Unsupervised Learning-Based Spatio-Temporal Vehicle Tracking and Indexing for Transportation Multimedia Database Systems

Prepared by:
Chengcui Zhang
School of Computer Science
Florida International University

Outline

- Introduction
- Learning-Based Object Tracking and Indexing for Traffic Video Sequences
  - The SPCPE segmentation algorithm
  - Object Tracking
  - Self-adaptive Background Subtraction
  - Multimedia Augmented Transition Networks (MATNs) and Multimedia Input Strings
- Experimental Analysis
- Conclusions

Introduction

- Intelligent Transportation Systems (ITS)
- ITS technologies include
  - Advanced Traveler Information Systems (ATIS):
    - Providing network-wide routing information to users
  - Advanced Traffic Management Systems (ATMS)
    - Using advanced sensor systems for on-line surveillance and detailed information gathering on traffic conditions

Introduction (cont.)

- It is very important for the traffic control center to collect, analyze, and store large-scale multimedia traffic flow data for real-time usage
- For this purpose, a learning-based object tracking and indexing framework for traffic video sequences is proposed
  - To enable the vehicle identification process for object tracking and indexing for transportation multimedia database systems

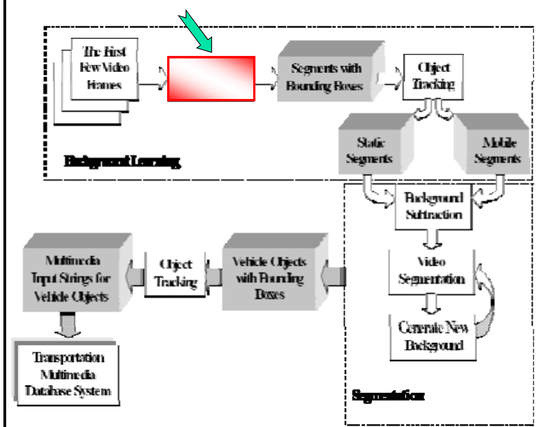
Introduction (cont.)

- To identify and track the temporal or relative spatial positions of vehicle objects in video sequences
  - needs to have object-based representation of video data
    - an unsupervised image segmentation method
  - needs an effective background learning algorithm
  - incorporates the background learning algorithm with the image segmentation method
    - the initial inaccurate background information can be refined and adjusted in a self-adaptive way throughout the segmentation
  - achieves traffic video indexing by representing and modeling the temporal and spatial relations of vehicle objects
    - multimedia augmented transition network (MATN) model
    - multimedia input strings (MIS)

Introduction (cont.)

- Possible applications of this framework:
  - intersection traffic monitoring
  - incident detection
  - vehicle queue length determination at signaled street intersections
  - intelligent traffic signal control

The basic workflow of the proposed framework

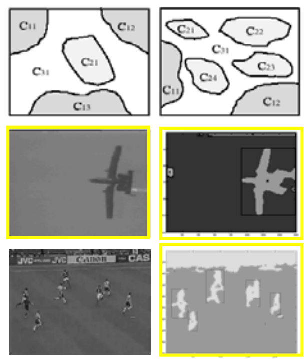


Unsupervised Segmentation Algorithm (SPCPE)

- Simultaneous partition and class parameter estimation (SPCPE) algorithm
  - Images and video frames segmentation
  - Identifying segments (objects) in an image or a video frame

SPCPE Algorithm -- Background

- Classes and Segments
  - Class: All regions with the similar statistical description. Eg. Houses, roads, parks, etc.
  - Segment: An instance of the class in the image



$C_i = \cup_j C_{ij}$

Pixels in segment C_{ij} belong to class C_i

2 classes, 2 segments
Airplane and Sky

2 classes, 7 segments
{Players, ball, sign boards}
And {ground}

SPCPE Algorithm

- Simultaneous Partition and Class Parameter Estimation (SPCPE) Algorithm
  - The partition and the class parameters ⇨ random variables
  - Partition variable ⇨ $\boldsymbol{c} = \{\boldsymbol{c_1}, \boldsymbol{c_2}\}$ (with 2 classes)
  - Classes ⇨ parameterized by $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$
    - $Y = \{y_{ij};\ 1 \le i \le Nr \text{ and } 1 \le j \le Nc\}$
    - $y_{ij} = a_{k0} + a_{k1} i + a_{k2} j + a_{k3} ij,\ \forall (i, j)\ y_{ij} \in \boldsymbol{c_k},\ k=1, 2$
    - $\boldsymbol{\theta_k} = (a_{k0}, \ldots, a_{k3})^T,\ k=1, 2$
  - Frame partition ⇨ a joint estimation of the partition and class parameters
  - MAP estimates of $\boldsymbol{c} = \{\boldsymbol{c_1}, \boldsymbol{c_2}\}$ and $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$

