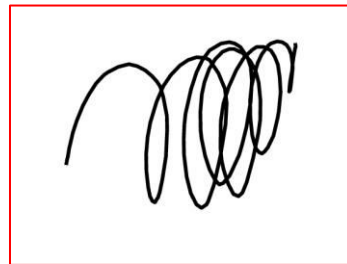
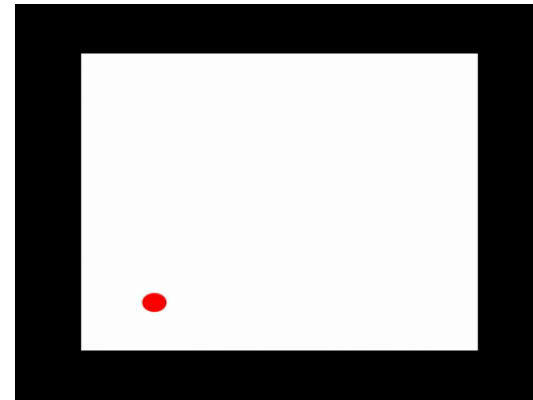


Challenges

Background Extraction

- Camera Motion
- Occlusion
- Illumination
- Background Noise
- and many more...

Perceptual Variability



Users perceive the same motion differently



Qualitative vs Quantitative Feature Space

Approximations instead of absolute numeric values to remove perceptual variability

Qualitative Spatio-Temporal Features

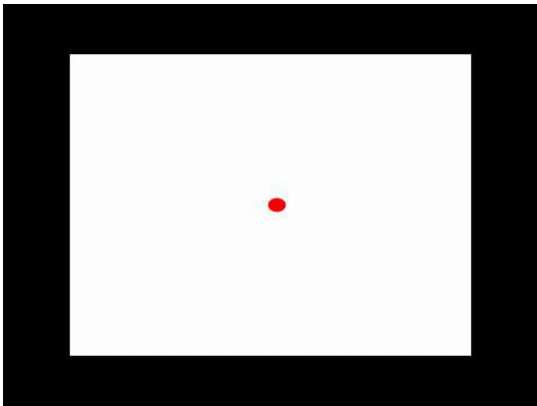
Features that capture “how” (qualitative) objects moved rather than “how much” (quantitative)

Problem Statement

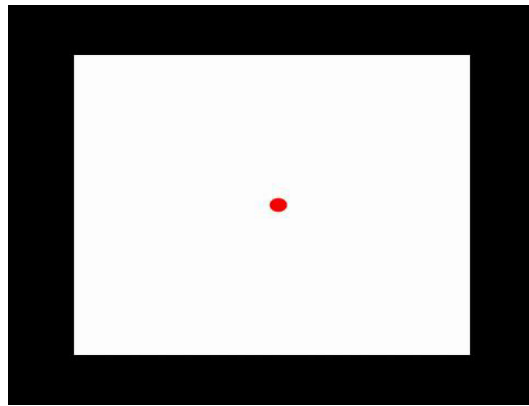
- **The Problem Statement** : Modelling the query (user sketch) and original trajectories using “*qualitative*” features instead of “*quantitative*” features to remove perceptual variability.
- **Assumption** : Trajectories are available for a set of videos.

Aspects of Motion

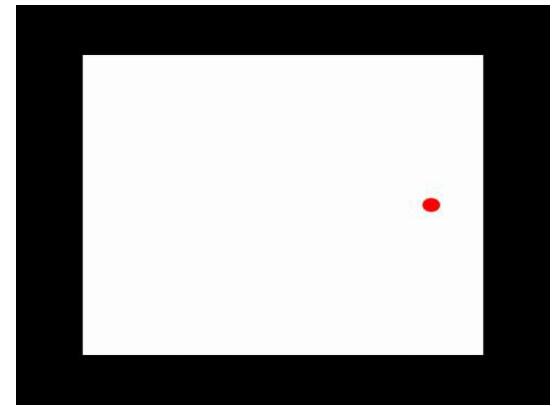
- **Shape:** Linear and non-linear sub-trajectories.
- **Direction:** Similar in shape but have different directions
- **Scale:** Different scales of same motion



