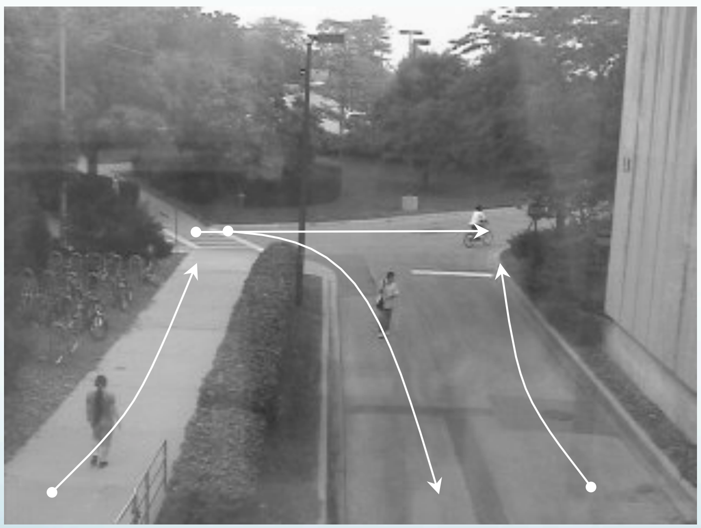
Abnormal Event Detection Based on Trajectory Clustering