    $$(\hat{c}, \hat{\theta}) = Arg \max_{(c,\theta)} P(c, \theta \mid Y) = Arg \max_{(c,\theta)} P(Y \mid c, \theta) P(c, \theta)$$

    $$(\hat{c}, \hat{\theta}) = Arg \min_{(c,\theta)} J(c_1, c_2, \theta_1, \theta_2)$$
  - Simplifying joint estimation

    $$J(c_1, c_2, \theta_1, \theta_2) = \sum_{y_{ij} \in c_1} -\ln p_1(y_{ij}; \theta_1) + \sum_{y_{ij} \in c_2} -\ln p_2(y_{ij}; \theta_2)$$

SPCPE Algorithm (Cont.)

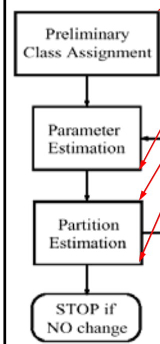
- Steps:

1: Starting with an arbitrary partition and computing class parameters

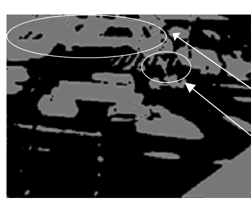
2: Estimating a new partition by class parameters and data

3: Iteratively refining partition and class parameters until no change

4: Obtaining the bounding box of each semantic object

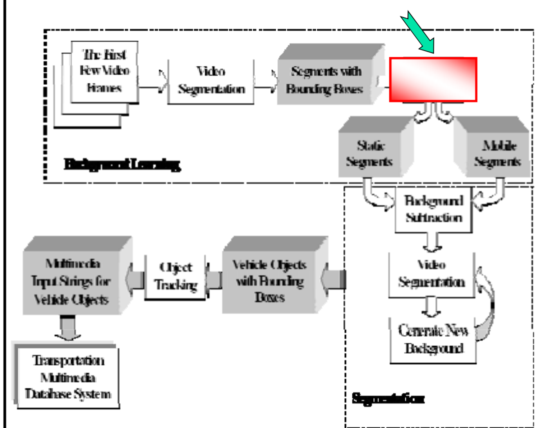


Problem: Objects Segmented Based Solely on Pixel Intensity



- SPCPE sensitive to image complexity
  - shadows grouped into vehicle class
  - fail to detect black car similar in color to the road

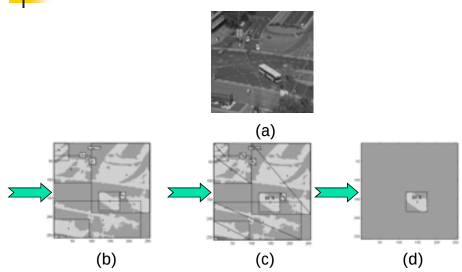
The basic workflow of the proposed framework



Object Tracking

- Object Tracking
  - Capturing the temporal and spatial relations of vehicle objects
  - Approach:
    - Partition each frame into 2D segments using SPCPE algorithm
    - Estimated partition of previous frame is initial condition for current frame
    - Connect 2D segments in adjacent frames with least distance to obtain the trace tube

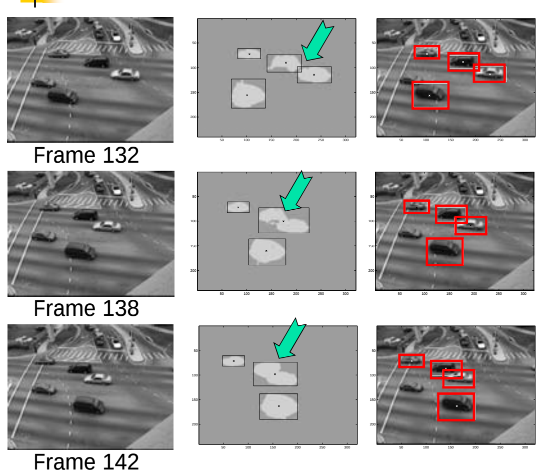
Object Tracking (Cont.)



