Intelligent Video Surveillance

Stan Z. Li

Center for Biometrics and Security Research (CBSR) &
National Lab of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences

ASI-07, Hong Kong, 13 Jan, 2007

Research Motivations

- Applications: Immediate needs for automated surveillance systems in commercial, law enforcement and military applications.
- Research: An active research area in computer vision and video processing

Topics Overview

- Background modeling
- Object Tracking
  - By Blob
  - By Mean-shift
  - By Active Contour
- Discriminative Feature Learning
- Face Tracking
- Object Classification in Video
- Anomaly Detection
- Multi-Camera Fusion

Background Modeling

- Automatically model and update background in video
- For separating foreground (moving objects) from background

Methods

- Temporal average or median
- Gaussian Mixture Model
- Kernel Density Estimation
- Subspace method: Eigenbackgrounds
- Multilevel method
- PDE based method

Challenges in Background Modeling

- Illumination changes
  - Gradual (eg day-night change)
  - Sudden (eg cloud)
- Motion changes
  - camera oscillation
  - High-frequency background objects (such as swaying vegetation, rippling water and flickering monitors)
- Change in background geometry
  - New objects introduced
  - Old objects removed

Mixture of Gaussians

- Mixture of K Gaussians $(\mu_i, \Sigma_i, \omega_i)$ (Stauffer and Grimson 1999)

$$P(X_t) = \sum_{i=1}^{K} \omega_i \times \eta(X_t, \mu_i, \Sigma_i)$$

$$\eta(x, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^{'}\Sigma^{-1}(x-\mu)}$$

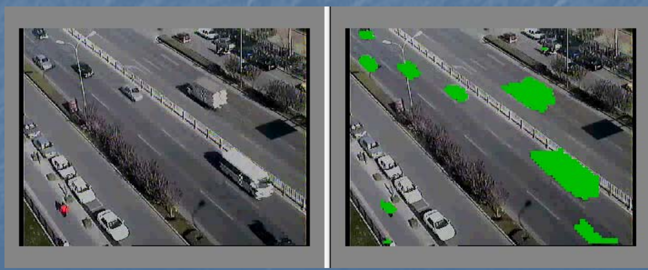
- In this way, the model copes with multimodal background distributions

Online Update Procedure

- The number of modes is arbitrarily pre-defined (usually from 3 to 5)
- All weights ω_i are updated at every new frame
- At every new frame, some of the Gaussians “match ” the current value,: for them, μ_i, Σ_i are updated by running average $\mu_i^{'} \Sigma_i^{-1} \mu_i$
- All distributions are ranked according to their and the first ones are chosen as “background”

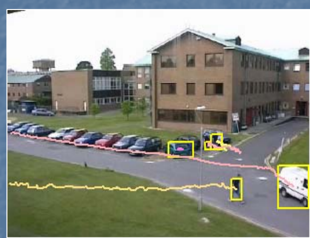
Motion Segmentation Methods

- Background subtraction
- Optical flow
- Temporal differencing



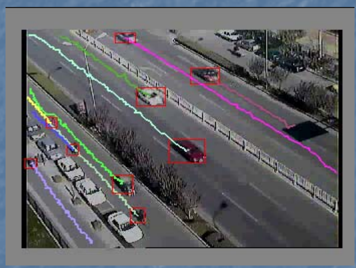
Objects Tracking – Blob Methods

- Background modeling
- Foreground segmentation
- Blob extraction
- Data association (based on: distance, color, velocity etc.)



(Yang et al., CVPR 2005)

Blob Tracking Results

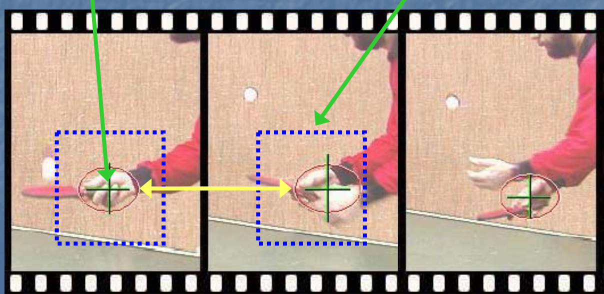


Object Tracking – Mean-Shift Method
(Comaniciu, CVPR 2000)

Start from the position of the model in the current frame

Search in the model's neighborhood in next frame

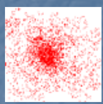
Find best candidate by maximizing a similarity function.



