Research Activities in Computational Media Group

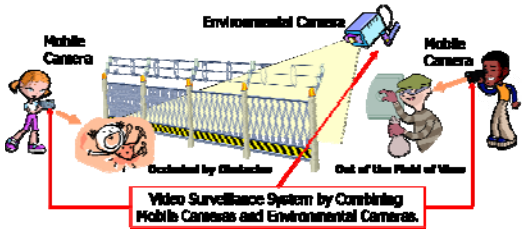
Yuichi Ohta, Yoshinari Kameda, Itaru Kitahara

Computational Media Group
Division of Computational Informatics
Center for Computational Sciences
University of Tsukuba
2007.10.31

Network Transmission and Interactive Display of Live 3D Video

Mixed Reality

Privacy Considering Video Surveillance



Topics Presented by Itaru Kitahara

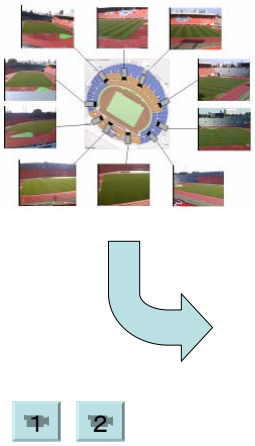
Network Transmission and Interactive Display of Live 3D Video

Purpose

- To realize a 3D free-viewpoint video system
  - Multiple audiences
  - Live sports events in a distant large-scale space (e.g., a soccer stadium)
  - Freely moving viewpoints
- To create an advanced contents technology.
  - Developing an interactive 3D video display.
  - An ordinary person can enjoy without any special knowledge about video capturing and editing.

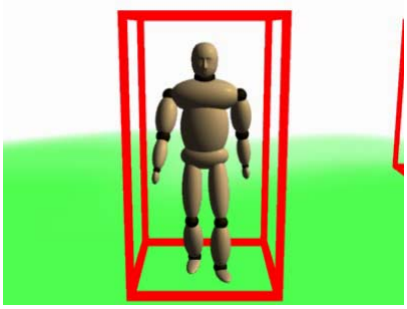
Demonstration

The world first live 3D free-viewpoint video system



Key Technology

- Small-Scale Space

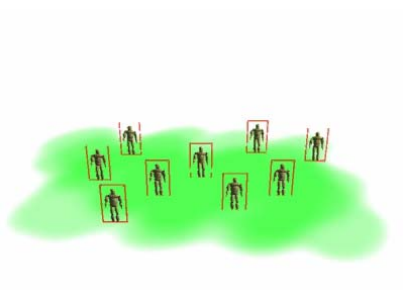


Ordinary Approach

Soccer Games:

Stadium is Large-Scale.
More than 22 players

- Large-Scale Space

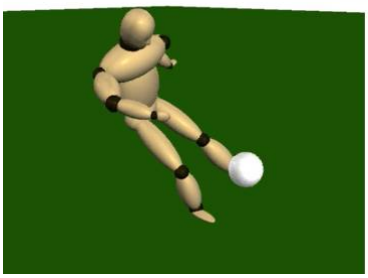


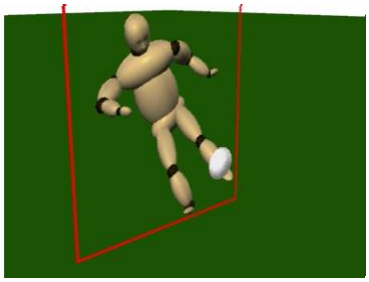
Player Billboard

A single plane and
The dynamic texture

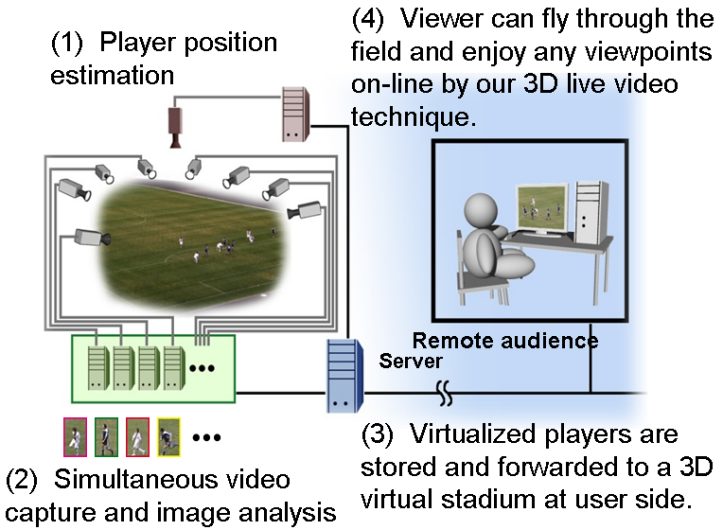
Physically Correct vs. Perceptually Correct

These two 3D videos is perceived almost equally in a large-scale space.