S-A-I-O



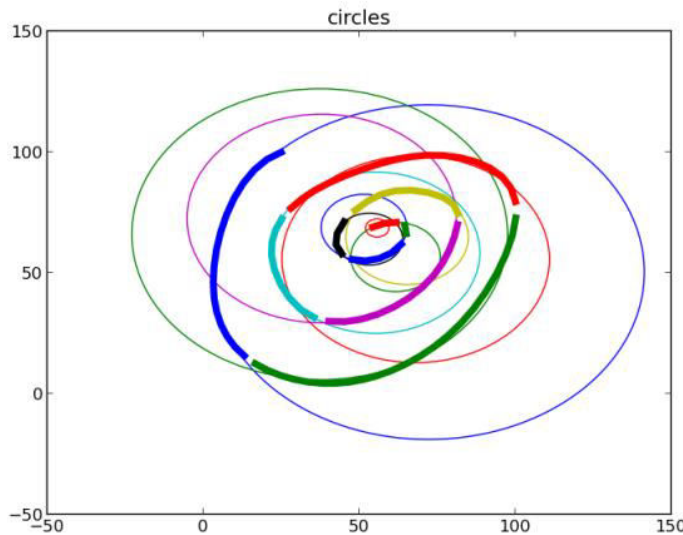
S-C-I-O



S-C-O-I

Shape : Circle Based Features

$$J = \min_{x_0, y_0, r} \sum_i^n x_i^2 + y_i^2 - 2x_0x_i - 2y_0y_i + x_0^2 + y_0^2 + r^2$$



$$S = (x_\mu, y_\mu, r, m, s)$$

(x_μ, y_μ) , r = Center , Radius

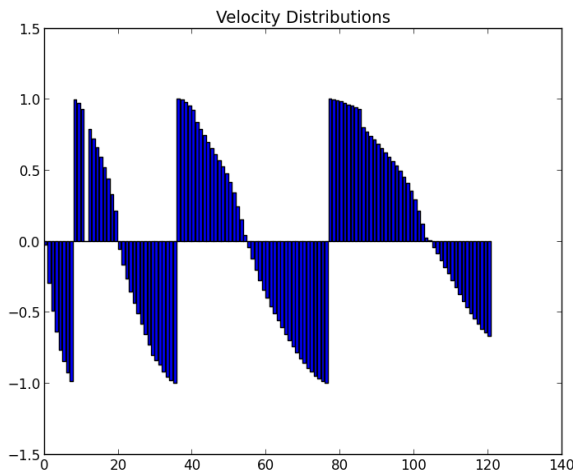
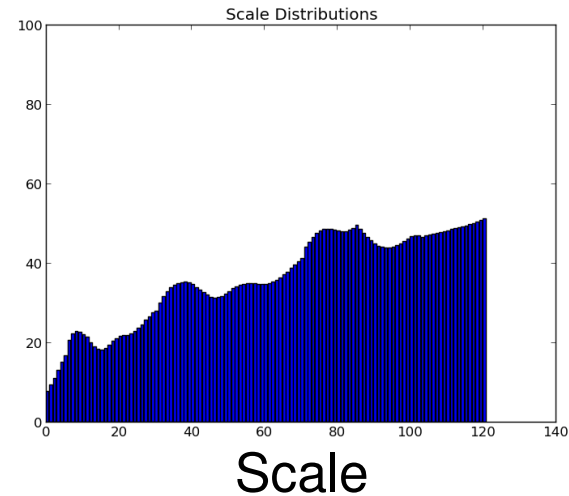
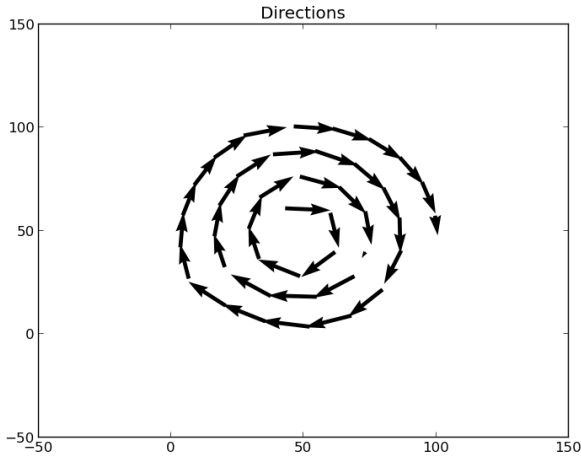
m = Slope

s = Normalized Length

K- Means , Bag-Of-Motion

Trajectory = Histogram of motion-segments = Loss of temporal Information !!

