Computational Media: Challenges to Enhance Computer Vision and Augmented Reality

Yoshinari Kameda

筑波大学
University of Tsukuba
Computational Media Group,
Division of Computational Informatics at CCS.

Department of Intelligent Interaction Technologies,
Graduate School of Systems and Information Engineering

Computer Vision and Image Media Lab.

Professor	Yuichi OHTA
Associate Professor	Yoshinari KAMEDA
Associate Professor	Itaru KITAHARA

Research themes
- 3D Free Viewpoint Video
- Mixed Reality (Shared / Outdoor)
- ITS (Intelligent Transportation Systems)
- Massive Sensing, Video Surveillance

3M310, The Science and Engineering Cluster,
University of Tsukuba,

Tennoudai 1-1-1 Tsukuba Science City, Ibaraki
Japan, 305-8573

Tel: +81 29 853 6556

Agenda

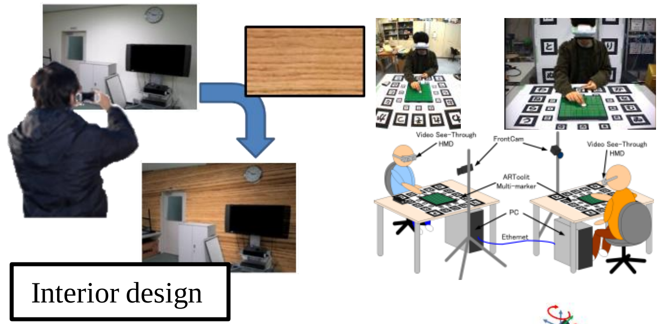
1. Overview of Computational Media Group
2. Localization of Pedestrian in town only by Computer Vision
3. Very precise reaction recording/analysis on interview scenario

Computer Vision and Image Media

Enhancing human ability throughout innovative media technology !

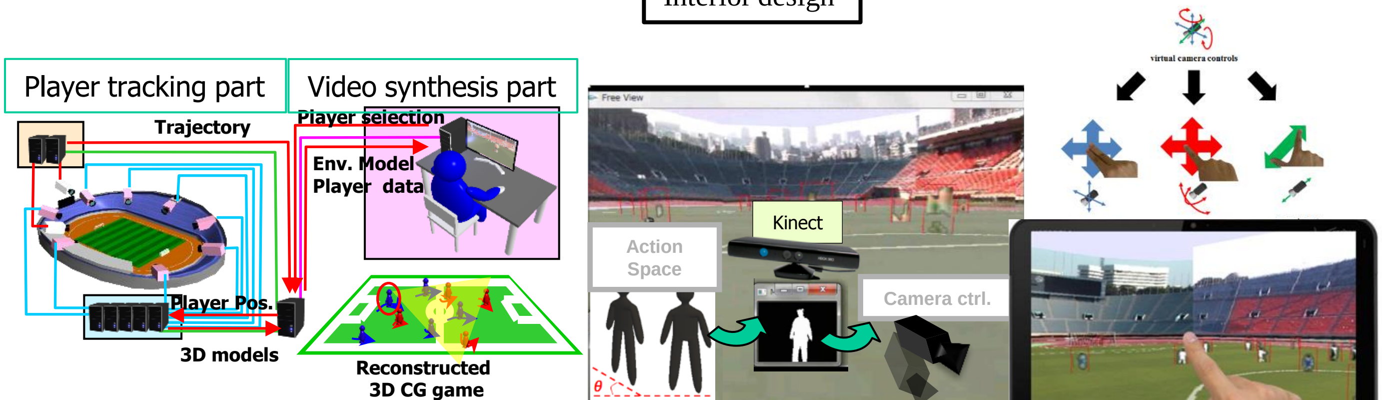
Collaborative Mixed Reality

Smooth communication in MR space
Sharing gaze, emotion, pointing, etc.



