Semantic-Based Surveillance Video Retrieval

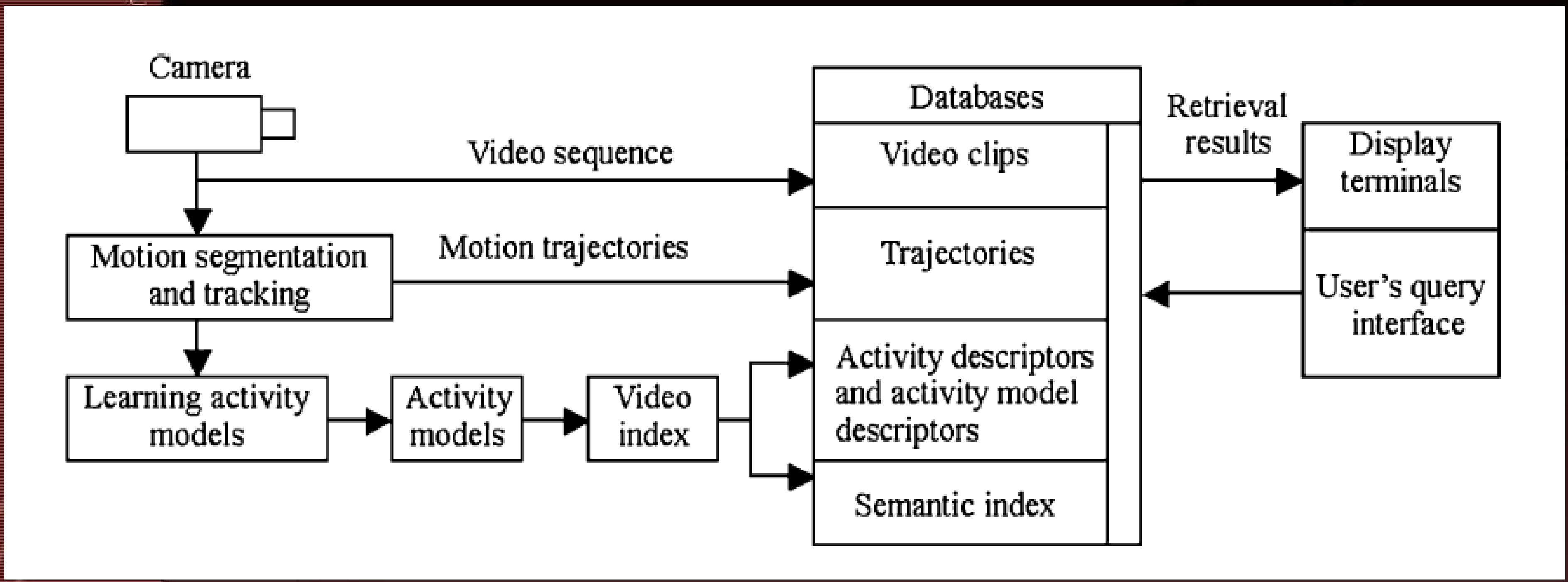
Weiming Hu, Dan Xie, Zhouyu Fu,
Wenrong Zeng, and Steve Maybank,
Senior Member, IEEE