Direction And Scale



Direction

$$\text{Trajectory} = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

$$\alpha_k = \sin(\theta_k) \quad \dots (1)$$

$$\text{Trajectory} = (d_1, d_2, \dots, d_n)$$

$$d_k = \text{Distance from mean} \quad \dots (2)$$

Summary of Features

1. Bag-Of-Motion

2. Ordered Motion Segments (s_1, s_2, \dots, s_m) where
 $s_k = (x_\mu, y_\mu, r, m, s)$

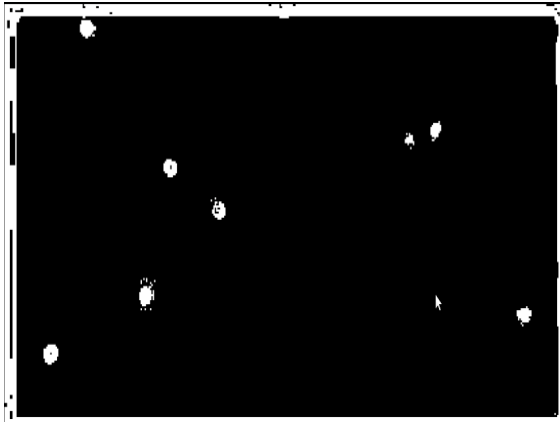
3. Change of Direction $(\alpha_1, \alpha_2, \dots, \alpha_n)$

4. Change of Scale (d_1, d_2, \dots, d_n)

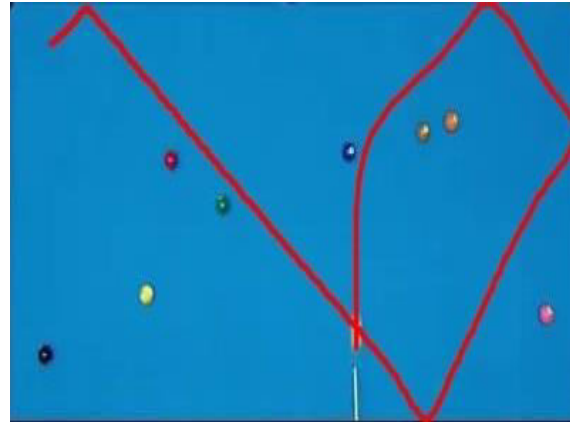
Both For User-Sketches and Videos

Dataset

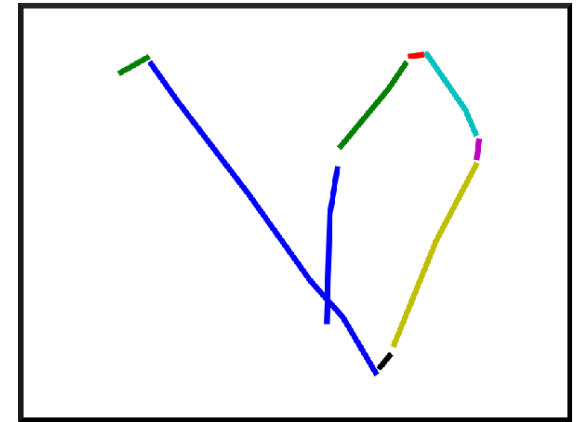
- 100 Synthetic videos and 100 pool videos