Fan Jiang, Ying Wu, Aggelos K. Katsaggelos

Dept EECS, Northwestern University

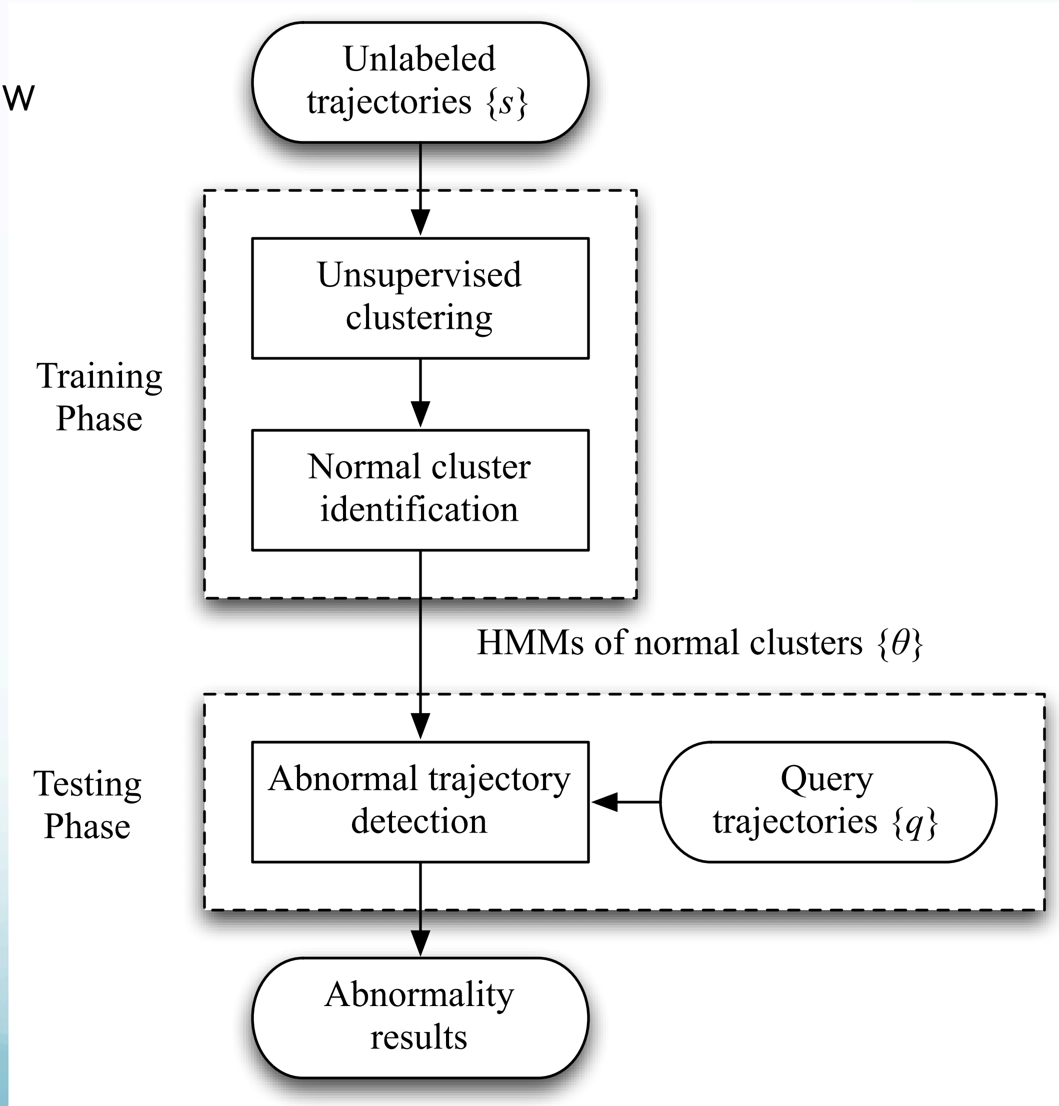
Introduction

- Suspicious / abnormal events in surveillance scene
- Trajectory-based representation




Clustering-Based Approach

System Overview



Dissimilarity Measure

- Bayesian information criterion (BIC)

$$BIC = -\log L + \frac{1}{2} K \log N,$$

- For N trajectories:

$$BIC(i, j, \cdots) = -\sum_{n=1}^{N} \log L_n + \frac{1}{2} N K_0 \log N,$$

- Merge i and j

$$BIC(ij, \cdots) = -\sum_{\substack{n=1 \\ n \neq i,j}}^{N} \log L_n - \log L_{ij} + \frac{1}{2}(N-1) K_0 \log N,$$

- Information-based dissimilarity

$$\begin{aligned} d(i, j) &= BIC(ij, \cdots) - BIC(i, j, \cdots) \\ &= \log L_i + \log L_j - \log L_{ij} - \frac{1}{2} K_0 \log N. \quad (5) \end{aligned}$$