IEEE Transactions on Image Processing,
Vol. 16, No. 4, April 2007

National Taiwan University
CMLAB , since 1991

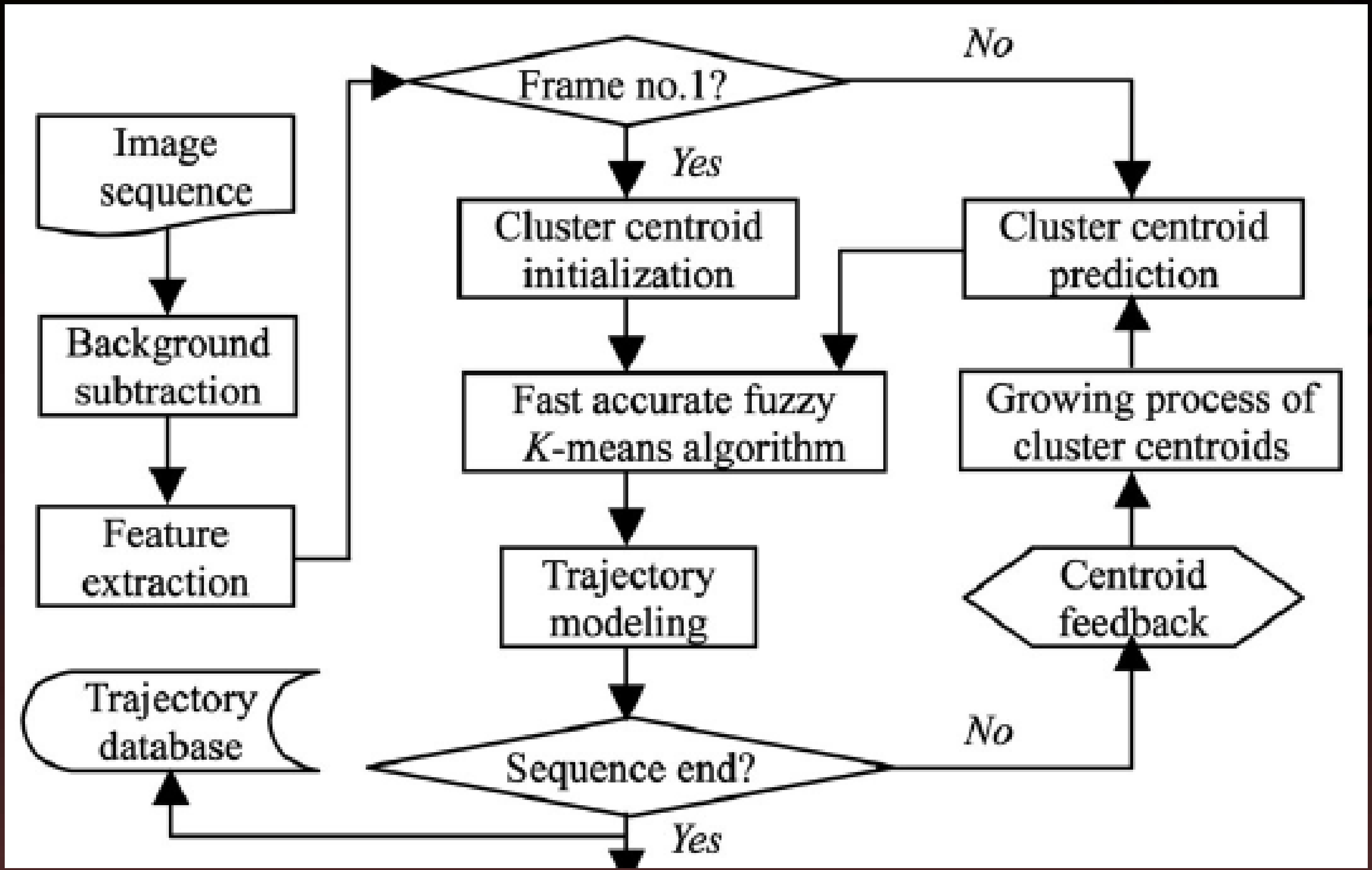
Present by komod

Introduction



Object Tracking

Multiple object tracking



Pixel Features

- Foreground pixels are acquired by a self-adaptable background update model
- Each foreground pixel is described with a feature vector f

$$f = (x, y, v_x, v_y, r, g, b).$$

Component Quantization Filtering

- The image plane is partitioned into square regions with equal size
- X_i - sample vector, mean of the feature vectors of the foreground pixels in the region
- w_i - weight, number of foreground pixels in this region

Fuzzy c-means Clustering

- In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster
  - $u_k(x)$: the degree of being in the k_{th} cluster

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\mathrm{center}_k, x)}{d(\mathrm{center}_j, x)} \right)^{2/(m-1)}}.$$

$$\mathrm{center}_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m}.$$

- For m = 2, this is equivalent to normalising the coefficient linearly to make their sum 1. When m is close to 1, algorithm is similar to k-means

Fuzzy c-means Clustering

- X_l – sample feature vectors
- V_i - vector of cluster centroid
- M - number of sample feature vectors
- N - dimension of the sample feature vectors
- K - number of cluster centroids
- Fuzzy membership

$$R_{lj}(t) = \frac{1/d_{lj}^2(t)}{\sum_{m=1}^{K} (1/d_{lm}^2(t))},\ 1 \le l \le M,\ 1 \le j \le K.$$

Fuzzy c-means Clustering

- cluster centroid initialization
  - first frame : random select
  - otherwise : prediction from previous frame
- cluster centroid update

$$V_{ji}(t+1) = V_{ji}(t) + \frac{\sum_{l=1}^{M} R_{lj}(t) \cdot w_l \cdot (X_{li} - V_{ji}(t))}{\sum_{l=1}^{M} R_{lj}(t) \cdot w_l}$$

$$1 \leq i \leq N,\ 1 \leq j \leq K$$

Dynamic Growing of Centroids

- entering and leaving regions are manually defined
- Creation
  - we find a subset of samples where the Euclidean distance between each of these samples and its associated cluster centroid j exceeds a threshold Φ_j
- Erasure
  - The position of cluster centroid j is within a leaving region
  - The number of the samples corresponding to cluster centroid j is too small to represent the smallest object in the scene.

Modeling of Cluster Centroids

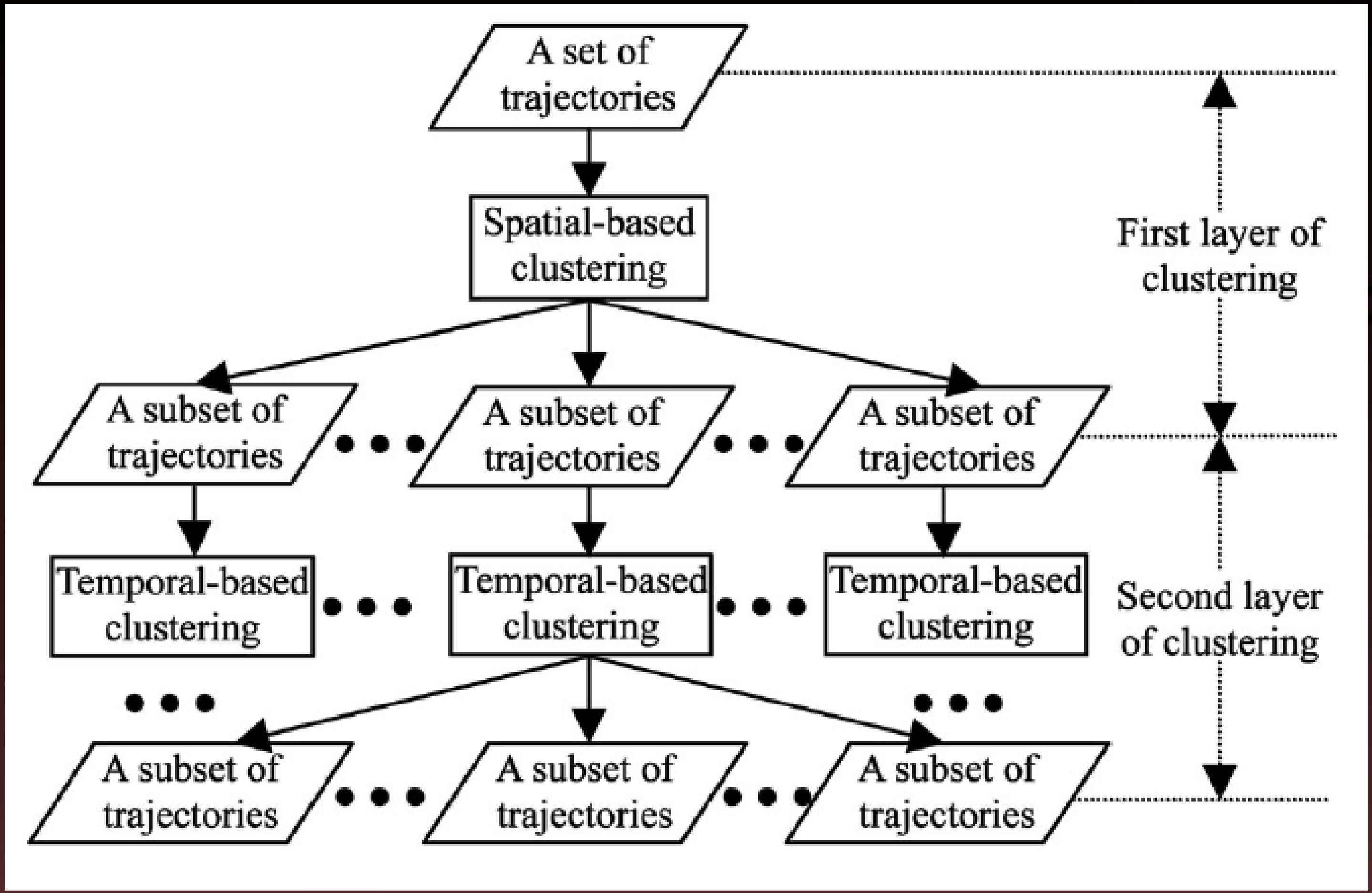
- There may be objects which correspond to two or more cluster centroids in one frame
- For two centroid trajectories exist over the same sequence of frames, if the differences between the centroids in each frame are approximately constant and small, two trajectories are merged