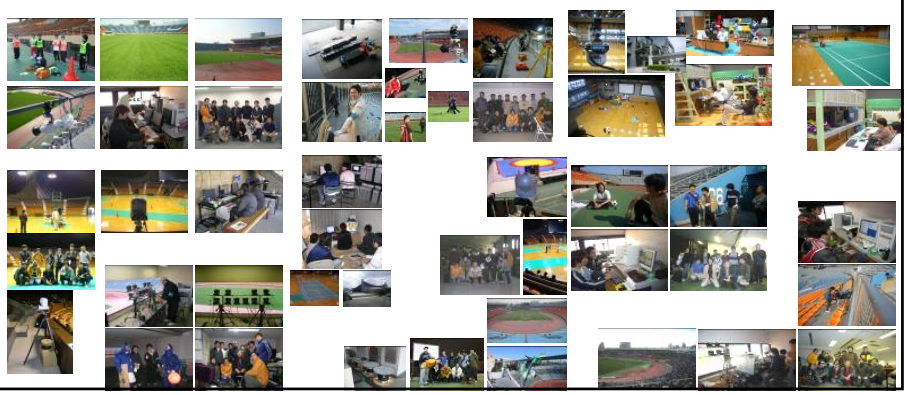


Live 3D Video System



Experiments in a Real Stadium

- Conducting more than 20 experiments
  - Tokyo National Olympic Stadium
  - Yoyogi National Gymnasium
- An unexemplified system has been developed.



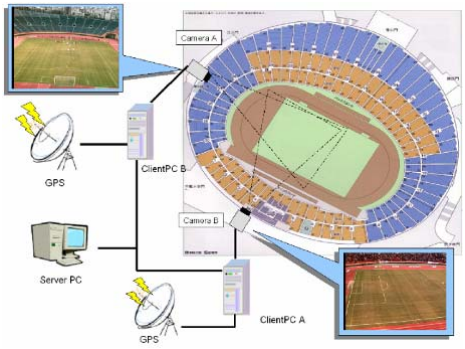
Remarkable Achievements

- International Journal of Computer Vision (IJCV)
  - The flagship journal in computer vision
- “2006 Image Electronics Technology Award”
  - The Institute of Image Electronics Engineering of Japan

Live 3D Video in Soccer Stadium

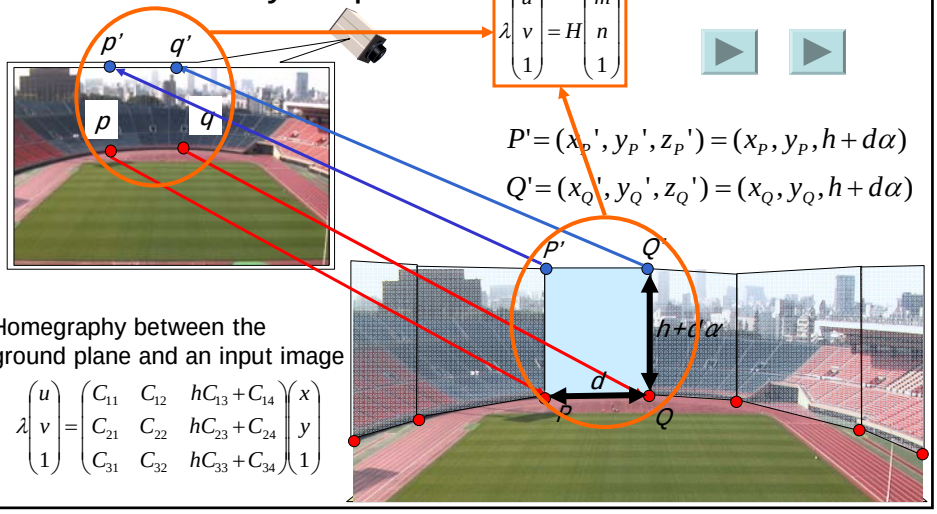
Soccer Ball Tracking

- The 3D position of the ball is calculated by the two camera images.
- Kalman Filter predicts the lost 3D ball position.
- Tracking the ball in 3D at almost on-line speed.