Filtering the 'static segments'

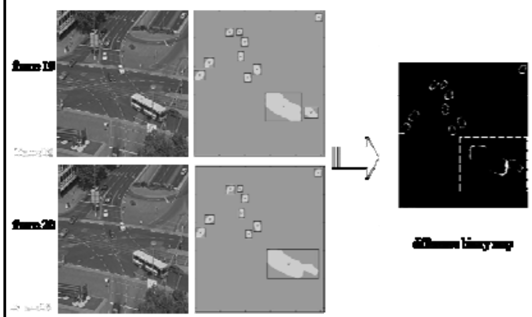
- (a) the original video frame
- (b) the segmentation result along with the bounding boxes and centroids for (a)
- (c) the segments with diagonals are identified as 'static segments'
- (d) the final segmentation result for frame 3 after filtering the 'static segments'

Object Tracking (Cont.)



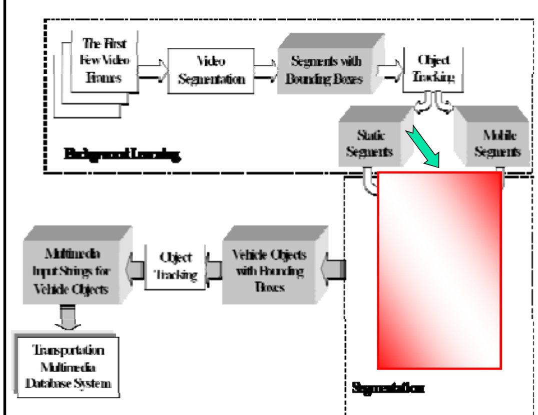
Handling object occlusion in object tracking
Backtrack-Chain-Update-Split algorithm[Chen00]

Object Tracking (Cont.)

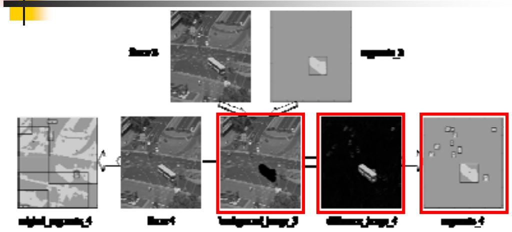


Handling object occlusion in object tracking [Chen01a]

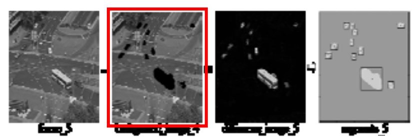
The basic workflow of the proposed framework



Self-Adaptive Background Learning and Subtraction

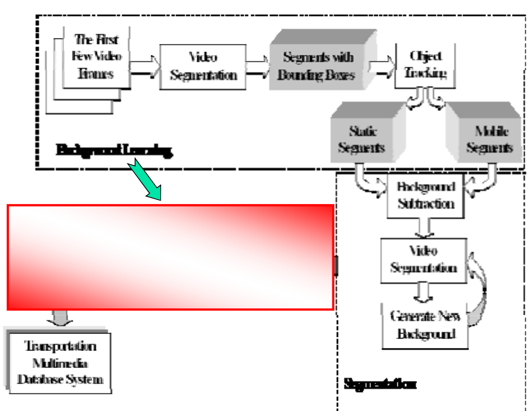


• Generating the initial reference background image (background_image_3)
• Subtracting it from the next frame (frame 4)
• Applying the segmentation on the obtained difference image (difference_image_4)



Self-adaptive adjustment in generating new background images

The basic workflow of the proposed framework



MATN and Multimedia Input Strings

- Multimedia Augmented Transition Network (MATN) Model [Chen01]
  - represented as a labeled directed graph, called a *transition graph*
  - uses multimedia input strings as the inputs
  - to index and model the spatio-temporal relations of the vehicle objects in the video sequence

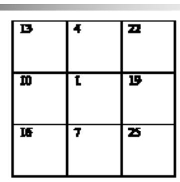
MATN and Multimedia Input Strings (Cont.)

