Detecting and Tracking Moving Objects for Video Surveillance

Isaac Cohen and Gerard Medioni
University of Southern California

Their application sounds familiar.

- Video surveillance
- Sensors with pan-tilt and zoom
- Sensors mounted on moving airborne platforms
- Requirement for detection and tracking of moving objects and the relationship of their trajectories
- Requirement for high-level description of the whole video sequence

Approach

- First find optical flow and perform motion compensation to stabilize.
- Find large numbers of moving regions.
- Use the residual flow field and its normal component to detect errors.
- Define a attributed graph whose nodes are detected regions and edges are possible matches between two regions detected in two different frames.
- Use the graph as a dynamic template for tracking moving objects.

Their Idea

- Detection after stabilization doesn’t work well.

- So integrate the detection into the stabilization algorithm by locating regions of the image where a residual motion occurs using the normal component of the optical flow field.

What is “Normal Flow”?

- The optical flow equation constrains the image velocity in the direction of the local image gradient, but not the tangential velocity.
- Normal flow corresponds to the image velocity along the image gradient.
- It is computed from both image gradients and temporal gradients of the stabilized sequence.
- The amplitude is large near moving regions.
- The amplitude is near zero near stationary regions.

Computation of Normal Flow

Let τ_{ij} denote the warping of image to reference frame j.

$$\mathcal{T}_{ij} = \prod_{k=i,..,j+1} \mathcal{T}_{k,k-1}$$

The stabilized image sequence is defined by $I_i(\tau_{ij})$.

Given reference image I_0 and target I_1, image stabilization consists of registering the two images and computing the geometric transform τ that warps I_1 so it aligns with I_0.

Parameter Estimation of the Geometric Transform

- Minimize the least squares equation

$$E = \sum_i \{I_0(x_i, y_i) - I_1(\mathcal{T}(x_i, y_i))\}^2$$

- Detect and remove outliers through an iterative process.

Definition of Normal Flow

$$w_{\perp} = -\frac{\left(I_{i+1}(\mathcal{T}_{i+1,j}) - I_i(\mathcal{T}_{i,j})\right)}{\|\nabla \mathcal{T}_{ij} \nabla I_i(\mathcal{T}_{ij})\|} \cdot \frac{\nabla \mathcal{T}_{ij} \nabla I_i(\mathcal{T}_{ij})}{\|\nabla \mathcal{T}_{ij} \nabla I_i(\mathcal{T}_{ij})\|}$$

- The warping function is integrated into the formula so that the image gradients are computed on the original image grids and not the warped ones.

- This simplifies the computation and allows for a more accurate estimation of the residual normal flow.

Finding Moving Objects

- Given a pair of image frames, find the moving objects by thresholding the normal flow.

- Does this overcome problems of other approaches?