Back Ground Modeling

Developing a 3D background modeling method that needs only simple manual operations.



$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = H \begin{pmatrix} m \\ n \\ 1 \end{pmatrix}$$

$$P' = (x_P', y_P', z_P') = (x_P, y_P, h + d\alpha)$$

$$Q' = (x_Q', y_Q', z_Q') = (x_Q, y_Q, h + d\alpha)$$

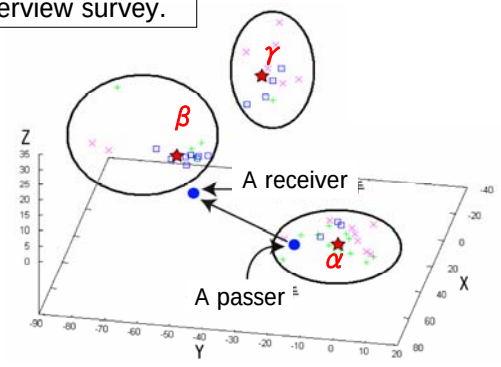
Homegraphy between the ground plane and an input image

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & hC_{13} + C_{14} \\ C_{21} & C_{22} & hC_{23} + C_{24} \\ C_{31} & C_{32} & hC_{33} + C_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Advance Contents Providing

An empathic 3D video system that counts users' personal preference.

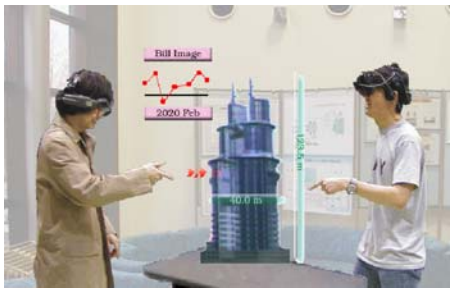
Typical 3 viewpoints where many users prefer are estimated by interview survey.



Mixed Reality

Purpose

- Shared Mixed Reality Supports Collaboration Tasks.
  - Augmenting Visual Information by CG Objects
  - Smooth Communication by Sharing Real Object's Appearance



MR Face

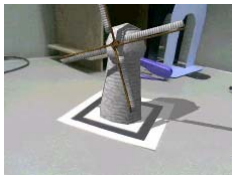
- MR Face
  - A mosaic face with real and virtual facial parts
- Restoring the eye expressions lost when wearing HMDs

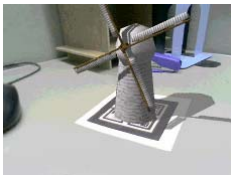
Nested Marker

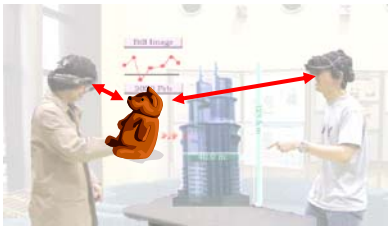
- The distance between each user and an object is different.
- Nested Marker is effective for enlarging the visual range of observers.

Single Marker

Nested Marker

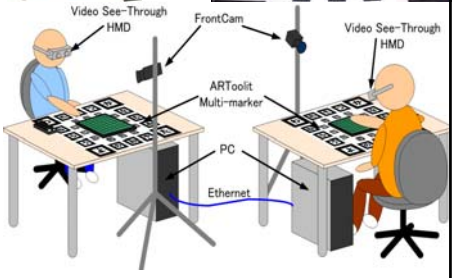






Remote Face-to-Face Mixed Reality

- Advanced Tele-Communication
  - Two people in distant places to share MR space
  - Real-time computing is key issue



Remarkable Achievements

- "University President Award 2005"
  - received by a PhD student (2006.3)
- "Microsoft Innovation Award 2007 (IT Division) "
  - Innovation Japan 2007 (2007.9)

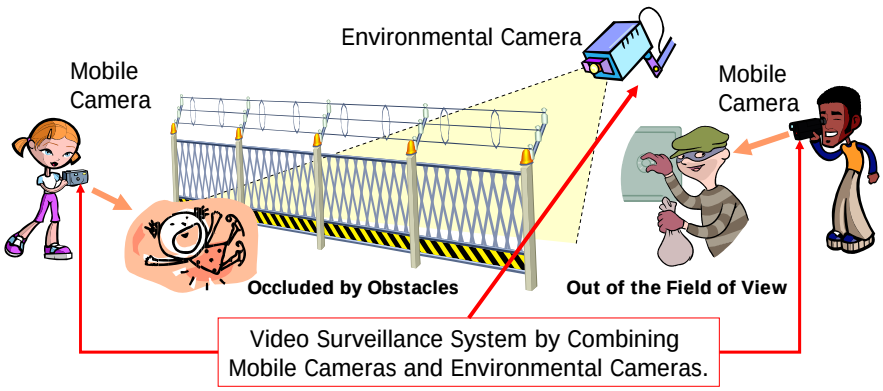
Privacy Considering Video Surveillance System by Combining the Advantages of Mobile and Environmental Cameras