Trajectory Clustering

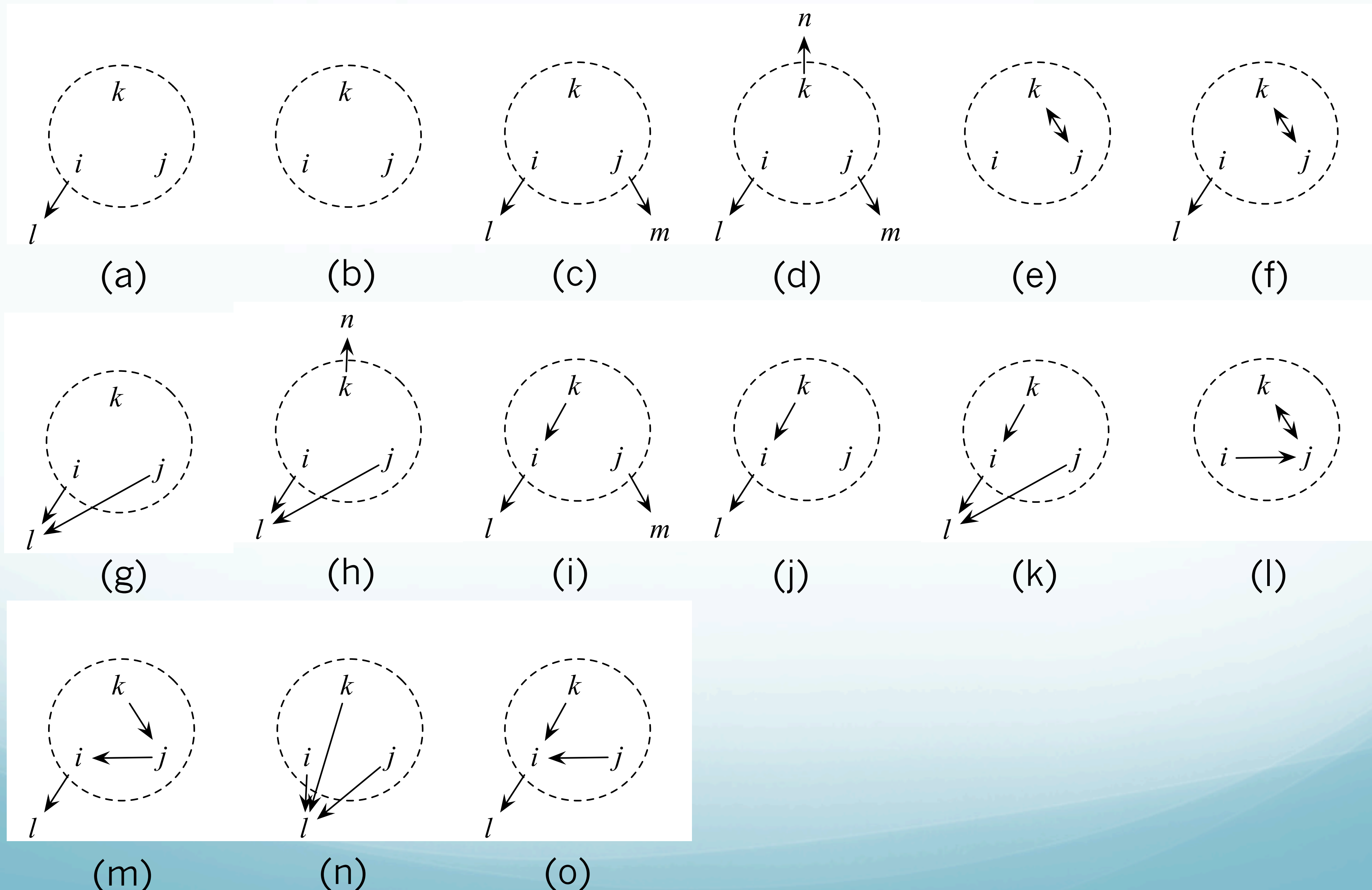
- Dynamic Hierarchical Clustering (DHC)

0) *Initialization:* each trajectory in the dataset forms a group and is fitted with an HMM. There are N groups with N HMMs;
1) *Dissimilarity Measurement:* calculate $d(i, j)$ of any two groups i and j in the dataset by (5);
2) *Merging:* the two groups $\hat{i}$ and $\hat{j}$ with smallest $d(\hat{i}, \hat{j})(d(\hat{i}, \hat{j}) < 0)$ are merged; if no $d(i, j) < 0$, the clustering terminates;
3) *Reclassification:* a new HMM $\theta_{\hat{i}\hat{j}}$ is trained, replacing $\theta_{\hat{i}}$ and $\theta_{\hat{j}}$; then based on the $(N-1)$ HMMs, all trajectories are reclassified into $(N-1)$ groups by the maximum likelihood (ML) criterion;
4) *Retraining:* $(N-1)$ HMMs are retrained based on the updated $(N-1)$ data groups respectively;
5) *Update:* $N = N - 1$; go back to step 1).

2-Depth Search Strategy

- At each merging step, consider all pairs and all triplets:
  - Make two merges that decrease BIC the most
- However, need to calculate all trajectory pairs (d(i,j)), all trajectory triplets (d(i,j,k))
  - Huge computation!

15 Categories of trajectory groups according to nearest neighbors



Branch Pruning

- Given pairwise meausement, the triple measurement can be possibly avoided
- Example: for category (f), merging of i,j,k together can be rejected if
  - BIC (ij, k, l, ···) > BIC (il, jk, ···)
  - Sufficient condition:

$$d(i,j) - d(j,k) + d(i,k) - d(i,l) > \frac{1}{2} K_0 \log N.$$

Proof of Sufficient Condition for (f)