Semantic Activity Models

Activity Model

- Activity models are learned from the trajectories obtained by tracking
- Spatiotemporal trajectory T_{ST}
  - $T_{ST} = \{f_1, f_2, \ldots, f_i, \ldots, f_n\}$
  - $f_i = (x_i, y_i, v_{x_i}, v_{y_i})$
- An activity model describes a category of activities with similar semantic meanings

Hierarchical clustering



Hierarchical clustering

- Spatial
  - Activities observed in different road lanes or routes are assigned to different activity models
- Temporal
  - Objects which pass along the same lane or route may have different activities producing different activity models

Spatial-based Clustering

- Spectral Clustering

Step 1: Compute the similarity matrix A for the data set X by (6).

Step 2: Construct matrix L [23]

$$L = D^{-1/2} A D^{-1/2} \tag{7}$$

where D is a diagonal matrix whose ith diagonal element is the sum of all elements in the ith row of matrix A.

Step 3: Apply the eigenvalue decomposition to matrix L to find out the K eigenvectors $q_1, q_2, \ldots, q_K$, corresponding to the K largest eigenvalues. The eigenvectors are represented as column vectors.

Step 4: Form a new $M \times K$ matrix $Q = [q_1, q_2, \ldots, q_k]$ by stacking the K eigenvectors in columns, and normalize each row of Q to unit length.

Step 5: Cluster the M row vectors of Q into K clusters, using the fuzzy c-means algorithm by treating each row as a new feature vector corresponding to the vector in the original data set X.

Temporal-based Clustering

- Spatiotemporal trajectories, rather than spatial trajectories, are required in temporal-based clustering
- Assume that trajectory *i* contains *n* sampling points, trajectory *j* contains *n* sampling points, and *m* > *n*

$$\overline{d_{ij}} = \frac{1}{m}\left(\sum_{k=1}^{n} \|f_{i,k} - f_{j,k}\| + \sum_{k=1}^{m-n} \|f_{i,n+k} - f_{j,n}\|\right)$$

Semantic Indexing and Retrieval

Semantic Indexing

- Activity descriptor

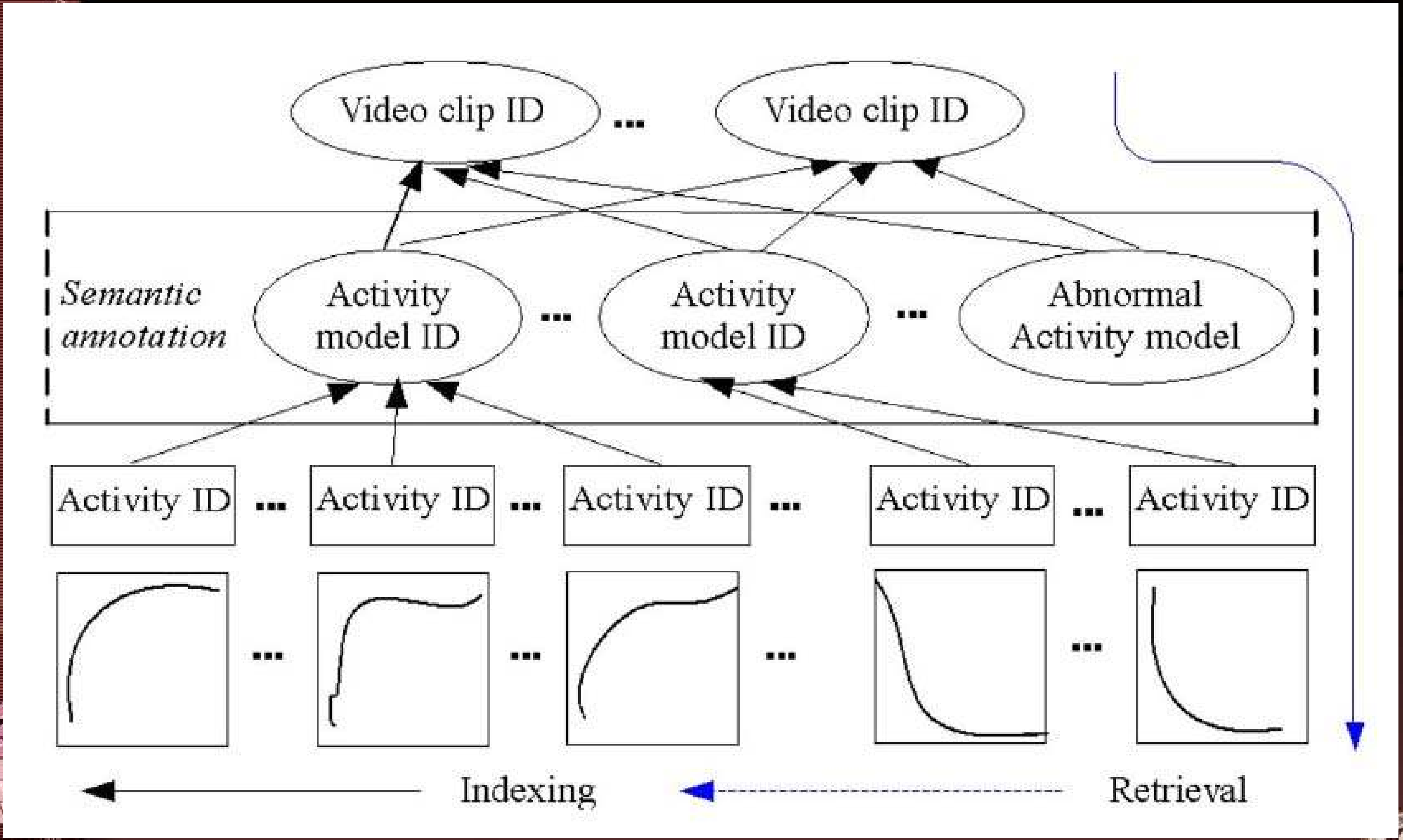
Components	Value
ACT_ID	ID of an activity
VIDEO_ID	ID of a video clip
Birth_Time	Frame number
Death_Time	Frame number
Spatio-Temporal Trajectory	$T_{ST} = (f_1, f_2, \ldots, f_i, \ldots, f_n)$ $f_i = (x_i, y_i, v_{x_i}, v_{y_i})$
Obj_Color	Object color (*R*, *G*, *B*)
Obj_Size	Object size (height, width)