Repeat the same process in the next pair of frames

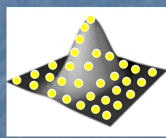
Mean Shift Tracker

- Mean Shift – A Method for Mode Seeking from data samples



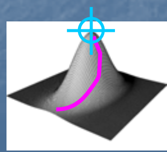
Data

Non-parametric Density Estimation



Discrete PDF Representation

Non-parametric Density GRADIENT Estimation (Mean Shift)



Kernel Density Estimation

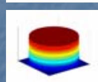
Parzen Window Method: Estimating A function of some finite number of data points *x1…xn* in *d*-dimensional Euclidean space R^d

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)$$

Various Kernels:

- Epanechnikov Kernel $K_E(\mathbf{x}) = \begin{cases} c\left(1-\|\mathbf{x}\|^2\right) & \|\mathbf{x}\| \le 1 \\ 0 & \text{otherwise} \end{cases}$

- Uniform Kernel $K_U(\mathbf{x}) = \begin{cases} c & \|\mathbf{x}\| \le 1 \\ 0 & \text{otherwise} \end{cases}$



- Normal Kernel $K_N(\mathbf{x}) = c \cdot \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right)$

Estimation of Density Gradient

$$\nabla \hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \nabla K(\mathbf{x} - \mathbf{x}_i)$$

Using the Kernel form:

$$K(\mathbf{x} - \mathbf{x}_i) = ck\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)$$

Profile of Kernel K

We get :

$$\nabla \hat{f}(\mathbf{x}) = \frac{2c}{nh^{d+2}}\left[\sum_{i=1}^{n} g_i\right] \cdot \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g_i}{\sum_{i=1}^{n} g_i} - \mathbf{x}\right]$$

$$g_i = -k'\left(\left\|\frac{x - x_i}{h}\right\|^2\right)$$

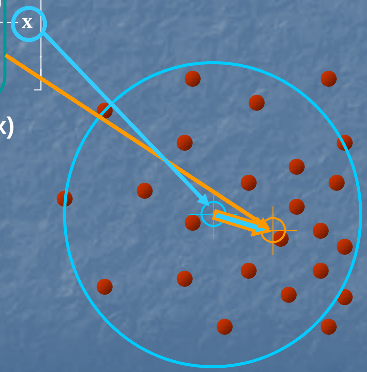
Computing The Mean Shift

Simple Mean Shift procedure (Comaniciu, CVPR 2000)

- Compute mean shift vector

$$\mathbf{m}(\mathbf{x}) = \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}\right]$$

- Translate the Kernel window by **m(x)**

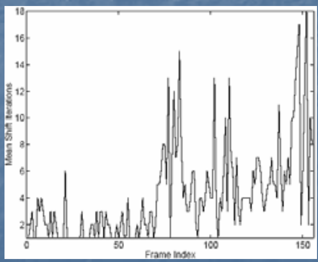


Mean-Shift Object Tracking Results

Partial occlusion

Distraction

Motion blur



(D. Comaniciu et al., CVPR 2000)

Moving Camera Tracking

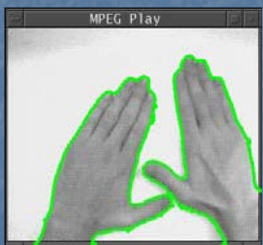
Tracking moving objects from a moving camera

Widely used method is combining appearance features (e.g. color histograms) with mean shift tracker, which is resilient to changes in object appearance due to non-rigidity and viewpoint

Active Camera Tracking

Active Contour Tracker

- For extracting object boundaries in tracking
- Minimizing an energy defined based on the shape and location



From Julien Jomier 2002

Snake Modeling

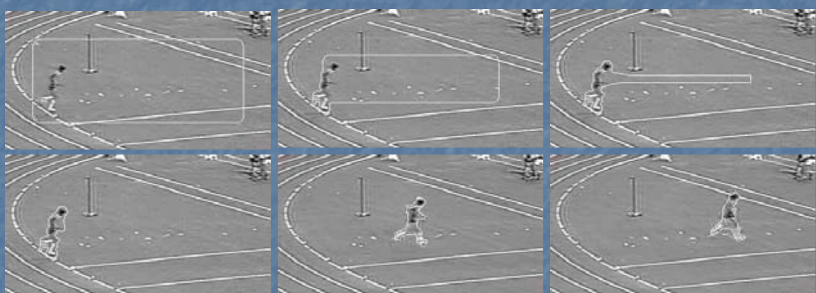
- Contour is said to possess an energy Esnake which is defined as the sum of three potential terms

$$E_{\text{snake}} = E_{\text{internal}} + E_{\text{external}} + E_{\text{constraint}}$$