Outdoor Mixed Reality

Make invisible visible in outdoor scene

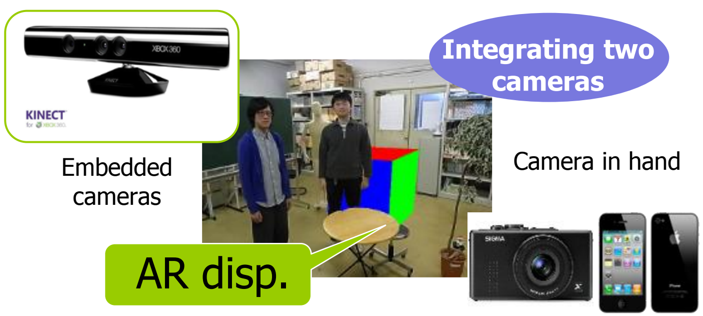


Browsing from Free Viewpoint

Live viewing of sport game, controlling the viewpoint as you wish

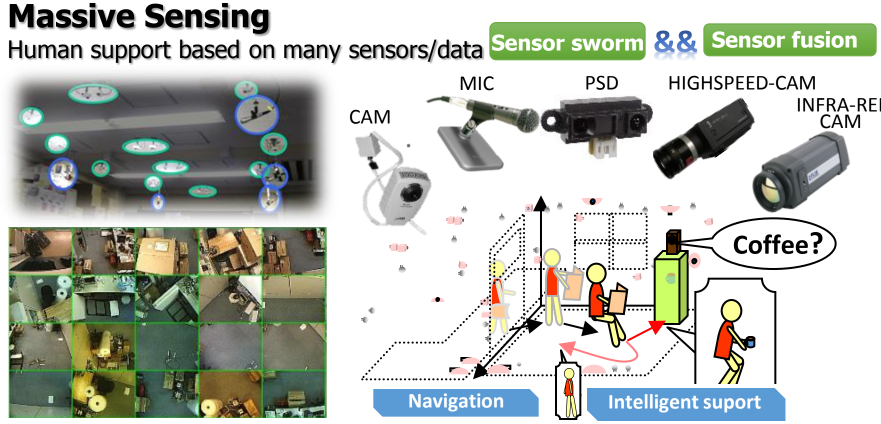
Intelligent Transport System

Visual Support by using Environmental and Mobile cameras



Massive Sensing

Human support based on many sensors/data



Hostile Intent Detection Using Video Signal

Pedestrian Location Estimation by a Single Image from Human-borne Camera

Yoshinari Kameda

University of Tsukuba, Japan

Where am I in town?