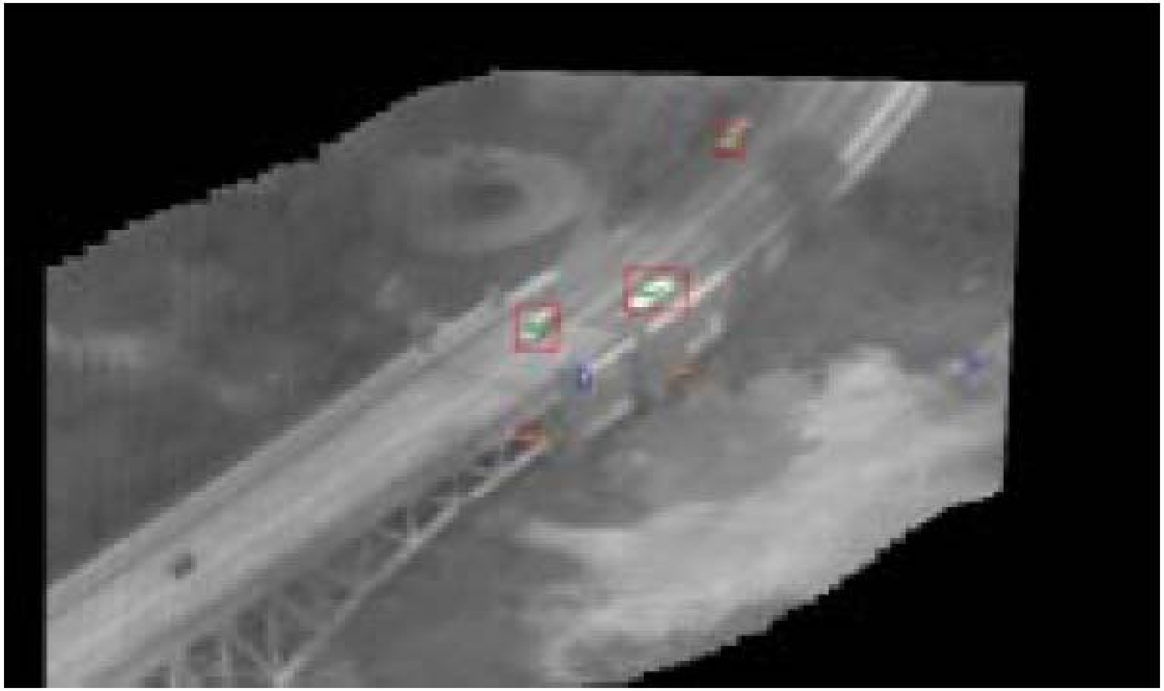
Graph Representation of Moving Objects

- Nodes are the moving regions.
- Edges are the relationships between 2 moving regions detected in 2 separate frames.
- Each new frame generates a set of regions corresponding to the moving objects.
- We want to know which moving objects in the new frame are the same as those in the old.

Why is this Difficult?

- Little information about the objects is known.
- Objects tend to be of small size in aerial imagery.
- Large changes in object size are possible.

Sample Detection



Detection Videos

- seq1.avi
- seq1_det.avi

- seq6.avi
- seq6_det.avi

Example Graph: What's going on?