Purpose

- An advanced video surveillance system
  - Combining the advantages of mobile and environmental cameras
- A Privacy Considering Video Media
  - Appropriately capture, record and display the video information with considering about the privacy information.

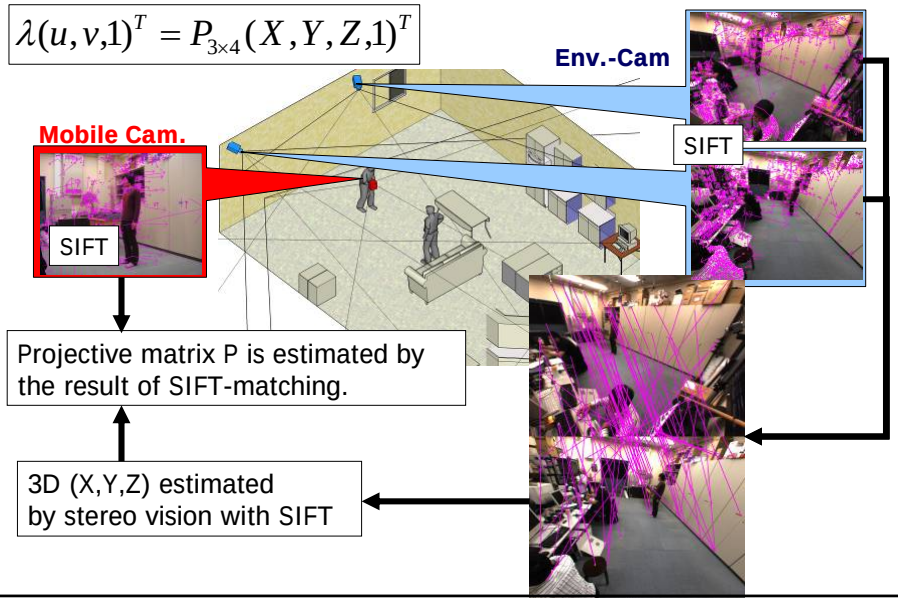
Advanced Video Surveillance

	Advantage	Weak Point
Env. Cam.	Accurate Object Tracking by easy camera calibration	Limitation of FOV
Mob. Cam.	Wide Field of View (FOV)	Camera calibration issue