$$d(i,j) - d(j,k) + d(i,k) - 2d(i,l) > \frac{1}{2}K_0 \log N$$

$$\Longrightarrow (\log L_i + \log L_j - \log L_{ij} - \frac{1}{2}K_0 \log N)$$
$$- (\log L_j + \log L_k - \log L_{jk} - \frac{1}{2}K_0 \log N)$$
$$+ (\log L_i + \log L_k - \log L_{ik} - \frac{1}{2}K_0 \log N)$$
$$- 2(\log L_i + \log L_l - \log L_{il} - \frac{1}{2}K_0 \log N)$$
$$> \frac{1}{2}K_0 \log N$$

$$\Longrightarrow -\log L_{ij} - \log L_{jk} - \log L_{ik}$$
$$+ 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0$$

$$\Longrightarrow -\log L_{ijk}^{ij} - \log L_{ijk}^{jk} - \log L_{ijk}^{ik}$$
$$+ 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0$$

$$\Longrightarrow -\log(L_{ijk})^2 + 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0$$

$$\Longrightarrow -\log L_{ijk} - \log L_l + \log L_{il} + \log L_{jk} > 0$$

$$\Longrightarrow BIC(ijk, l, \cdots) - BIC(il, jk, \cdots) > 0$$

$$\Longrightarrow BIC(ijk, l, \cdots) > BIC(il, jk, \cdots)$$

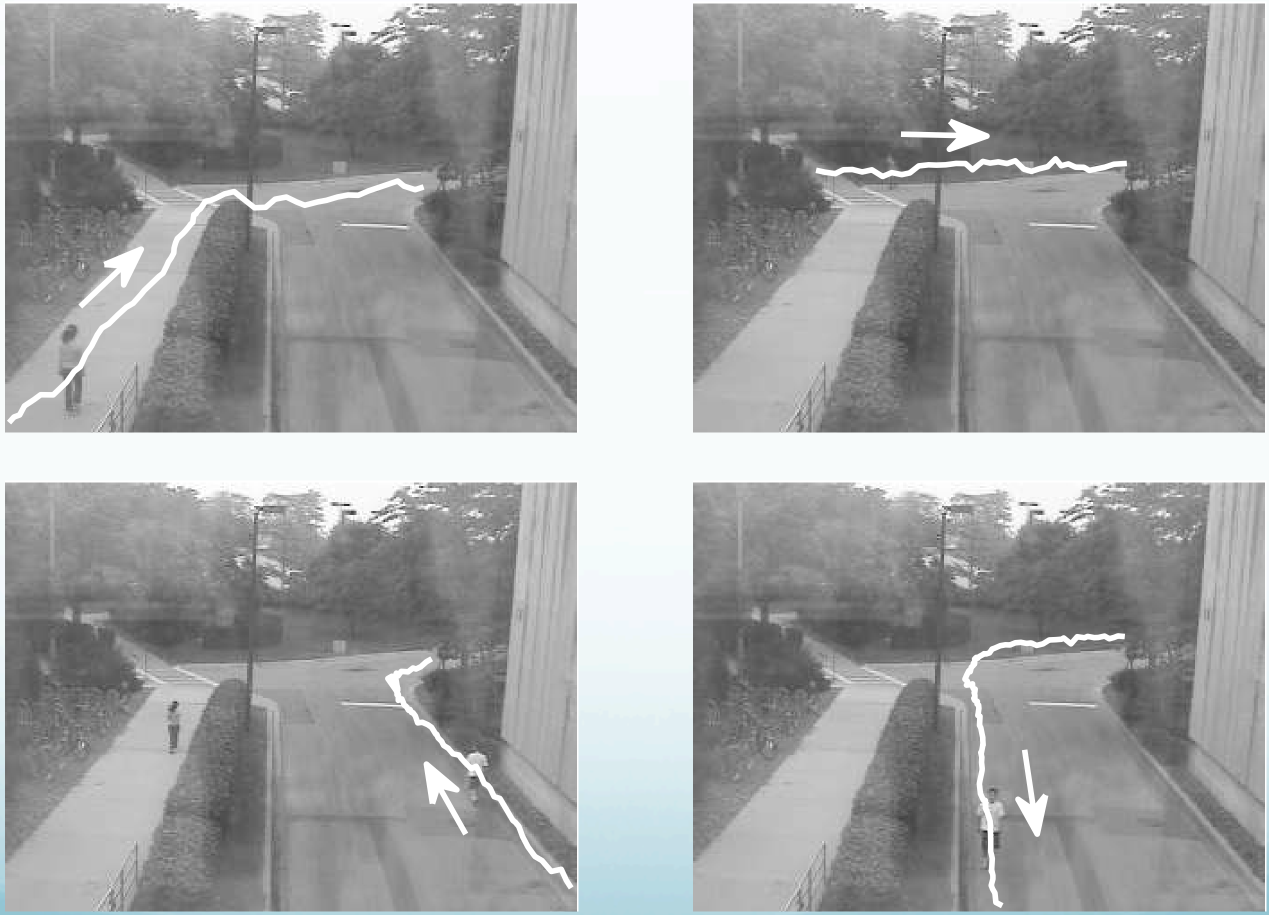
Category	Exclusions	Sufficient conditions
(a)	$BIC(ijk,\cdots) > BIC(i,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k) > \frac{1}{2}K_0 \log N$
(b)	$BIC(ijk,l,\cdots) > BIC(il,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
(c)	$BIC(ijk,l,m,\cdots) > BIC(il,jm,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(j,m) > \frac{1}{2}K_0 \log N$