- Multimedia Input Strings
  - Adopt the notations from regular expressions
    - Capture the spatio-temporal relations of objects in video frames
  - Each segment (object) is bounded by a minimal bounding rectangle (MBR)
    - Using centroids of MBRs for spatial reasoning
    - divide each image or video frame into nine equal size region (can have more regions)

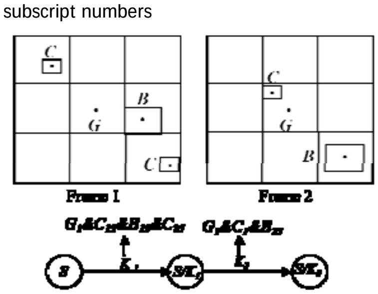
Symbol Definitions

- the ground (***G***) is selected as the target object
- the segments are denoted by ***C*** for cars or ***B*** for buses
- the "*overlapping*" objects are denoted by symbol ***O***
  - has the corresponding links to the related segments in the preceding frame

MATN and Multimedia Input Strings (Cont.)

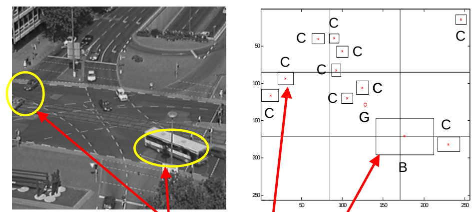


(a) the nine sub-regions and their corresponding subscript numbers



(b) an example MATN model and the multimedia input strings

Experimental Analysis (Cont.)



$G_1\&C_{10}\&C_{10}\&C_{13}\&C_4\&C_4\&C_4\&C_1\&C_1\&B_{25}\&C_{25}\&C_{22}$

- From left to right and top to bottom
- two gray cars ($C_{10}\&C_{10}$) are in the left middle area
- one white car is located in the upper left area (C_{13})
- three cars are in the upper middle area ($C_4\&C_4\&C_4$)
- two cars are located in the middle area ($C_1\&C_1$)
- one bus (B_{25}) and one dark gray car (C_{25}) are in the lower right corner
- one white car driving towards northeast is located in the upper right area (C_{22})

Experimental Analysis (Cont.)

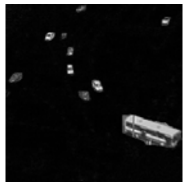
frame 19: $G_1\&C_{10}\&C_{10}\&C_{13}\&C_4\&C_4\&C_4\&C_1\&C_1\&B_{25}\&C_{25}\&C_{22}$

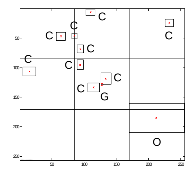
frame 25: $G_1\&C_{10}\&C_{10}\&C_{13}\&C_4\&C_1\&C_4\&C_1\&C_1\&O_{25}\&C_{22}$

frame 28: $G_1\&C_{10}\&C_{10}\&C_{13}\&C_4\&C_1\&C_4\&C_4\&C_1\&C_1\&O_{25}\&C_{22}$

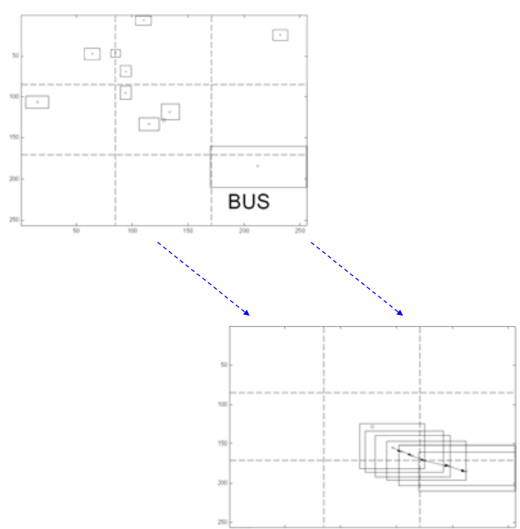
frame 34: $G_1\&C_{10}\&C_{13}\&C_{13}\&C_1\&C_4\&C_4\&C_1\&C_1\&O_{25}\&C_{22}$

Experimental Analysis (Cont.)





Experimental Analysis (Cont.)