Foreground



Kalman Filter Tracking

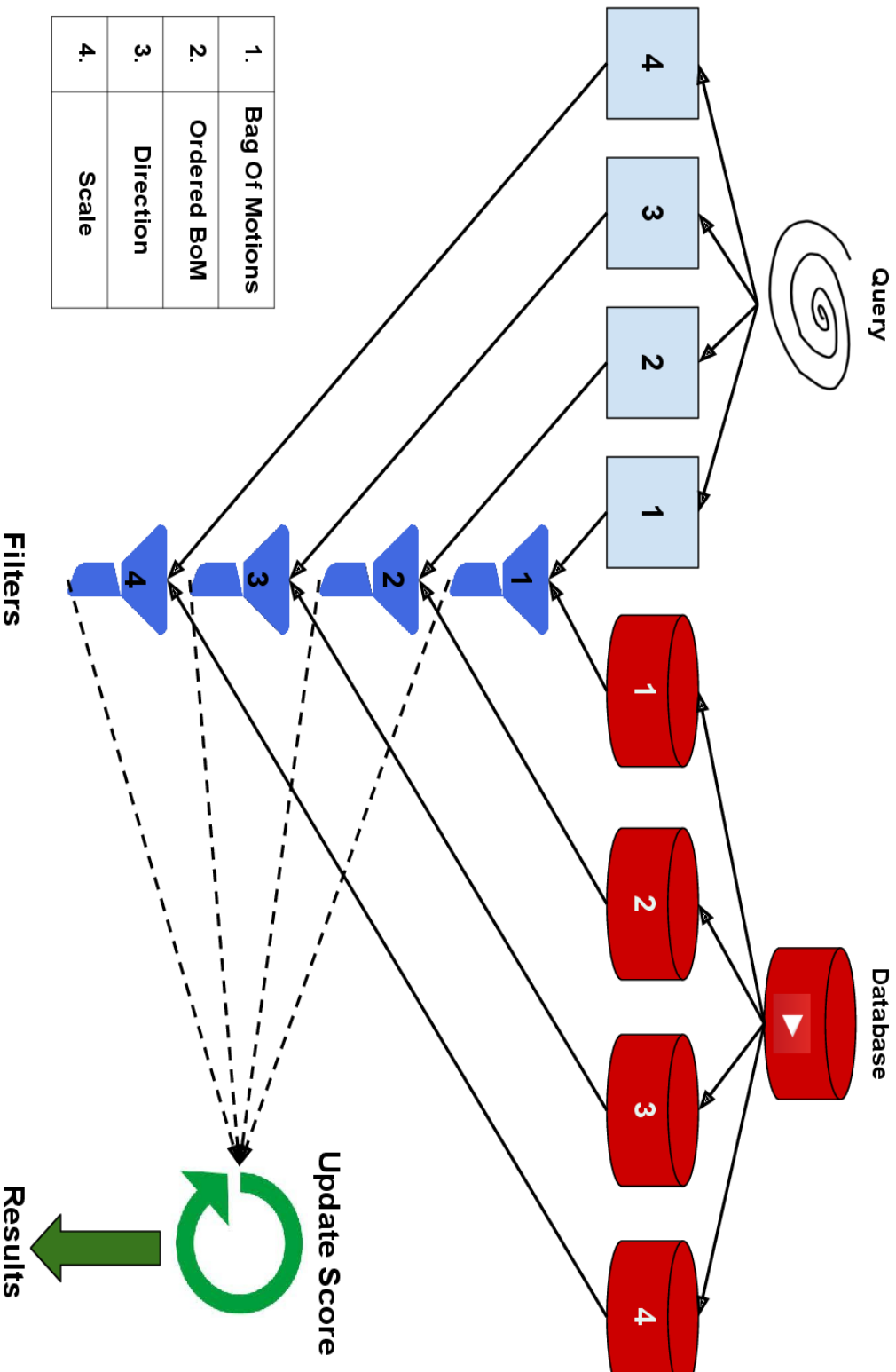


M- Segments

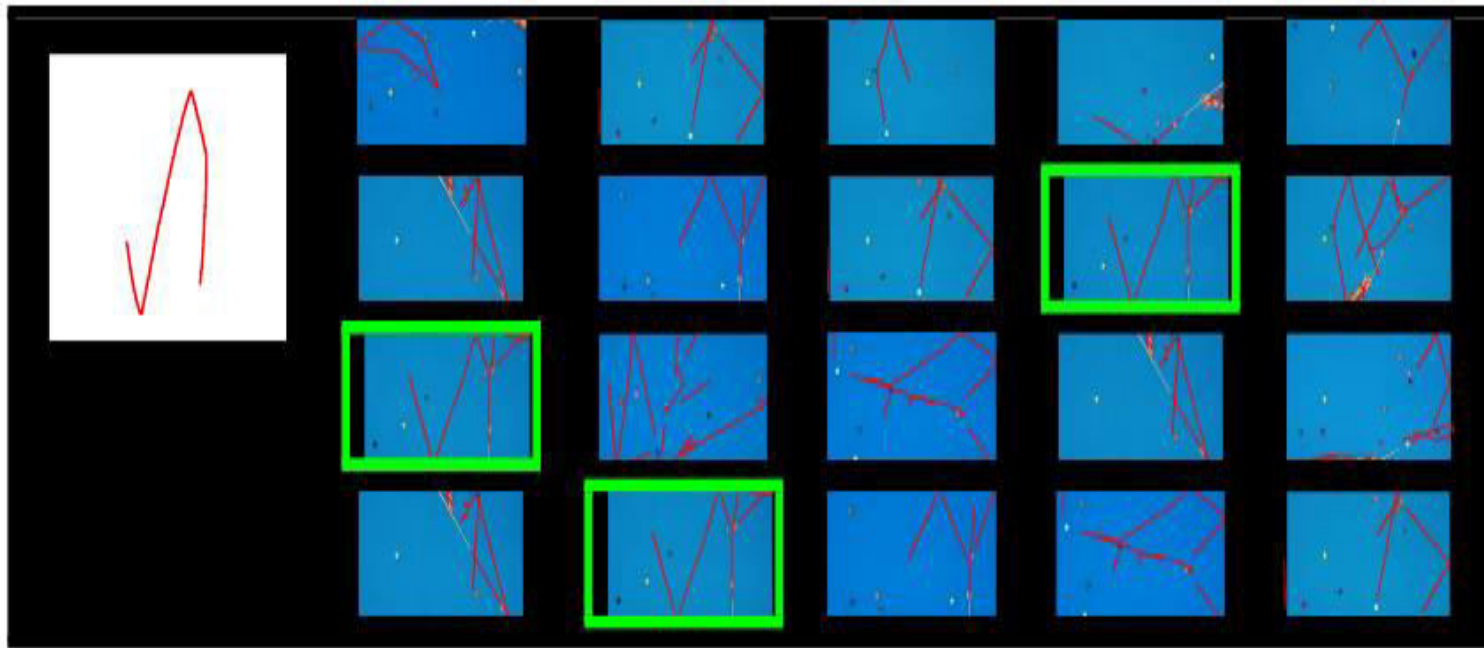
- Data was collected from 50 different users