- When you are out
  - Not to get lost
  - Find some nice store
  - Navigation

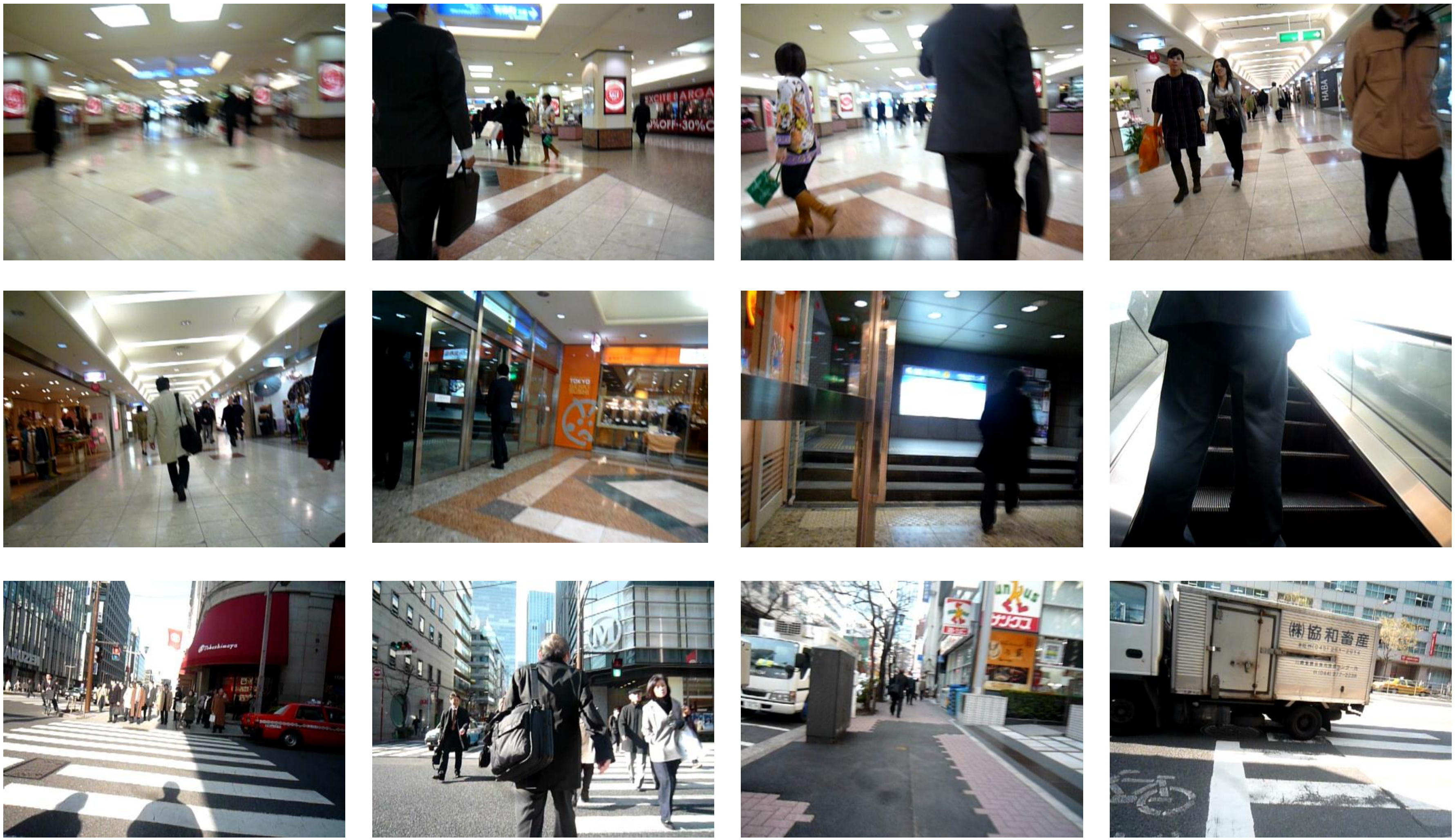
Motivation

- Get position at anywhere, both in indoor and outdoor scenes
  - GPS ··· You need sky.
  - Beacon ··· Need to embed some hardwares
  - Radio waves ··· Actually not so accurate
    - Cell phone stations, WiFi access points
- Sensor Fusion ···Yes, definitely.
- But we want one single solution that covers everywhere.

Computer Vision (What you see tells where you are!)

Examples of Walking Path

Looks difficult?

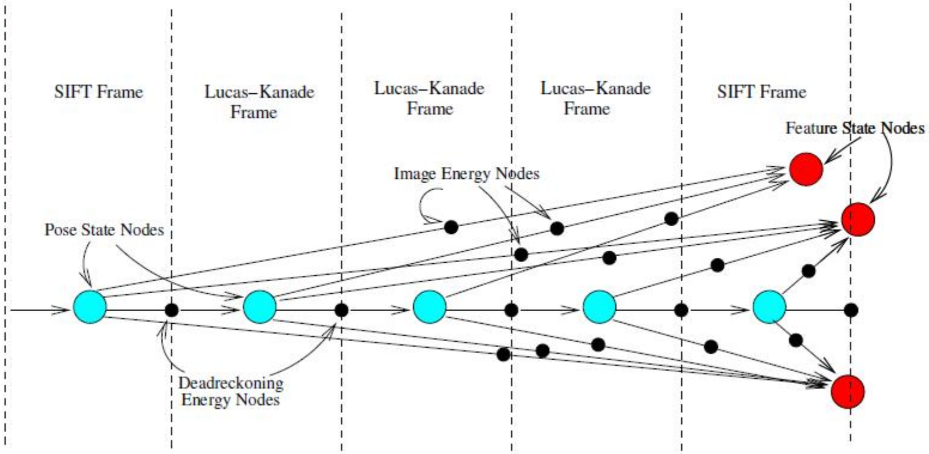


Non-Visual Outdoor Localization Techs.