Conclusions

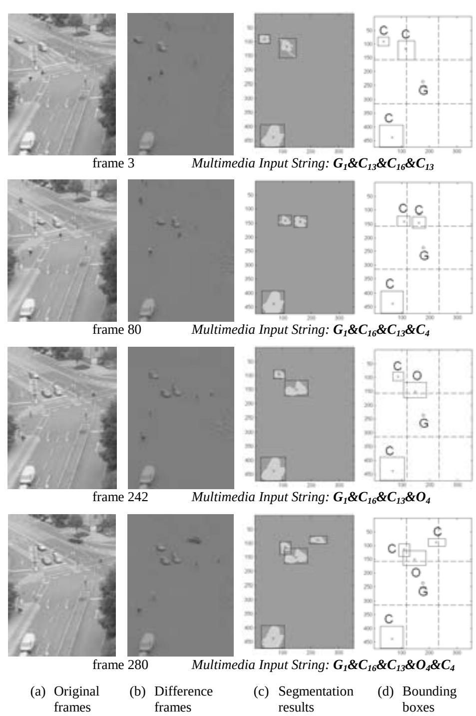
- A learning-based spatio-temporal vehicle tracking and indexing framework
  - seeks to bridge an important missing link between transportation data management and multimedia information technology
  - provides unsupervised video data storage and access for real-time traffic operations
  - can deal with very complex situations vis-à-vis intersection monitoring

Conclusions (Cont.)

- The proposed framework incorporates
  - SPCPE algorithm
  - self-adaptive background learning and subtraction method
  - object tracking
  - MATN model and Multimedia input strings

References

1. S.-C. Chen, M.-L. Shyu, S. Peeta, and C. Zhang, "Unsupervised Automated Learning-Based Spatio-Temporal Vehicle Tracking and Indexing for Transportation Multimedia Database Systems," *Transportation Research Board (TRB) 2002 Annual Meeting*, January, 2002, USA.
2. S.-C. Chen and R. L. Kashyap, "A Spatio-Temporal Semantic Model for Multimedia Database Systems and Multimedia Information Systems," *IEEE Trans. on Knowledge and Data Engineering*, vol. 13, no. 4, pp. 607-622, July/August, 2001.
3. S.-C. Chen, M.-L. Shyu, C. Zhang, and R. L. Kashyap, "Object Tracking and Augmented Transition Network for Video Indexing and Modeling," 12th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2000), pp. 428-435, November 13-15, 2000, Vancouver, British Columbia, Canada.



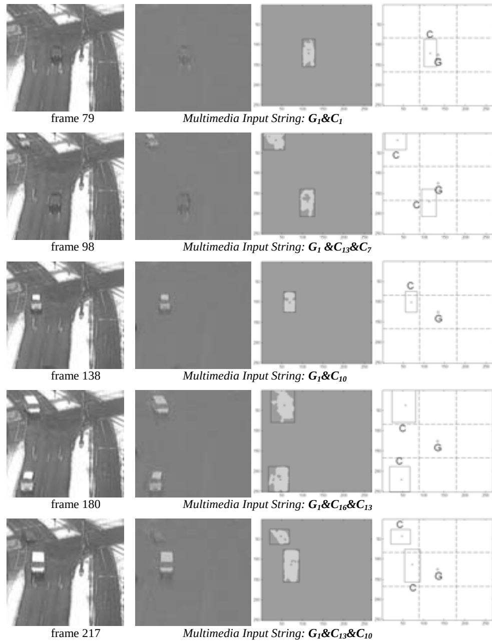
frame 3 *Multimedia Input String:* $G_1\&C_{13}\&C_{16}\&C_{13}$

frame 80 *Multimedia Input String:* $G_1\&C_{16}\&C_{13}\&C_4$

frame 242 *Multimedia Input String:* $G_1\&C_{16}\&C_{13}\&O_4$

frame 280 *Multimedia Input String:* $G_1\&C_{16}\&C_{13}\&O_4\&C_4$

(a) Original frames (b) Difference frames (c) Segmentation results (d) Bounding boxes



(a) Original frames (b) Difference frames (c) Segmentation results (d) Bounding boxes

Self-Adaptive Background Learning and Subtraction (Cont.)

- Self-adaptive adjustment
  - Motivation
    - The trail of a moving vehicle will be identified as part of the vehicle object
    - When an object moves, a small part of the background area will appear in the current frame though this area was identified as part of the vehicle object in the preceding frame
    - Causes inaccurate segmentation results after processing a number of frames
  - Solution
    - by shrinking the size of the bounding box of each segment before constructing a new background image for use in the next frame's segmentation based on the current segmentation results
    - can be thought as the prediction of the changes in the background area