Semantic Indexing

- Activity Model descriptor

Components	Value
AM_ID	ID of an activity model
ACT_List	A list of activities
Spatio-temporal template trajectory	$T_{ST} = (f_1, f_2, \ldots, f_i, \ldots, f_n)$ $f_i = (x_i, y_i, v_{x_i}, v_{y_i})$
Conceptual_Descriptions	Keywords {turn left; low speed; north ahead; traffic violation; …}

Hierarchical structure



Dynamic Adaption

- The distance from the spatiotemporal trajectory to the template trajectory
  - small enough: add to the activity list of this activity model
  - Otherwise: treated as a temporary abnormal activity and added to the abnormal activity model
- The spectral algorithm is used periodically to cluster the activities in the abnormal activity model
  - If there is a cluster which contains enough activities, the activities in this cluster are considered to be normal

Semantic Retrieval

- The object activities and their associated video clips are found, and the subvideo between birth and death frame is supplied to users for browsing
- Applicable Query Types
  - Query by keywords
  - Multiple object queries
  - Query by sketch

Query by Keywords

- Example: “a blue car ran from south to north at a high speed”
- Assume that an activity model contains a set of keywords A and there is a set of keywords in the query sentence(s)
- Degree of matching

$$\frac{|A \cap B|}{\sqrt{|A||B|}}.$$

Multiple object queries

- Two temporal restrictions are considered
  - Succession
  - Simultaneity
- the precision-recall cures are affected by the permutation order of retrieval results
  - BFS
  - DFS

Query by sketch

- trajectory drawn by a user
  - $A = (X_{A,1}, Y_{A,1}),(X_{A,2}, Y_{A,2}),\ldots,(X_{A,m}, Y_{A,m})$
- the spatial template trajectory in an activity model
  - $B = (X_{B,1}, Y_{B,1}),(X_{B,2}, Y_{B,2}),\ldots,(X_{B,n}, Y_{B,n})$
- Three step before calculate distance
  - Re-sampling
  - Scaling
  - Translation

Query by sketch

- Re-sampling (A_1)
  - point *i* in trajectory A_1 is prorated in the line segment

$$[(X_{A,\lfloor (m/n)\times i \rfloor}, Y_{A,\lfloor (m/n)\times i \rfloor}) - (X_{A,\lfloor (m/n)\times i \rfloor+1}, Y_{A,\lfloor (m/n)\times i \rfloor+1})]$$

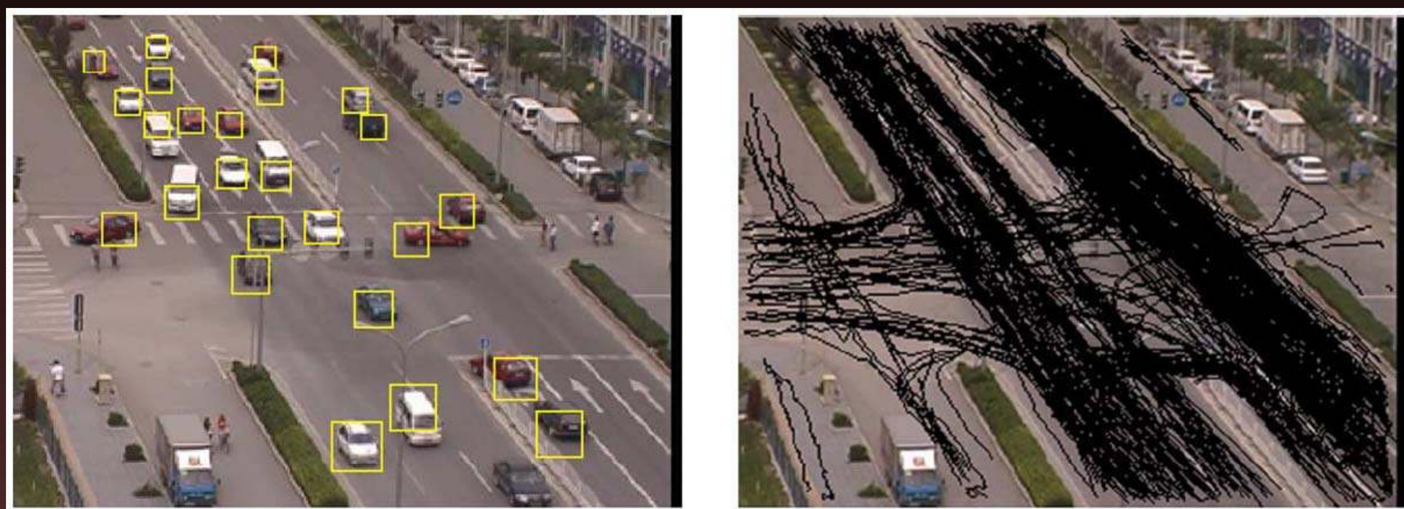
- Scaling (A_2)
  - Trajectory A_1 is scaled by L_B/L_{A_1}, to form trajectory
- Translation (A_3)
  - Trajectory A_2 is translated to match B
- Distance

$$f(\Delta x, \Delta y) = \sum_{i=1}^{n} \Big((x_{A_2,i} + \Delta x - x_{B,i})^2 + (y_{A_2,i} + \Delta y - y_{B,i})^2 \Big)$$