- Differential GPS
  - Satellite Based Augmentation System (SBAS)
  - Multi-Functional Satellite Augmentation System (MSAS)
- Beacons scattered in the area
  - “Indoor” GPS (Indoor messaging system)
  - Radio-wave beacons
- Dead Reckoning
  - Inertia Measurement Unit(IM(), electronic compass

Relate Works

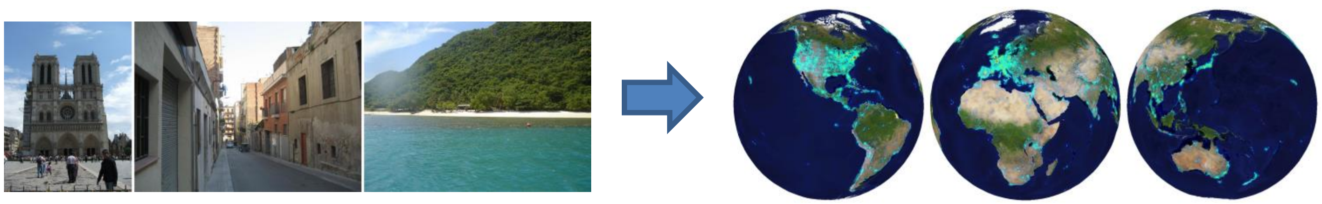
- Visual SLAM
  - Robotics（for vehicles）
    - (ex.) Outdoor SLAM using stereo vision and SIFT features
    - (ex.) SIFT Based Graphical SLAM on a Packbot



【Assumption】 Precise and dense localization (for automated driving)
Fewer moving obstacles

Related works (Image based)

- IM2GPS: estimating geographic information from a single image (CVPR2008)



- Landmark-Based Pedestrian Navigation with Enhanced Spatial Reasoning

Final Goal

The ultimate technology toward pedestrian navigation by Computer Vision (only)

- Anytime (24 hours, 365 days)
- Anywhere (indoor, outdoor, where people can go)
- Anyone (with ordinary hardware/software)

- At anywhere and by anyone -

Target Application

★Walking support for visually impaired people

To serve as alternative "Vision"

Demand for Going Out

・ High quality GIS for shopping, navigation, etc.

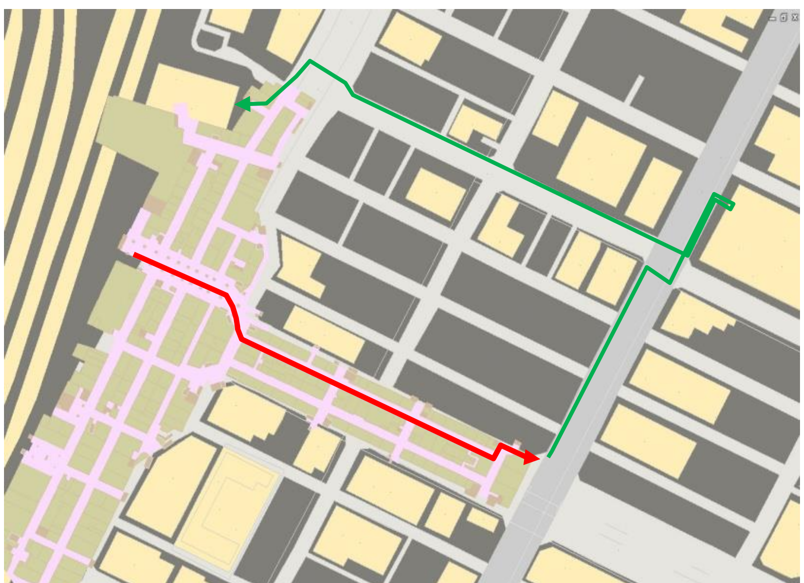
■ Collaboration with visually impaired professor