-- E_snake= E(snake)
-- E_int: smoothness constraint on snake bending
-- E_ext: gradient, color, texture
-- E_con: Other constraints

- Estimate snake model in the current frame with snake in previous frame as initialization

Snakes Detection and Tracking Results

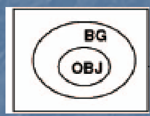


(N. Paragios & R. Derriche, PAMI 2000)

Discriminative Feature Selection

A tracking failure due to similarity of colors between object and background

(R.Collins et al., PAMI 2005)



- Fixed BG/FG features vs.
  Dynamic Discriminative Features
- Hypothesis: the image features that best discriminate an object from its background are the best features to use for tracking

Ideas

- Treating tracking as a binary classification problem (FG vs BG)
- Finding Discriminant Image, ie Optimal Combination of RGB that Best Separate FG/GB
- Finding features that maximize cross-class variance and minimize within class variance

$$VR(p,q) = \frac{\mathrm{var}((p+q)/2)}{\mathrm{var}(p) + \mathrm{var}(q)}$$

p: foreground class distribution (histogram)

q: background class distribution (histogram)

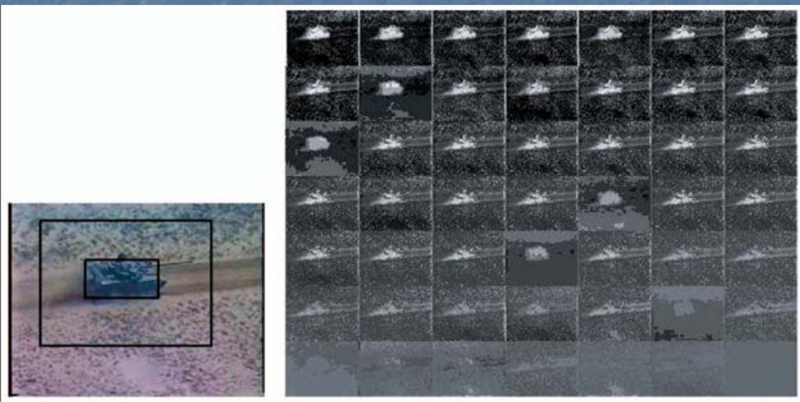
Color Feature Candidates

- Feature is a linear of combinations of R, G, B, ie F={w1R+w2G+w3B}
- Restricting wi∈{-2,-1,0,1,2} leads to 49 candidate combinations
- Finding w_i to maximize VR(p,q)

Tracking

- Online Feature Selection using p,q of the previous frame
- Compute L [x,y]=p[I(x,y)] / q[I(x,y)]
- Tracking by Applying Mean-shift on L [x,y]

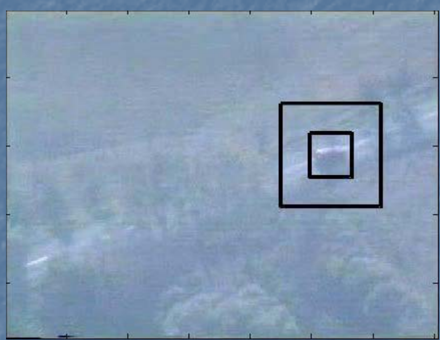
Weight Image Examples



(a) (b)

(R.Collins et al., PAMI 2005)

Tracking Hard-to-See Objects



(R.Collins et al., PAMI 2005)

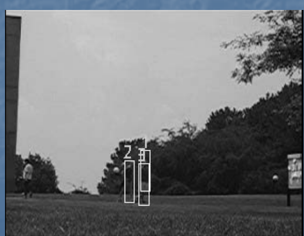
Occlusion Handling

- Feature Correspondence
  - Color , texture, geometry, et al.
  - Spatial-temporal information
- Stochastic method
  - Joint probabilistic data association
  - Bayesian reasoning
- Ensemble Method
  - Discriminative feature selection
- Multi camera methods
  - 3D coordinates correspondence
  - 2D methods (key point, principle axis correspondence)