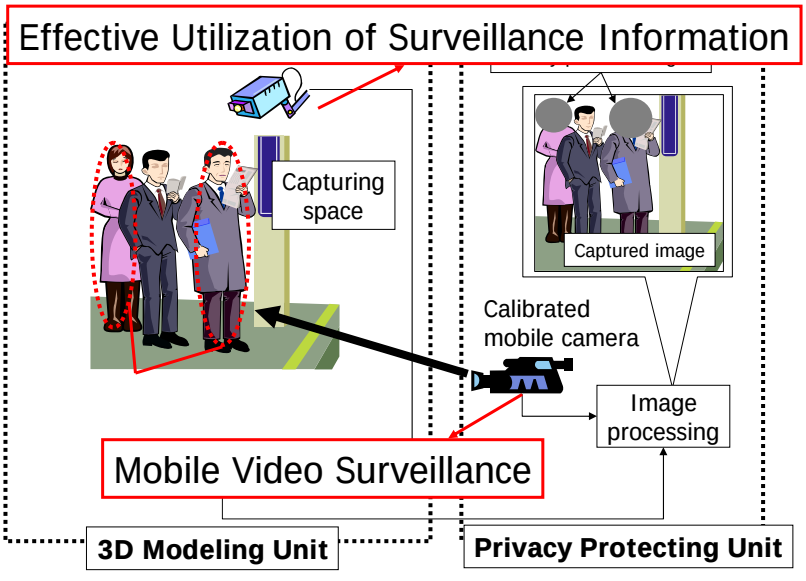


Mobile Camera Calibration

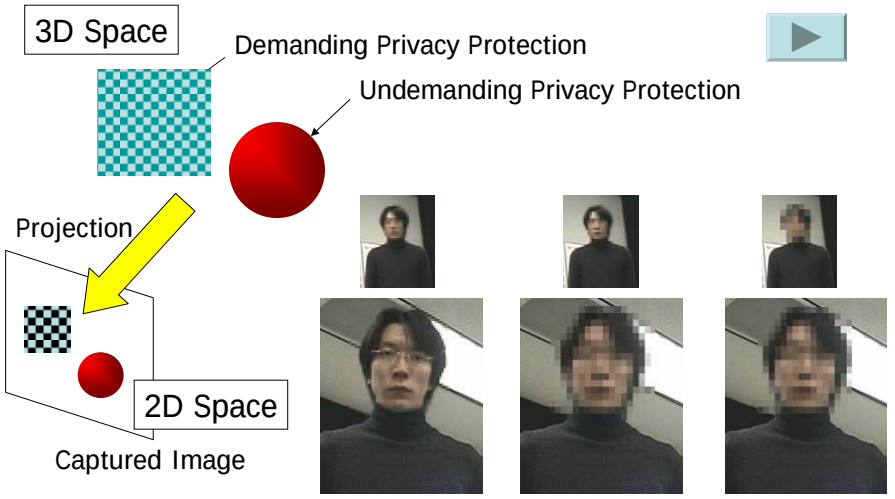
$$\lambda(u,v,1)^T = P_{3\times4}(X,Y,Z,1)^T$$



Privacy Considering Video Media



Privacy Level Control



Our method can control the privacy level by using 3D model.

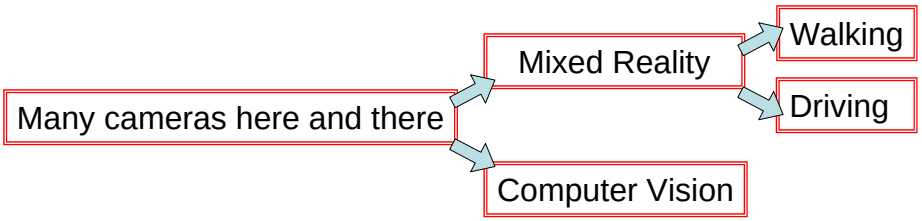
External Review on Center for Computational Sciences, University of Tsukuba

Computational Informatics
Research Division
Computational Media Group

Yuichi Ohta
Yoshinari Kameda
Itaru Kitahara

2007/10/31

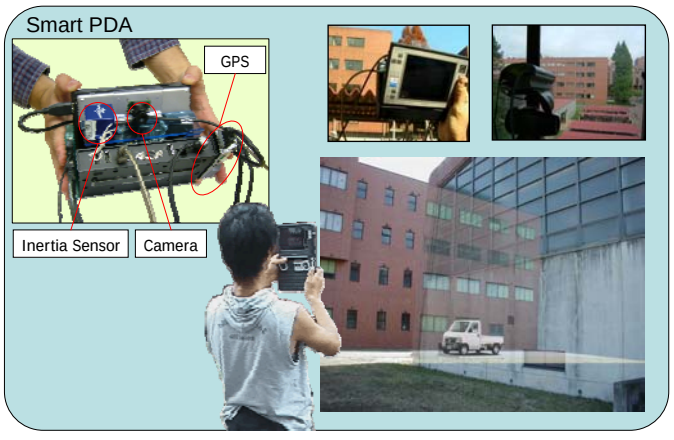
Visual Augmentation for Pedestrians
Visual Augmentation for Drivers
Massive Sensing

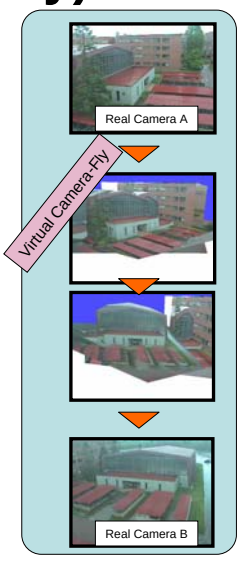


Outdoor Mixed Reality / See-Through Vision

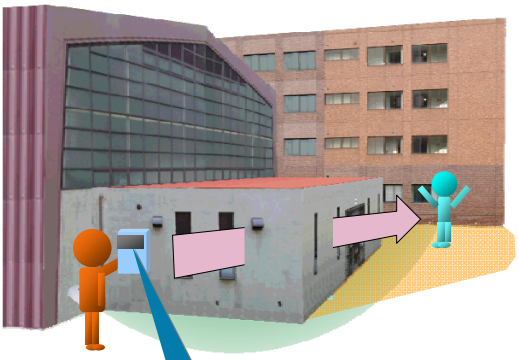
Visual Augmentation for Pedestrians (Outdoor Mixed Reality)