- Interval of 6 hours between video and sketch



Cascade



Results : An example query



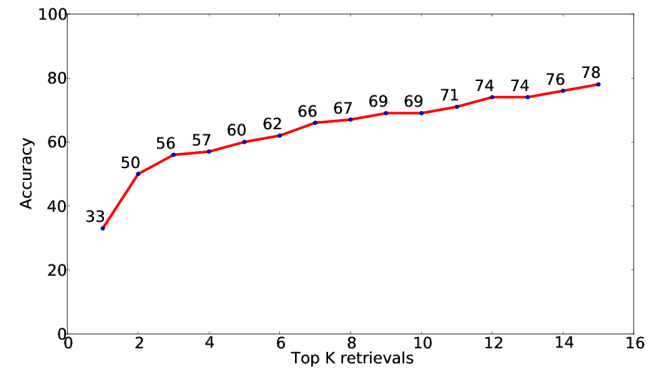
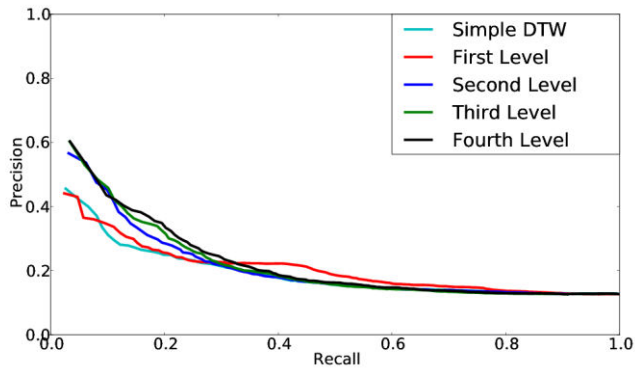
BOM