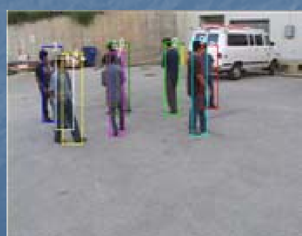
Occlusion Handling

(T. Yang et al., CVPR 2005)

(H. Grabner & H. Bischof, CVPR 2006)



(H.T. Nguyen et al., CVPR 2006)



(S.M. Khan & M. Shah, ECCV 2006)

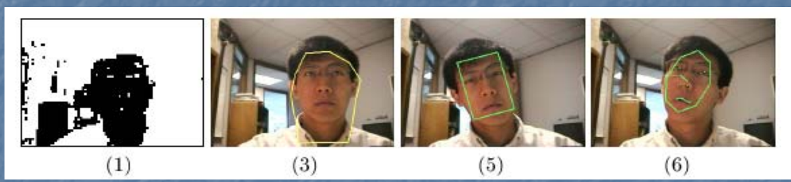
Face Tracking

Existing Techniques

- Detection based methods
  - Skin color-Based methods
  - Appearance-based methods
  - Template matching-based methods
- Tracking based methods
  - Skin color based methods (Toyama,1998)
  - Appearance model based methods (Birchfield ,1998)
  - Mean-Shift Search based methods (Comaniciu, 2000)

Existing Work: Face Tracking

- Toyama [1998] The Incremental Focus of Attention face tracking framework. (Sensitive to large head rotation)



- Comaniciu[2000] Mean-shift face tracking.

(Mean histogram based face detection model is not robust, and luminance gradient based scale estimation may fail in complex background)

Detection vs Tracking Based Methods

Issues	D—Based	T—Based
Auto Initialization	Yes	No
Scale changes	Stable	Not
Localization	Good	Average
Illumination changes	Not	Stable
Head Rotation	Not	Stable
Partial Occlusion	Not Good	OK

Take Advantages, and avoid disadvantages

Main Ideas

- Take Advantages, and avoid disadvantages:
  - Robust and Fast Face Tracker =
    Face Detector+ Face Tracker + Adaptive Switcher
- Multiview Face Detector [Li,TPAMI,2004]
  - Initializing tracker
  - Updating Tracker parameter
- Discriminant Based Face Tracker
  - Discriminant Index Map (DIM), learned online
  - Mean shift performed using DIM
- Adaptive Switcher in System
  - Management of switching between detector and tracker

Discriminant Index Map (DIM)

- In Mean shift alg: Sample weight image
  - Sample weight image = Foreground confidence map
  - Dynamic weight for each pixel
- In this work: Discriminant Index Map (DIM)
  - To discriminate foreground (face) from neighboring background
  - Track the most discriminative pixels within the surrounding region

Discriminative Color Feature Selection

Green color corresponds to Discriminant Index

- Features based on local color histograms of the tracked face and surrounding - insensitive to rotation
- Local window : Target (h=1.2w), Surrounding = 1.7*target
- Discriminant Index Histogram (DIH) - Equ.(5)
- Use most discriminative bin of DIH as the feature for tracking

Discriminant Index Histogram (DIH)

- q_t^k and q_b^k : local color histograms of the target and its surrounding background
- Online learning of DIH

$$q_d^k(n) = \begin{cases} \frac{1}{c}T & if \quad \frac{\max(q_t^k(n),\varepsilon)}{\max(q_b^k(n),\varepsilon)} > T \\ \frac{1}{c}\frac{\max(q_t^k(n),\varepsilon)}{\max(q_b^k(n),\varepsilon)} & if \quad T > \frac{\max(q_t^k(n),\varepsilon)}{\max(q_b^k(n),\varepsilon)} > 1 \\ 0 & otherwise \end{cases}$$

Discriminative Mean-Shift

- Most discriminative bins are those bins (n) for which q_d^k (n) =T/c
- Discriminant Index Map
  $w_d^k(x,y) = q_d^k(I(x,y))$
- Mean-shift tracking performed using $w_d^k(x,y)$ to updated target info

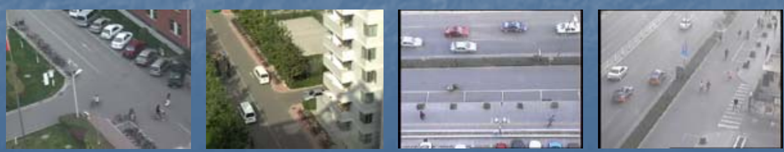
$$M = \sum_x \sum_y W_d^k(x,y)$$
$$M_x = \sum_x \sum_y x \cdot W_d^k(x,y)$$
$$M_y = \sum_x \sum_y y \cdot W_d^k(x,y)$$