- ・ Developer of GPS based audio navigation system

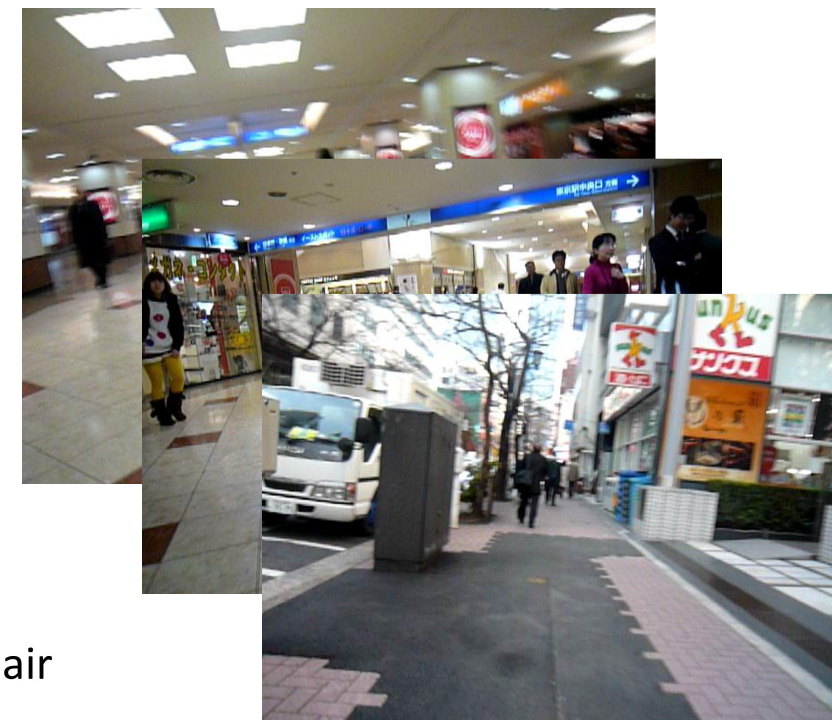
■ The method will be used as a part of sensor-fusion navigation system

An Example of Path Planning

- Visually impaired people go out after their plan is made
- Walking in downtown
  - Around Tokyo station
  - Many people, tall buildings, underground paths, etc



About 1km, red: underground, green: open air



Expected Environment

The route is set under supervision of visually impaired people

(They usually do not go out without route planning in unfamiliar place)

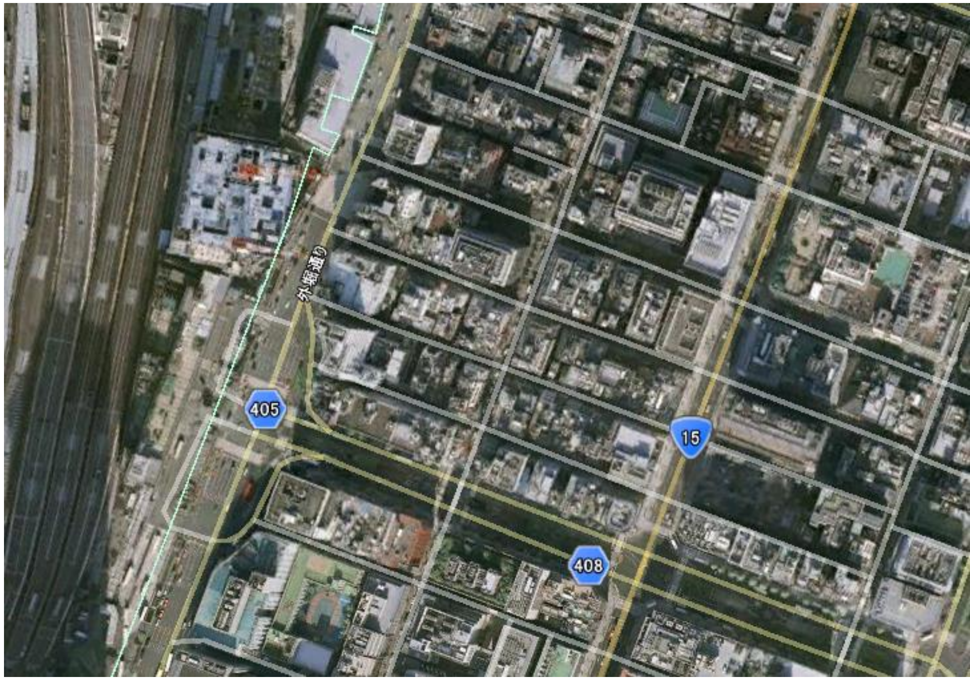
- Ground level, underground level, stairs, tall buildings, trees, people, etc
- Impossible to take shots without people, no tripod on recording
- GPS is not available or in low quality, WiFi/cell-phone based localization are also useless

Target Environment

About 900 m, daytime

From an underground exit of Tokyo Station

Route（red: underground, green: open air)



Aerial view (© Google Earth)

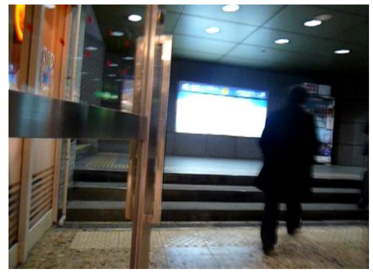
Human-borne (“First Vision”) Camera

- A big problem of camera mounting
  - Head（Cap）　×
  - Goggle/Glass　×
  - Walking stick
  - Electric pin display
  - Avoid hands and arms
- Camera direction