Experimental Results

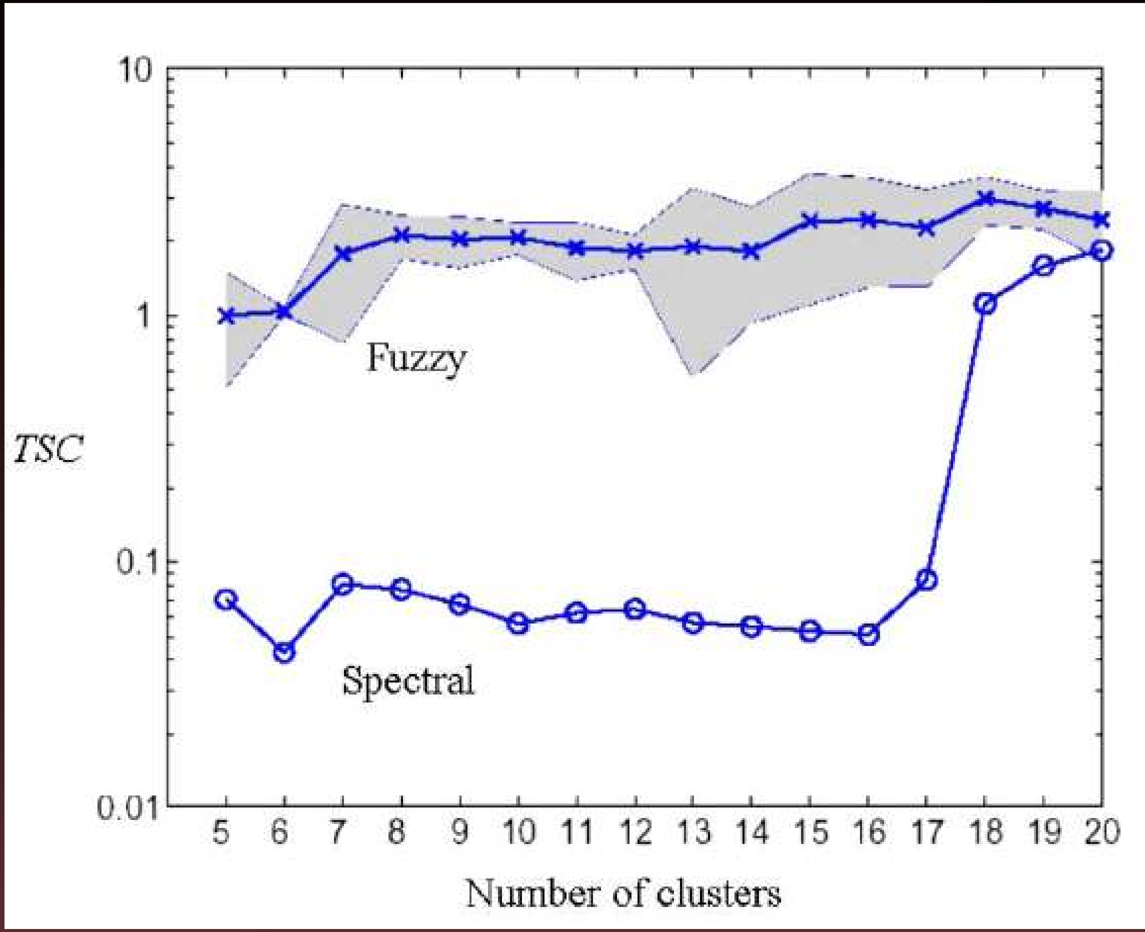
Tracking

- 320 * 240 RGB image
- 1184 /1216 = 97.4%
- 5–10 frame/s on P4-1.8-GHz

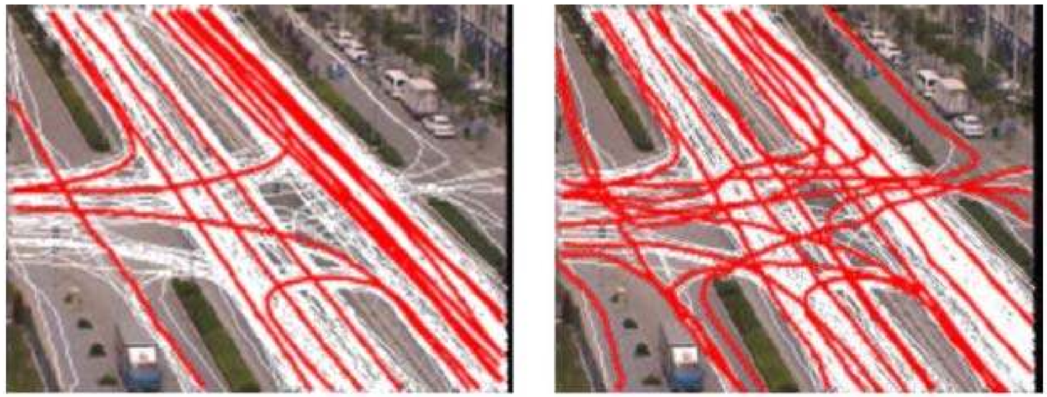


Learning of Activity Models

- first layer of spatial-based clustering

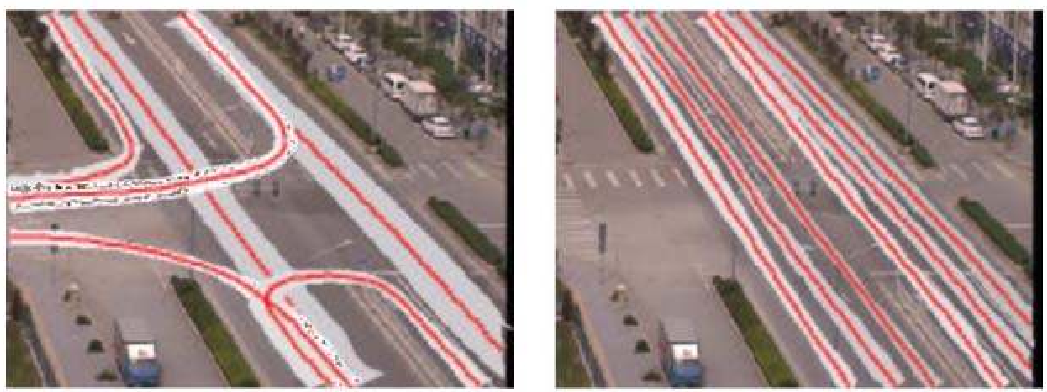


Learning of Activity Models



(a) (b)

Fig. 7. Results of second layer of spatial-based clustering: (a) fuzzy c-means; (b) spectral clustering.