Pedestrian + MR gadget + Surveillance cameras
- See-Through Vision
- Smooth Video Hopping

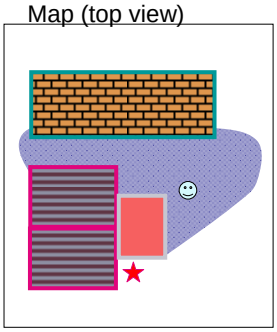




See-Through Vision

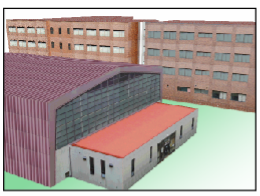


Map (top view)



"See-Through" on MR Display

Image Synthesis for MR Display



3D model

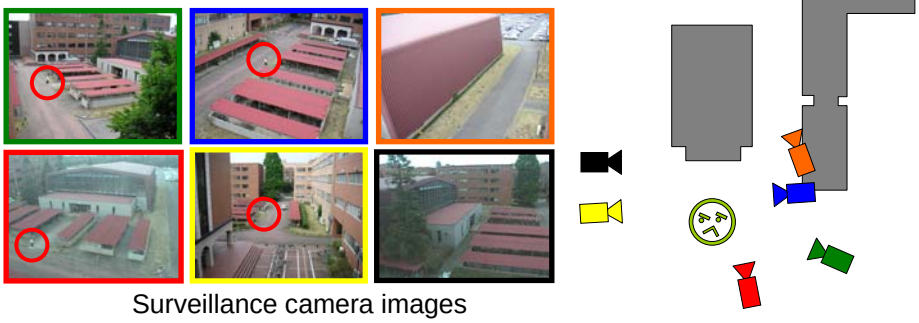
Live images captured by surveillance cameras

Intergrated view

Real image

Smooth Video Hopping

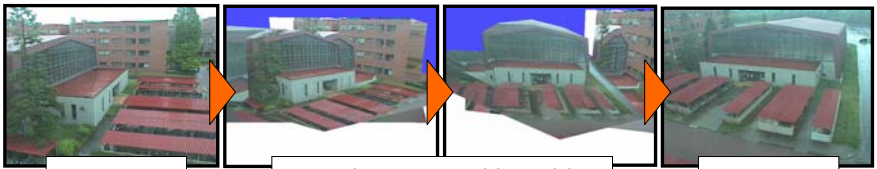
- Problem on monitoring via many cameras
  - On switching cameras, images are instantly changed
    - You need to understand how camera images change.
    - You have to learn the locations and directions of the surveillance cameras in the system beforehand.



Surveillance camera images

Idea of Smooth Video Hopping

- Visualize camera transition between cameras by pseudo-3D transition video sequence
  - Virtually show images seen from intermediate viewpoint on switching cameras



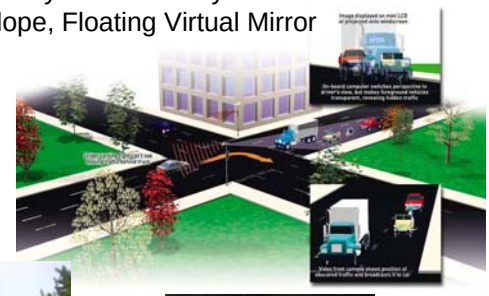
Camera A

Pseudo-3D transition video (Intermediate viewpoint)

Camera B

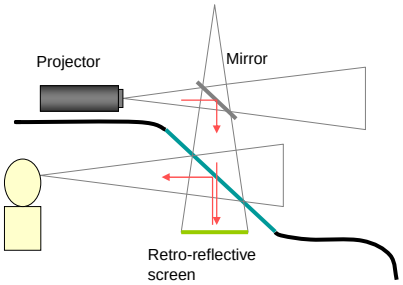
Visual Augmentation for Drivers (Mixed Reality in ITS)

Vision enhancement for drivers with support of roadside cameras
- Navigation by Mixed Reality
- Virtual Slope, Floating Virtual Mirror

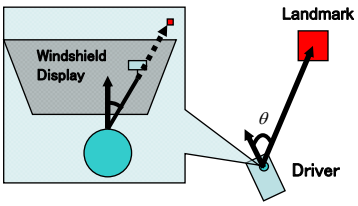


Windshield Display and MR Navigation

Windshield Display (WSD)