  = Body heading direction

Notes of Taking Snapshots





1. Camera is not squarely set to the world
2. It is impossible to take photos without people/obstacles
   - Doors, signs, etc
   - People
3. It is very hard to keep a straight (designed) path on walking
4. Wide variety of weather conditions and light conditions
5. The world is not static (construction etc)

Requirements

<u>For support of visually impaired navigation system</u>

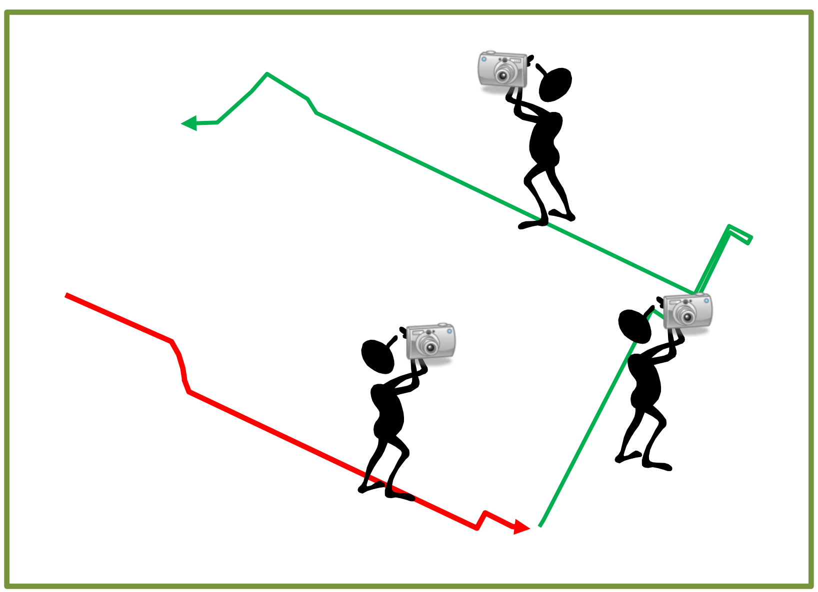
- Less than 1second for response time
  (no need of frame-rate processing)
- Standalone processing on wearable computer
  (note-size PC)
- High precision rate (low recall rate acceptable)
- Robustness
- Accuracy up to 1 meter (a door size)

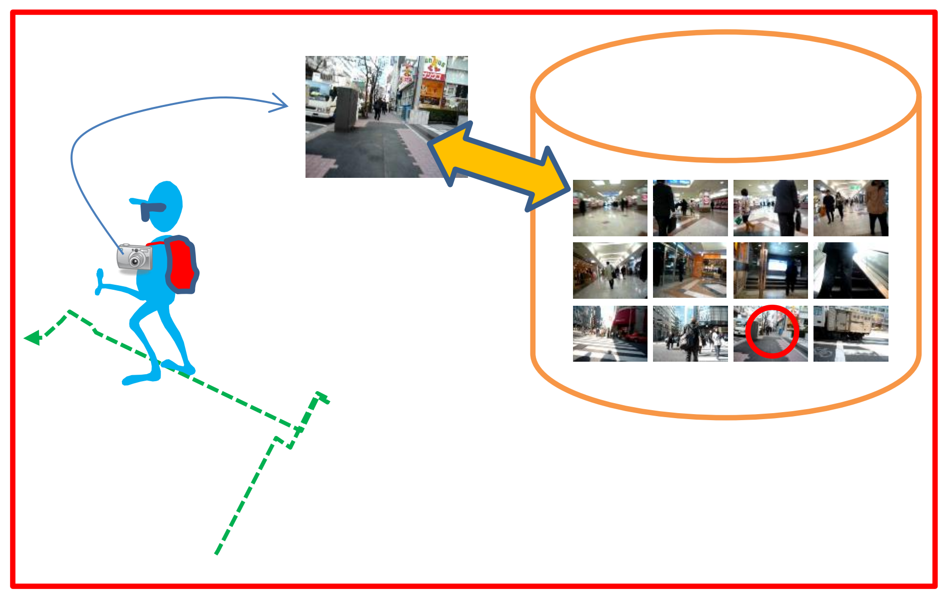
Image retrieval (outline)