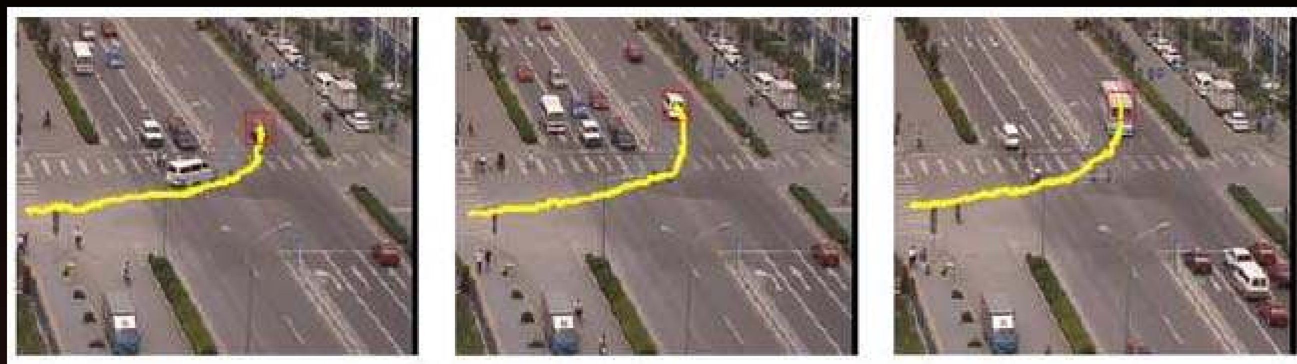
(a) (b)

Fig. 8. Contrast between first and second layers of spatial-based spectral clustering: (a) first layer; (b) second layer.

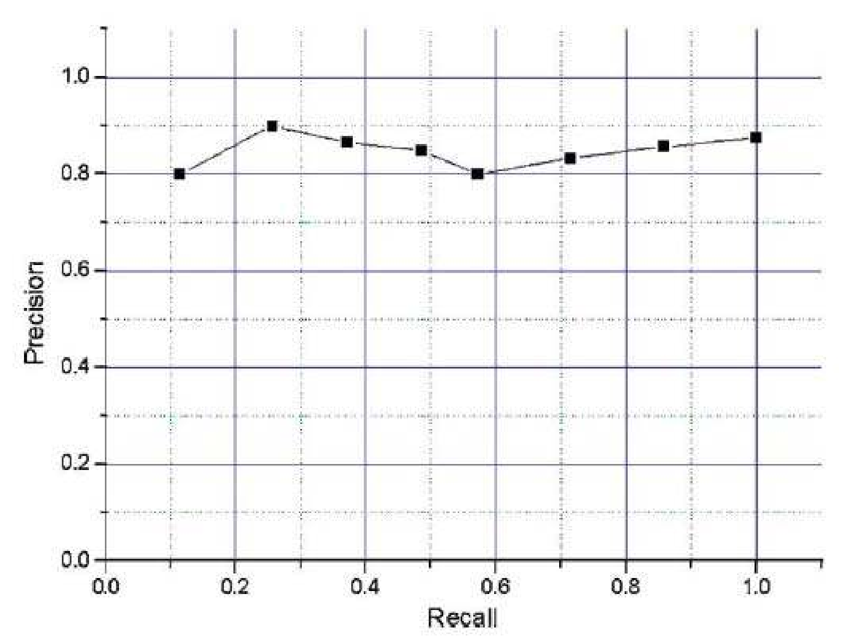
Learning of Activity Models

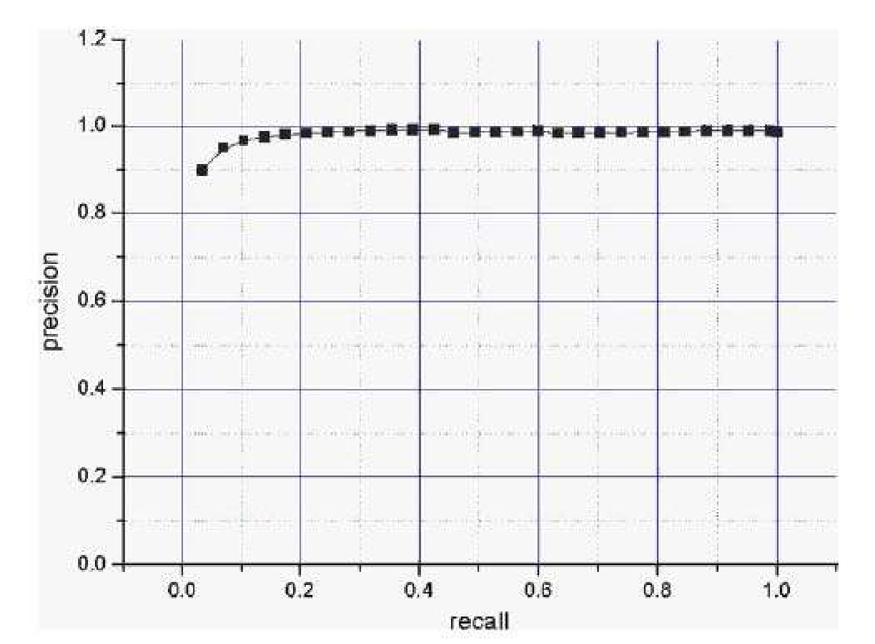
- 76 activity models are finally learned