(d)	$BIC(ijk,l,m,n,\cdots) > BIC(il,jm,kn,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(j,m)-2d(k,n) > \frac{1}{2}K_0 \log N$
(e)	$BIC(ijk,\cdots) > BIC(i,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k) > \frac{1}{2}K_0 \log N$
(f)	$BIC(ijk,l,\cdots) > BIC(il,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
(g)	$BIC(ijk,l,\cdots) > BIC(il,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,l,\cdots) > BIC(i,jl,k,\cdots)$	or $d(i,j)+d(j,k)+d(i,k)-2d(j,l) > \frac{1}{2}K_0 \log N$
(h)	$BIC(ijk,l,n,\cdots) > BIC(il,j,kn,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(k,n) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,l,n,\cdots) > BIC(i,jl,kn,\cdots)$	or $d(i,j)+d(j,k)+d(i,k)-2d(j,l)-2d(k,n) > \frac{1}{2}K_0 \log N$
(i)	$BIC(ijk,l,\cdots) > BIC(il,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,\cdots) > BIC(ik,j,\cdots)$	or $d(i,j)+d(j,k)-d(i,k) > \frac{1}{2}K_0 \log N$
(j)	$BIC(ijk,l,m,\cdots) > BIC(il,jm\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(j,m) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,m,\cdots) > BIC(ik,jm,\cdots)$	or $d(i,j)+d(j,k)-d(i,k)-2d(j,m) > \frac{1}{2}K_0 \log N$
(k)	$BIC(ijk,l,\cdots) > BIC(ik,jl,\cdots)$	$d(i,j)+d(j,k)-d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
(l)	$BIC(ijk,\cdots) > BIC(i,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k) > \frac{1}{2}K_0 \log N$
(m)	$BIC(ijk,l,\cdots) > BIC(il,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$