Sequence of
BOM

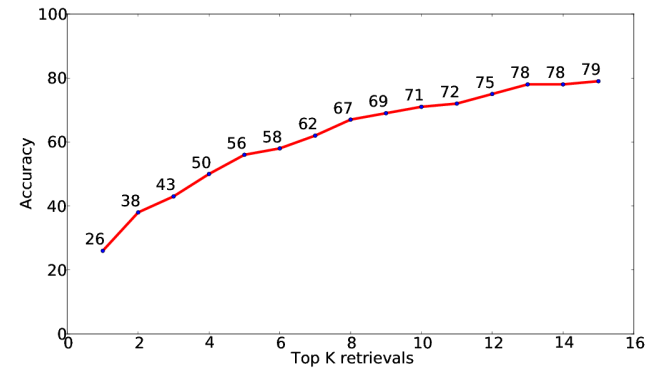
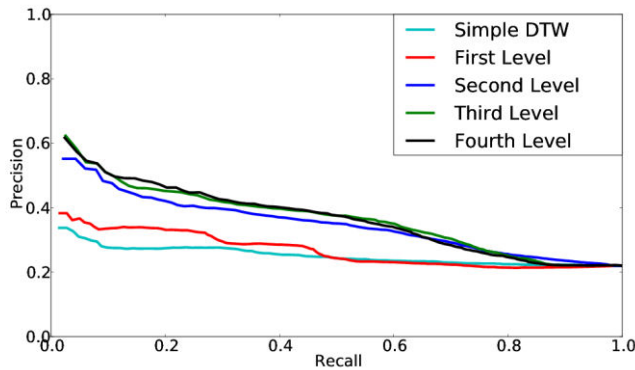
Sequence of
Directions

Sequence of
Scales

Results



Pool



Synthetic

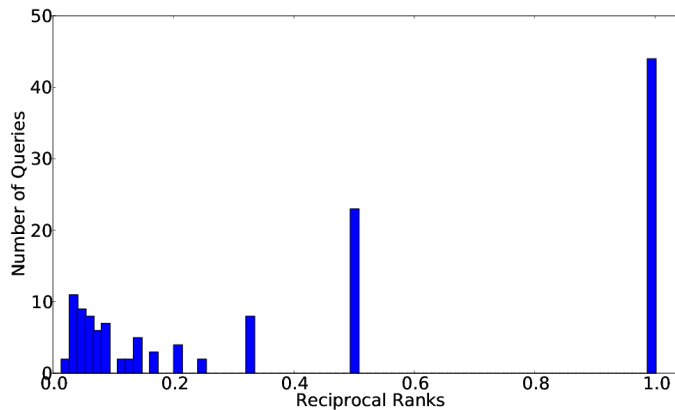
Precision - Recall

Top K Accuracy

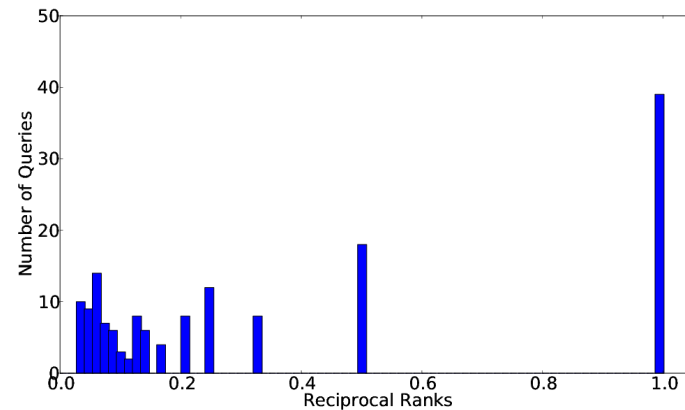
Results

Mean Reciprocal rank

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$



Pool



Synthetic

Summary

- Strengths
 - **Unconstrained Query** : No initial frame
 - **Qualitative Features** : Robust to user-level variations
 - **A novel retrieval strategy** : Cumulative Scoring Mechanism
- Limitations
 - **Dependency** : A strong trajectory extraction algorithm
 - **Not generic** : Doesn't work for videos where motion is not the most salient feature

Future Work

- Finding new features to expand the scope of this representation to more generic videos
- Using additional information about object color and shape to refine the search
- Generating consistent tracks from videos with camera motion and large perspective variations



Thank you