Keywords-Based Retrieval



Keywords-Based Retrieval

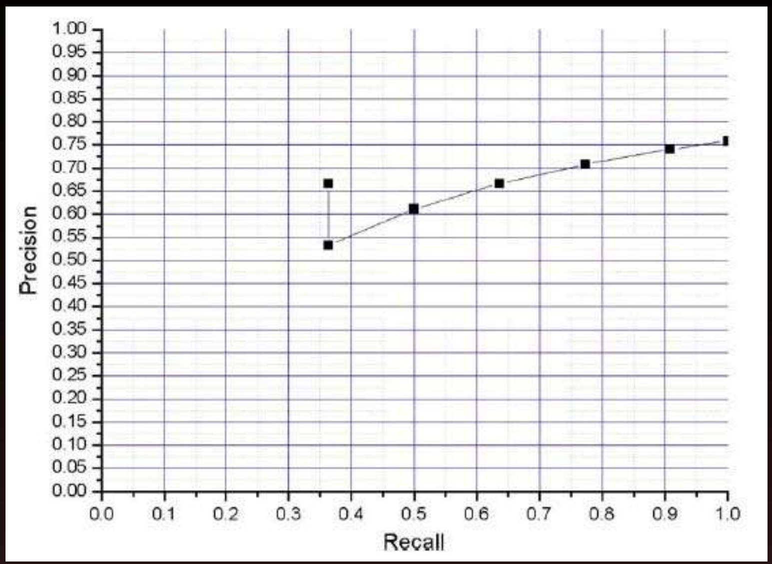


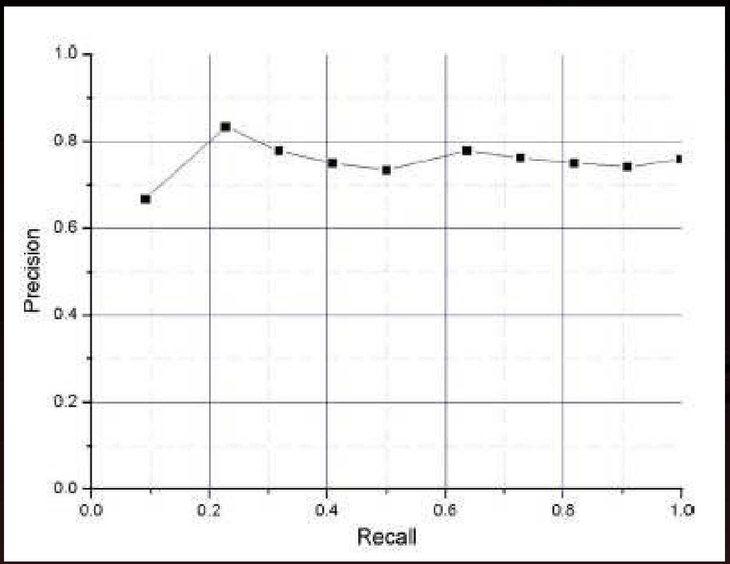


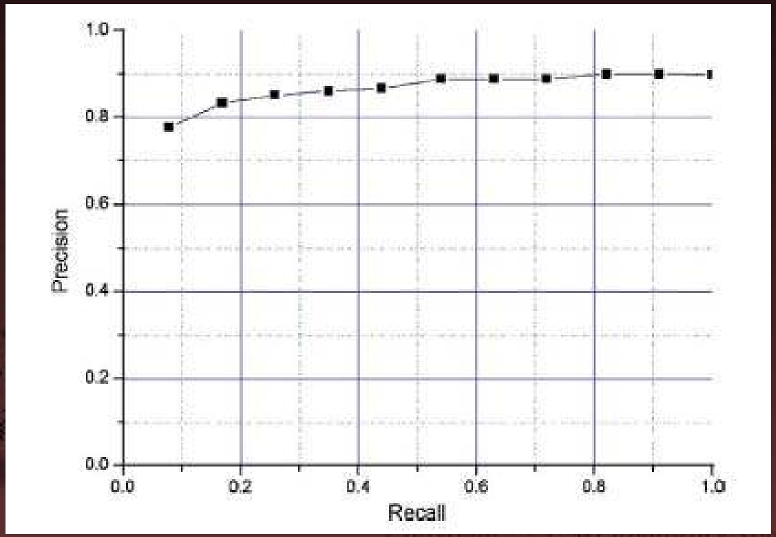
" a red car turned left "

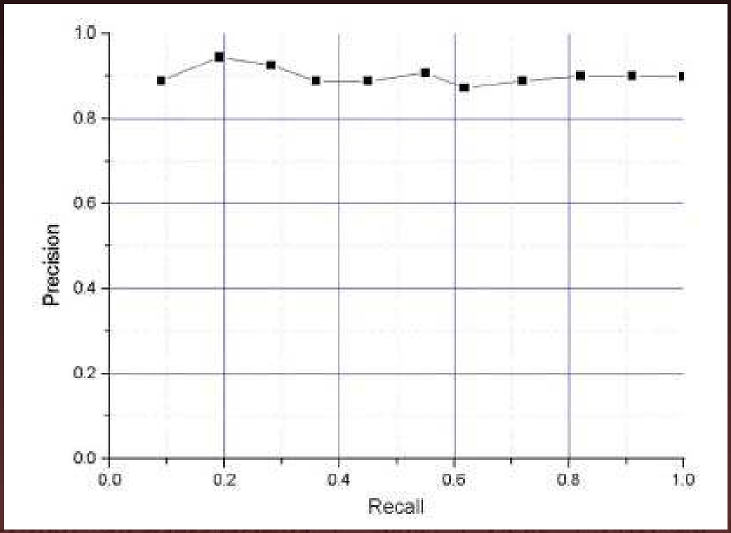
" a white car ran from south to north by the right lane "