Video Demo

- Single Face Tracking: Large Rotation and Jumping
- Multiple Faces Tracking: Large Rotation and Occlusion
- Active Camera Face Tracking

Object Classification

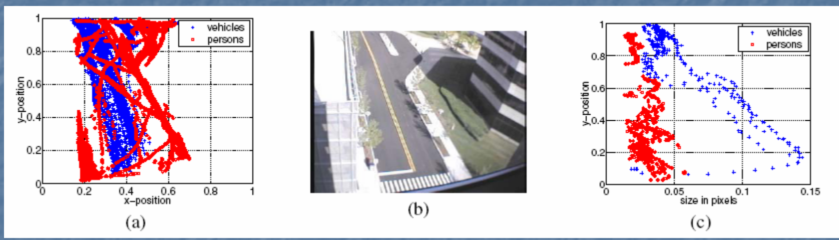
- To classify moving objects in Video, a standard pattern recognition issue.
- Challenges
  - Appearance variation due to changes in viewing angle and scale
  - real-time processing requirement.



Features for Object Representation

- Shape based classification
  - Image blob area size, compactness, apparent aspect ratio, etc.
- Motion based classification
  - Direction of motion, speed ,and periodicity
- 3D model based method
  - 3D geometry and structure. Practically difficult to implement
- Other constraints: eg x,y coordinates

Using Scene-Specific Features



- Spatial distribution of vehicles and persons
  - Left: in the x-y plane
  - in the y-size plane

Bootstrapping for Scene Transfer

- Train a low-performance baseline SVM classifier using the identified scene-independent features, on a set of labeled examples
- Apply this baseline classifier to a large unlabeled data set U in a novel urban scene. This produces ‘labels’ (and associated confidence values) for objects in the new scene.
- Treat the top 10 percent of the new labels as ‘labeled’ examples to re-train a new, scene-specific classifier, using both scene-independent and dependent features.

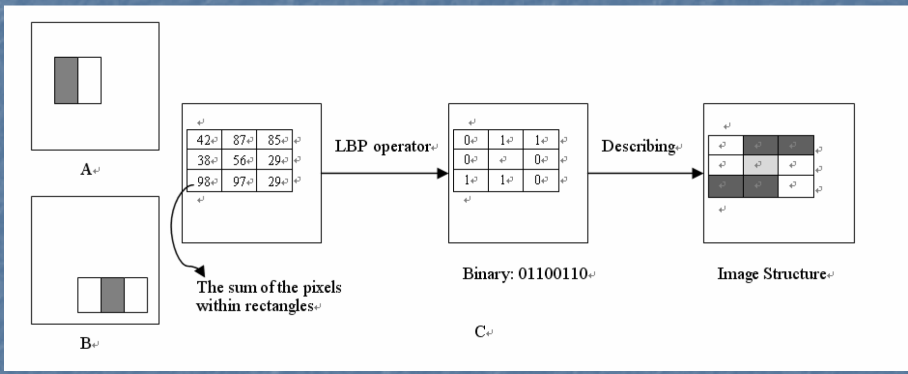
Vehicle classification (Ma and Grimson 2005)

- Vehicle type classification : All → cars vs. minivans; Cars → sedans vs. taxies.
- For fixed view angle.
- Features: edge points and modified SIFT descriptors.

Appearance Based Method

- Using appearance information to classify objects in different camera views
- Categories: car, van, truck, person, bike and group of people
- Take advantage of error correction property of Error Correcting Output Codes (ECOC) method to further deal with the challenge from large intra-class variations.

Multi-block Local Binary Pattern Feature for object representation



- MBLBP Feature can capture more information about the image structure than original haar-like feature.

Apply AdaBoost learning

- Learning effective features from a large feature set;
- Constructing weak classifiers each of which is based on one of the selected features;
- Boosting the weak classifiers into a stronger classifier.