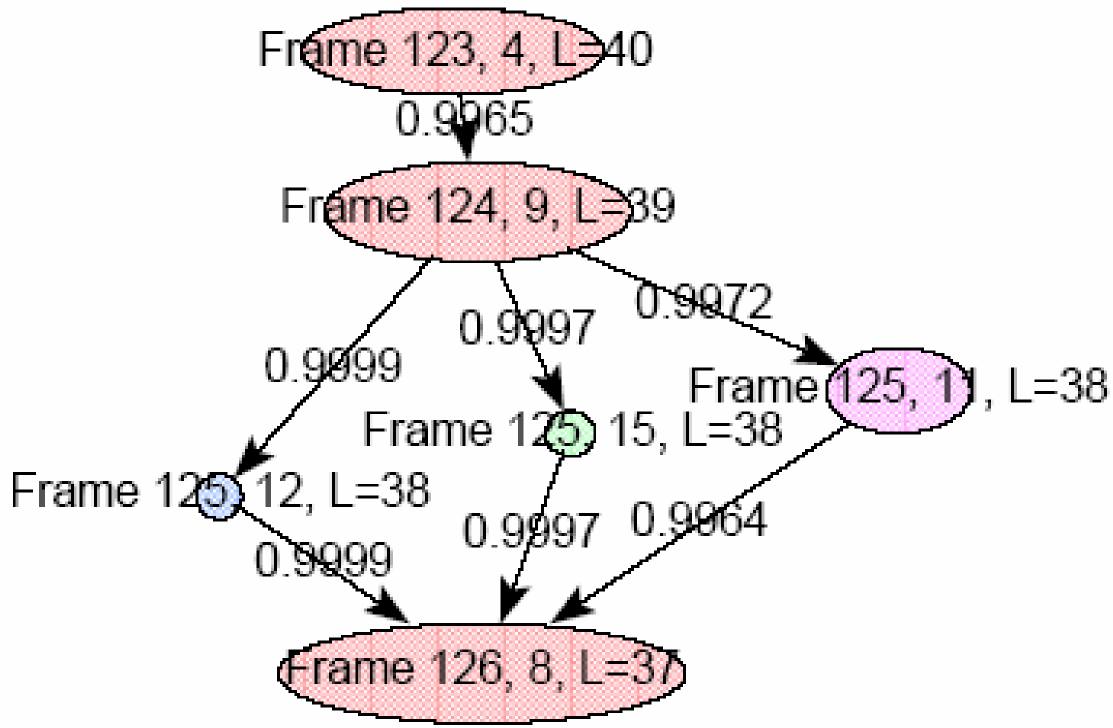


Figure 2: *Detected regions and associated graph.*

Attributes of a Node

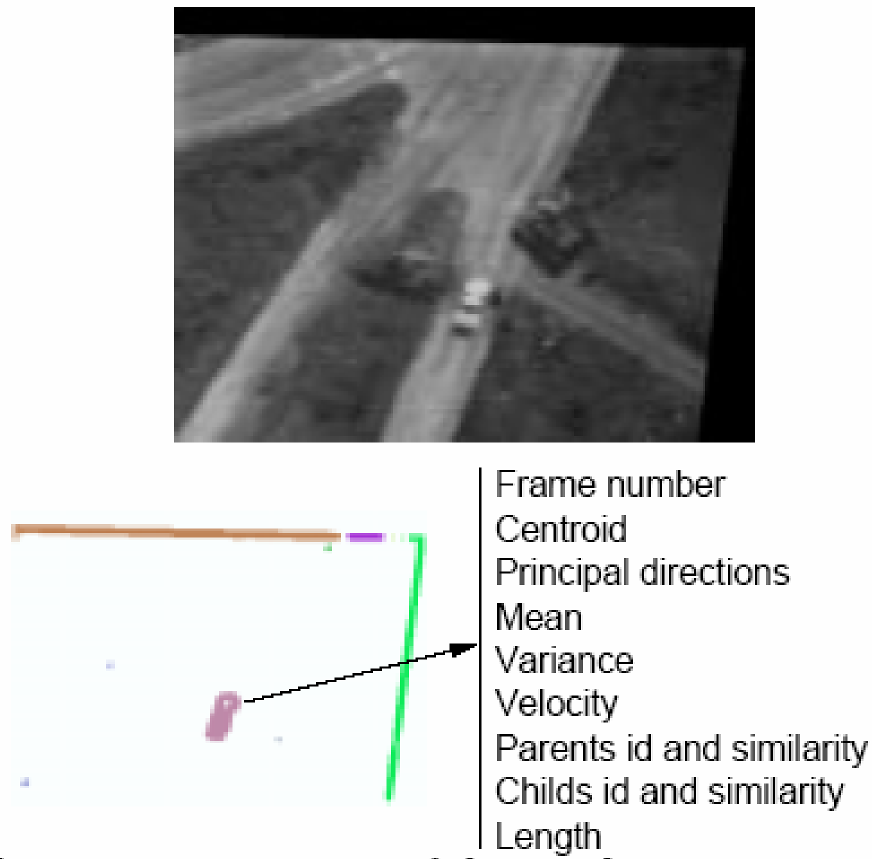
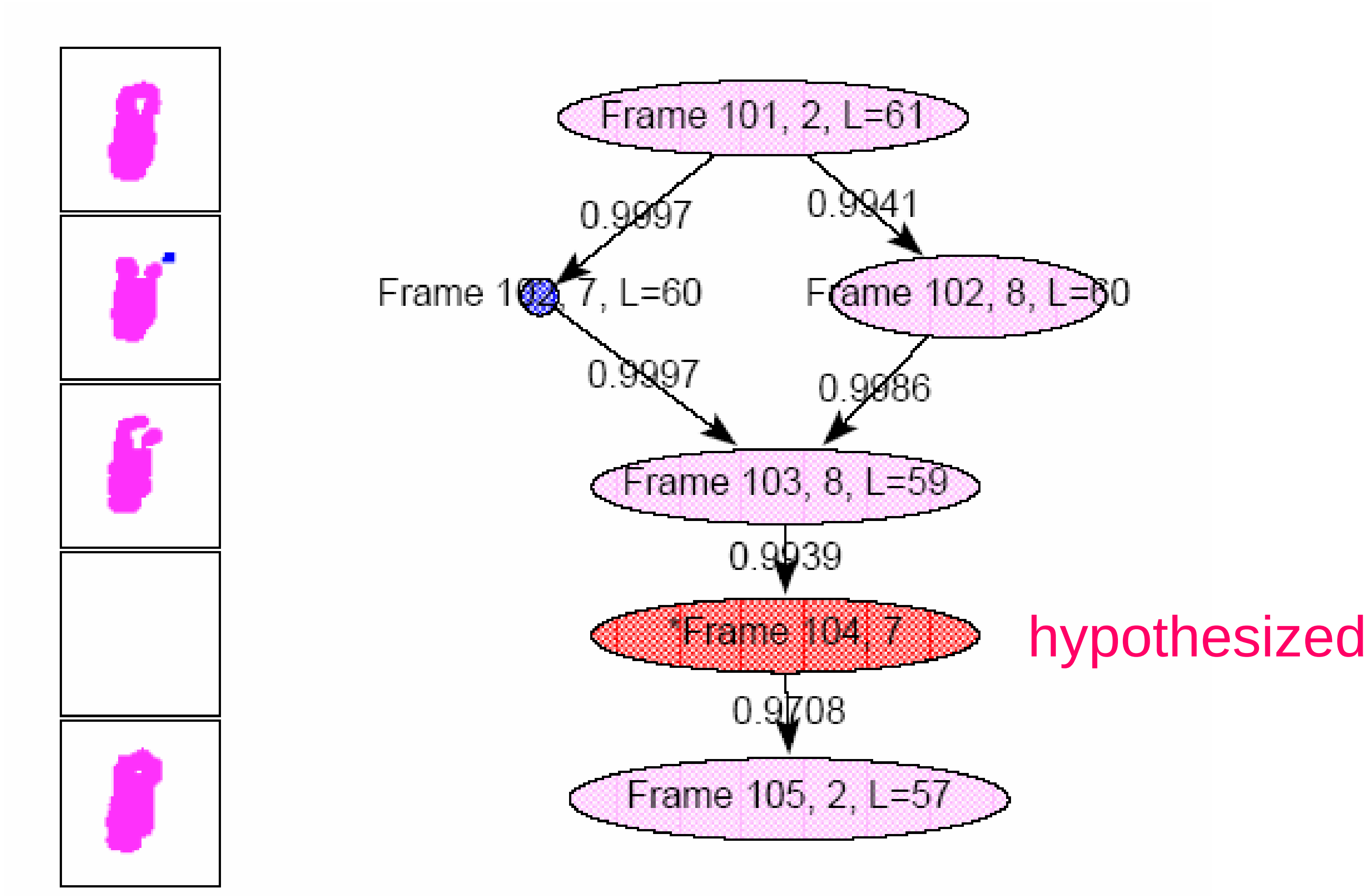