MR Navigation

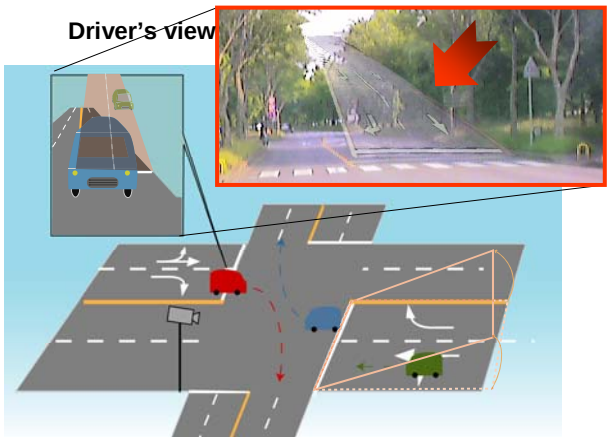


(GPS + Inertia sensor)

Nagivation icons on WSD

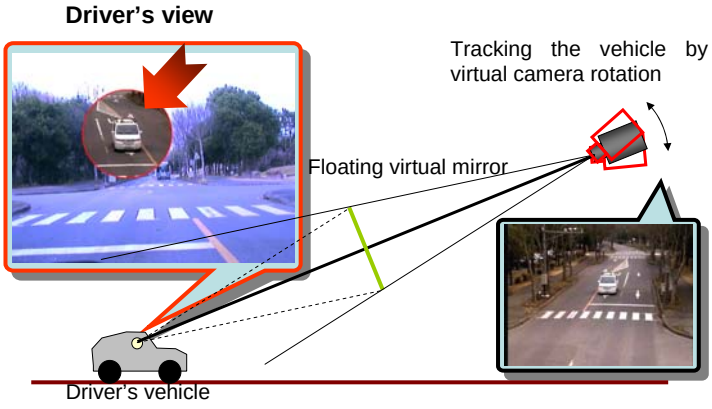
Virtual Slope

- Safety check on making right turn
- Visualize hidden vehicle on virtual slope



Floating Virtual Mirror

- Safety check on making left turn
- Visualize rear and surrounding area in form of mirror



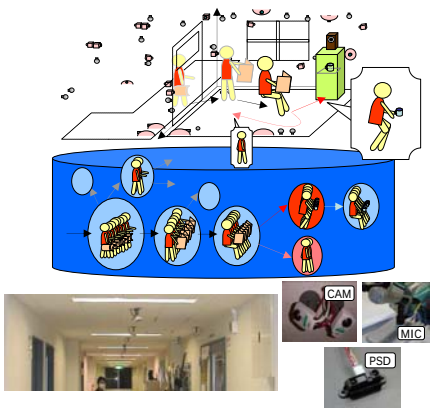
Environment and Simulator

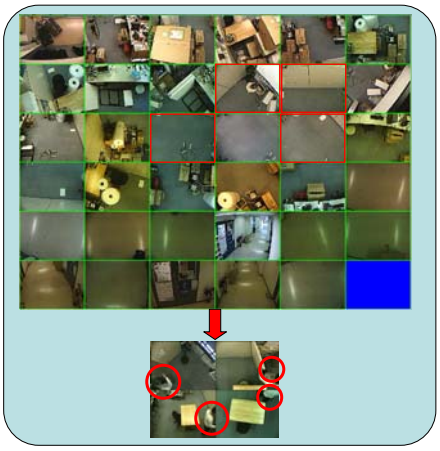
Cameras on real road commonly used in our campus

Simulator (1st, 2nd)

Massive Sensing

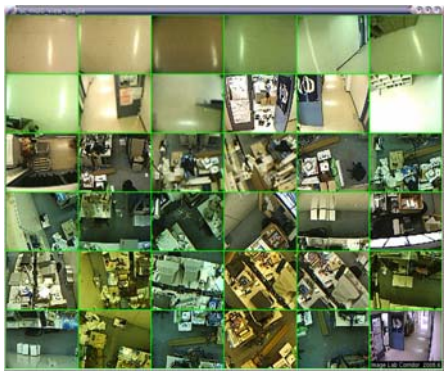
Sensing and understanding of human behaviors with a massive number of sensors (cameras, microphones, etc.)





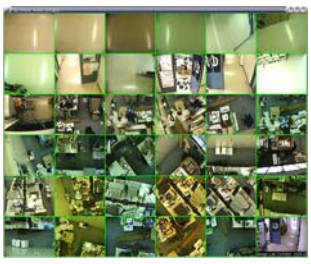
We use cameras.
MANY cameras.