Multiple Object Query

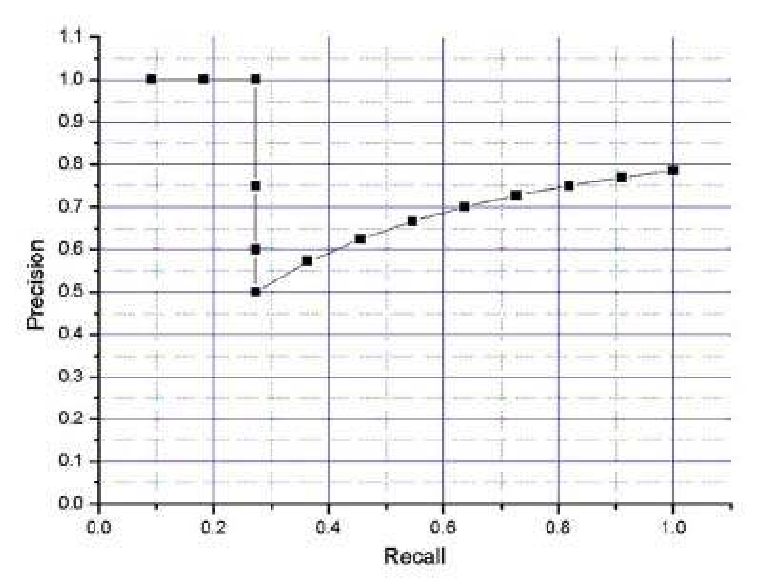


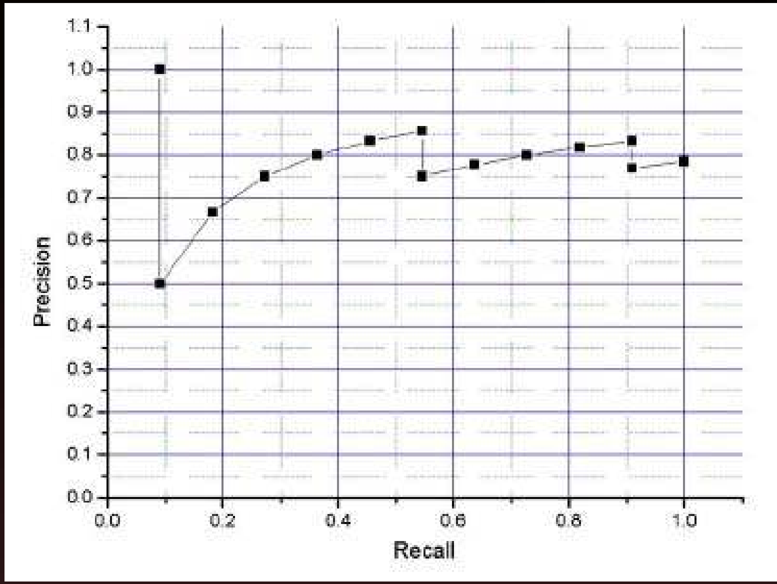


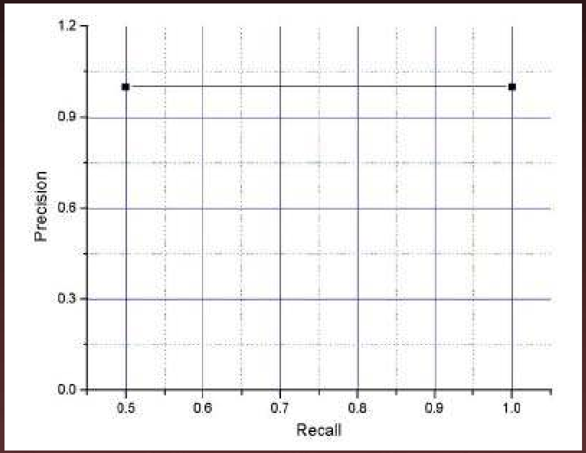




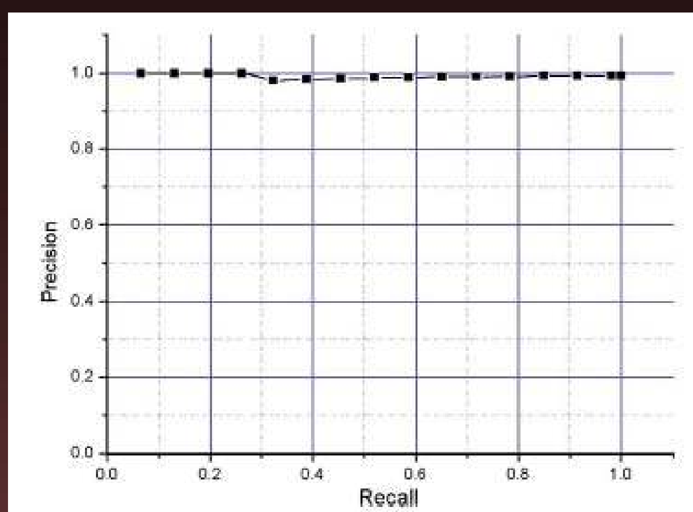
Multiple Object Query

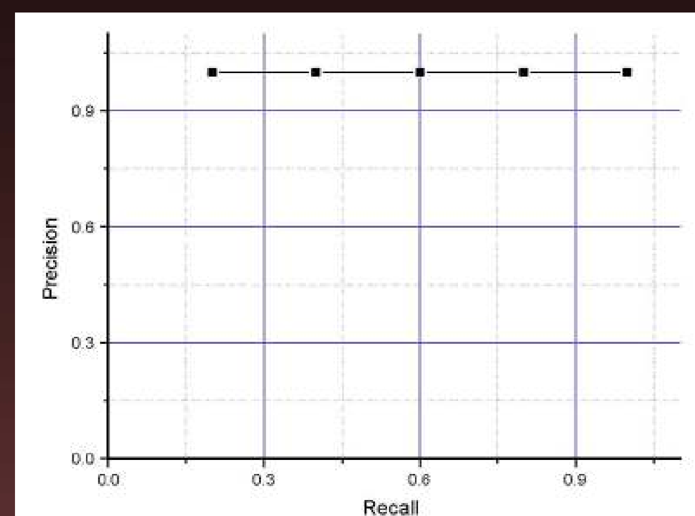






Sketch-Based Query





Conclusion

- A clustering-based tracking algorithm is used to obtain trajectories
- With semantic indexing, our retrieval framework provides a query interface at the semantic level
- the workload of manual annotation is greatly reduced
- The framework has been experimentally tested in a crowded traffic scene, with good results.

cmLAB NATIONAL TAIWAN UNIVERSITY

Thank You