2-Depth DHC

0) *Initialization:* each trajectory in the dataset forms a group and is fitted with an HMM. There are N groups with N HMMs;
1) *Pairwise Group Measurement:* calculate $d(i, j)$ of any two groups i and j in the dataset by (5);
2) *Triple Group Measurement:* calculate only $d(i, j, k)$ of those three groups i, j, k that do not satisfy the exclusions in Table I;
3) *Merging:* merge two group pairs or three groups together, based on which results in the greatest decrease of BIC. The clustering process terminates if no merge can decrease BIC;
4) *Reclassification:* new HMMs are trained for the merged groups; then based on the $(N-2)$ HMMs, all trajectories are reclassified into $(N-2)$ groups by the maximum likelihood (ML) criterion;
5) *Retraining:* $(N-2)$ HMMs are retrained based on the updated $(N-2)$ data groups respectively;
6) *Update:* $N = N - 2$; go back to step 1).

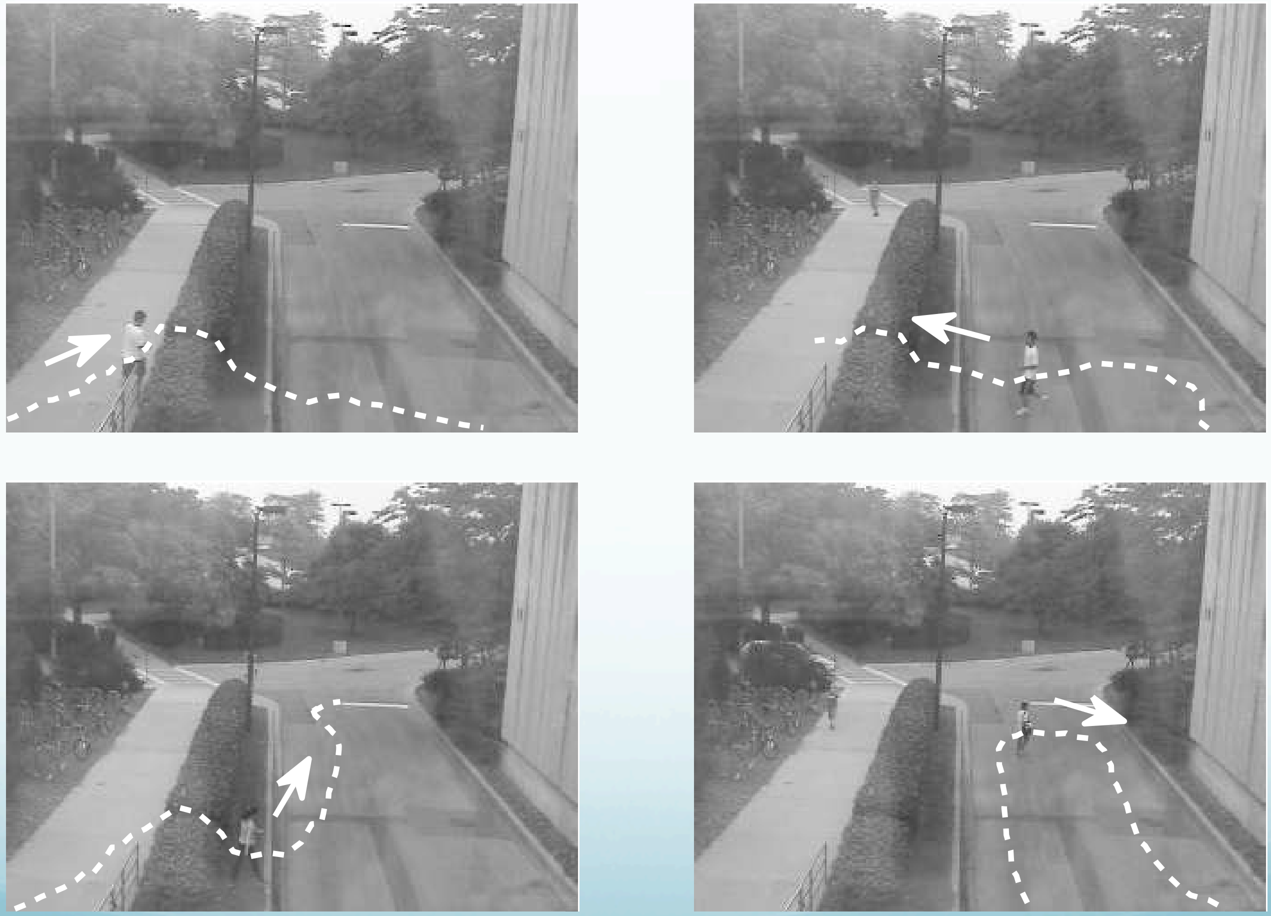
Normal Trajectories

- Examples:



Abnormal Trajectories

- Examples:



Comparison Results

- Compare 4 methods:
  - SC: spectral clustering (using CLR-based dissimilarity measure)
  - HC1: classic hierarchical clustering (1-depth, no dynamics)
  - DHC1: dynamic hierarchical clustering (1-depth)
  - DHC2: dynamic hierarchical clustering (2-depth)

Detection Error Rates