Figure 3: *Description of the attributes associated to each node of the graph. Each color represents a moving region.*

Keeping Track of Moving Regions

- Among the detected regions, some small ones should be merged into a larger one.
- They cluster the detected regions in the graph, instead of using single images.
- Use a median shape template to keep track of the different moving regions.

Moving objects that don't appear in some frames can be hypothesized.



Extraction of Object Trajectories

- The graph representation changes as new frames are acquired and processed.
- The goal is to find the full trajectory of each moving object.
- But we don’t know where each one starts or where it ends.
- So we have to consider each node with no predecessor a possible start and each with no successor a possible end.

Optimal Path

- Assign each edge of the graph a cost, which is the similarity between the connected nodes.

$$i \xrightarrow{c_{ij} = C_{ij} / (1 + d_{ij}^2)} j$$

C_{ij} is the gray-level and shape correlation between i and j.
d_{ij} is the distance between their centroids.

Optimal Path

- This formulation does not lead to the optimal solution.
- Instead they define the length l_j of node j as the maximal length of the path starting at that node.
- Then the modified cost function $C_{ij}=l_j c_{ij}$ is used to find the optimal paths from each node without predecessor, using a greedy search method.

Quantitative Evaluation

- TP = true positives of moving objects
- FP = false positives of moving objects
- FN = false negatives (not detected)

Metrics:

DR = TP / (TP + FN)	Detection Ratio
FAR = FP / (TP + FP)	False Alarm Ratio

Evaluation Results on 5 Shots

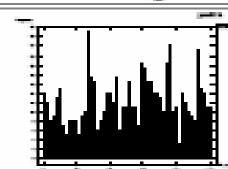
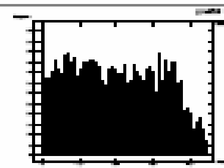
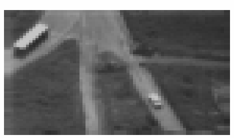
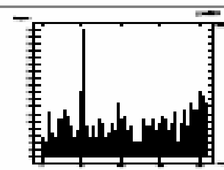
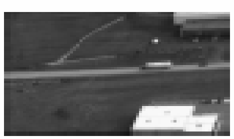
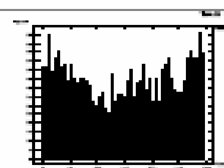
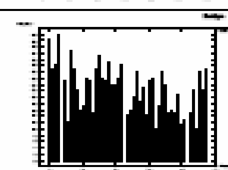
video stream	Moving Objects	Detection			Tracking		Metrics	
		detected regions	mean	σ	Regions	Paths	DR	FAR
	1		9	5	1	1	1.	0.
	2		29	15	3	3	1.	0.2
	4		6	3	4	5	1.	0.
	2		34	11	10	5	1.	0.8
	7		22	8	15	12	1.	0.53

Table 1: *Quantitative analysis of the detection/tracking modules*

Tracking Demos

- seq1_mos_track.avi

- seq6_track.avi

Questions/Comments

- Have they solved our problem?

- Has anyone done better? No one seems to use the metrics. But a 2005 paper by Nicolescu and Medioni compares results of 4 methods on fake sequences (and beats them, of course)

- Another recent paper by Xiao and Shah says they compared their results to those of Ke and Kanade, Wang and Adelson, and Ayer and Sawhney, but no numbers are given.

- Can we beat these people?