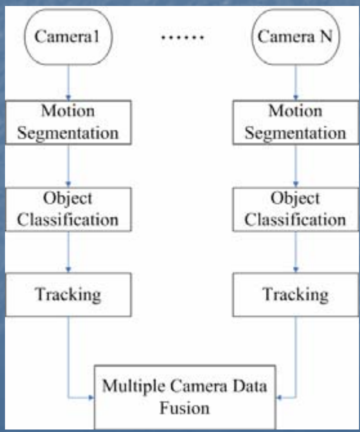
Evaluation the advantage of Error correction property of ECOC-based classifier

	Testing Sample	One-Vs-All	ECOC-based
Cars	27729	86.2%	92.6%
Vans	3516	67.6%	76.0%
Trucks	2662	66.1%	71.1%
Persons	4035	81.2%	85.2%
Bikes	7038	72.6%	78.8%
People	8058	71.6%	75.8%

Classification Results

Fusion of Multi-Camera Tracking

- Large Area Video Surveillance
- Issues
  - Sensor network
  - Camera calibration
  - Object handoff
  - Switching
  - Occlusion reasoning
  - Data fusion



Calibration

- **3-D calibration:** Compute the parameters of a camera by the relationship of 3D coordinates and the image coordinates of some known points
- **2-D calibration:** Determine the transformation matrix and then such matrix is decomposed to obtain the extrinsic parameters of the camera

Hand-off between Cameras

- Ability to Handle
  - humans as well as vehicles
  - overlapping cameras as well as non-overlapping cameras
  - indoor as well as outdoor scenes
- Explore various cues, such as geometric and kinematics, local and global appearances, for salient signatures of the objects.
- Multiple cues are fused to compute the optimal matches among all the moving objects.

Object Matching

- Region based methods
  - Color, texture, appearance, etc.
- Geometry based methods
  - 3-D method (correspondence of world coordinates)
  - 2-D method (correspondence of key points, principle axes)
- Other methods
  - Nonlinear manifold learning and correspondence
  - Trajectory based matching

Methods

- Calibration based methods
- Homography based methods

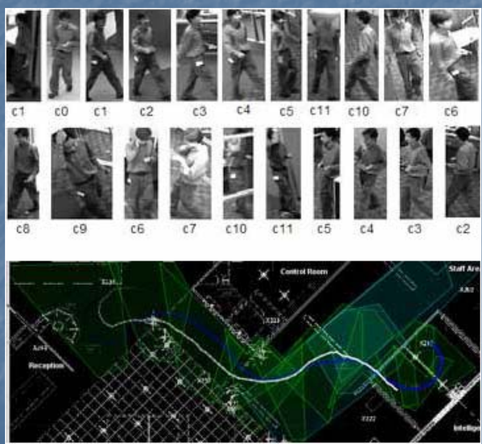
Suitable for overlapping field of views

- Appearance based methods
  - Height, size, silhouette
  - Color, texture
- Spatio-temporal methods
  - Arriving time、arriving situation
  - trajectory、velocity

Suitable for non-overlapping field of views

Object Tracking in Multi-Camera Network (Sarnoff)

- Two people
- area covered by 12 cameras
- human trajectories overlaid on a plane-view map



Calibrated Mapping

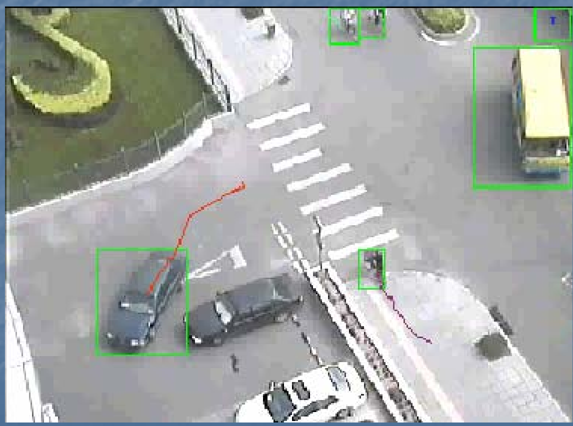
Anomaly Detection

- Detection of Add-on and Missing
- Intrusion to Forbidden Area
- Reverse Driving

Detection of Add-on and Missing

Crowd Density Estimation

Anomaly Detection and Alarming



Bad Doing

Challenges

- Occlusion Handling
- Illumination Handling
- Hand-off
- Behavior Analysis



CBSR Talents





Thank You

Contact:
Prof. Stan Z. Li
Center for Biometrics and Security Research
Institute of Automation, Chinese Academy of Sciences
szli@cbsr.ia.ac.cn