- As so many cameras are coming…
  - Hard to see/understand what is going on
  - Hard to maintain camera positions/directions



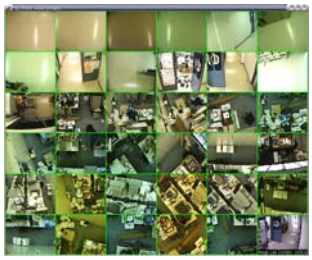
Address of Massive Sensing

- Many cameras (+ microphones, etc)
  - No calibration in advance
  - No special/uniform layout (scattered arbitrarily)
  - Various sensing ability
- Detection/Recognition
  - Relies heavy on data analysis
  - On-the-fly training



Visual Surveillance Using Multiple Cameras

- Monitoring videos
  - Inter-camera Region of Interest (ROI)
  - Dominant cameras
- No layout knowledge in advance
  - On-the-fly training
  - No heuristics of the scene



Visual Surveillance Using Multiple Cameras

- Monitoring videos
  - Inter-camera Region of Interest (ROI)
  - Dominant cameras

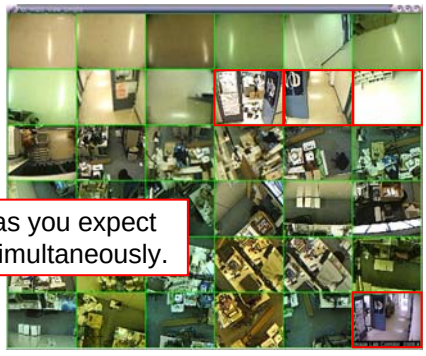
Visual Surveillance Using Multiple Cameras

- Monitoring videos
  - Inter-camera Region of Interest (ROI)
  - Dominant cameras

Find ROIs that share the same target space

Visual Surveillance Using Multiple Cameras

- Monitoring videos
  - Inter-camera Region of Interest (ROI)
  - Dominant camera set



Suggest cameras you expect to see objects simultaneously.

Video Process

- Web cameras (roughly synchronized)
  - ✕ 320[pix] × 240[pix] × 3[rgb] × 15[fps] × 60[sec] × 60[min] × 8[hour] × 36[unit] ≒ 3.6TB
  - △ 8[block] × 8[block] × 1[gray] × 10[fps] × 60[sec] × 60[min] × 8[hour] × 36[unit] ≒ 660MB
- Frontal Region Detection

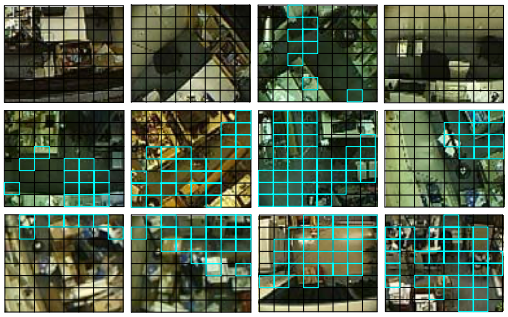
N-D vector for frame

$$\mathbf{x}(t) = \{x_1(t), \ldots, x_i(t), \ldots, x_N(t)\}$$
$$N = C \times R$$
C : # of cameras R : # of blocks

Redundant Block Elimination

Shrink the N-D space

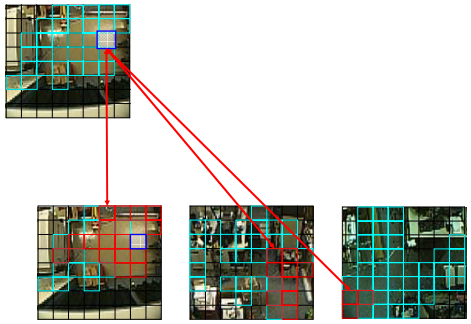
- Step1: Foreground Detection
  - Eliminate blocks that never find foregrounds
- Step2: Principal Component Analysis
  - Eliminate less important blocks that do not contribute to the calculation of values of principal axes



Example of 12 cameras:
ROIs/all blocks = 253/768

"Suggestion of Associative Blocks"

- Evaluate the co-occurrence degree between two ROIs for all the ROI pairs
- List up the highly correlated ROIs



"Dominant Camera Set"

- Evaluate the co-occurrence degree between two ROIs for all the ROI pairs
- Count the amount of correlation for any pairs of cameras
- Select cameras that belong to higher correlation pairs