a. Collect images around the area

b. Load the images based on the planned route

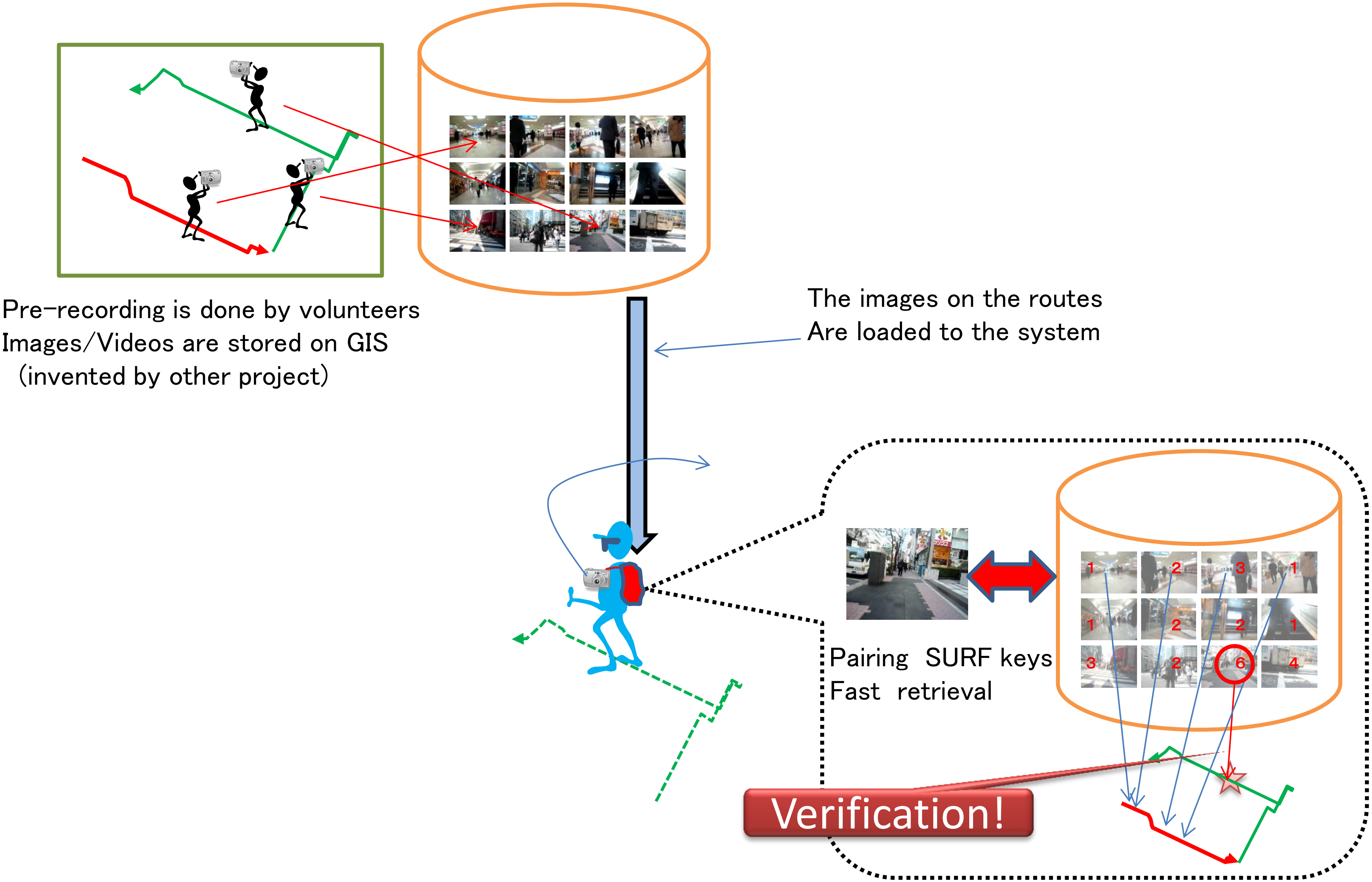
c. On-line image retrieval for every input frame





Example of SURF key matching

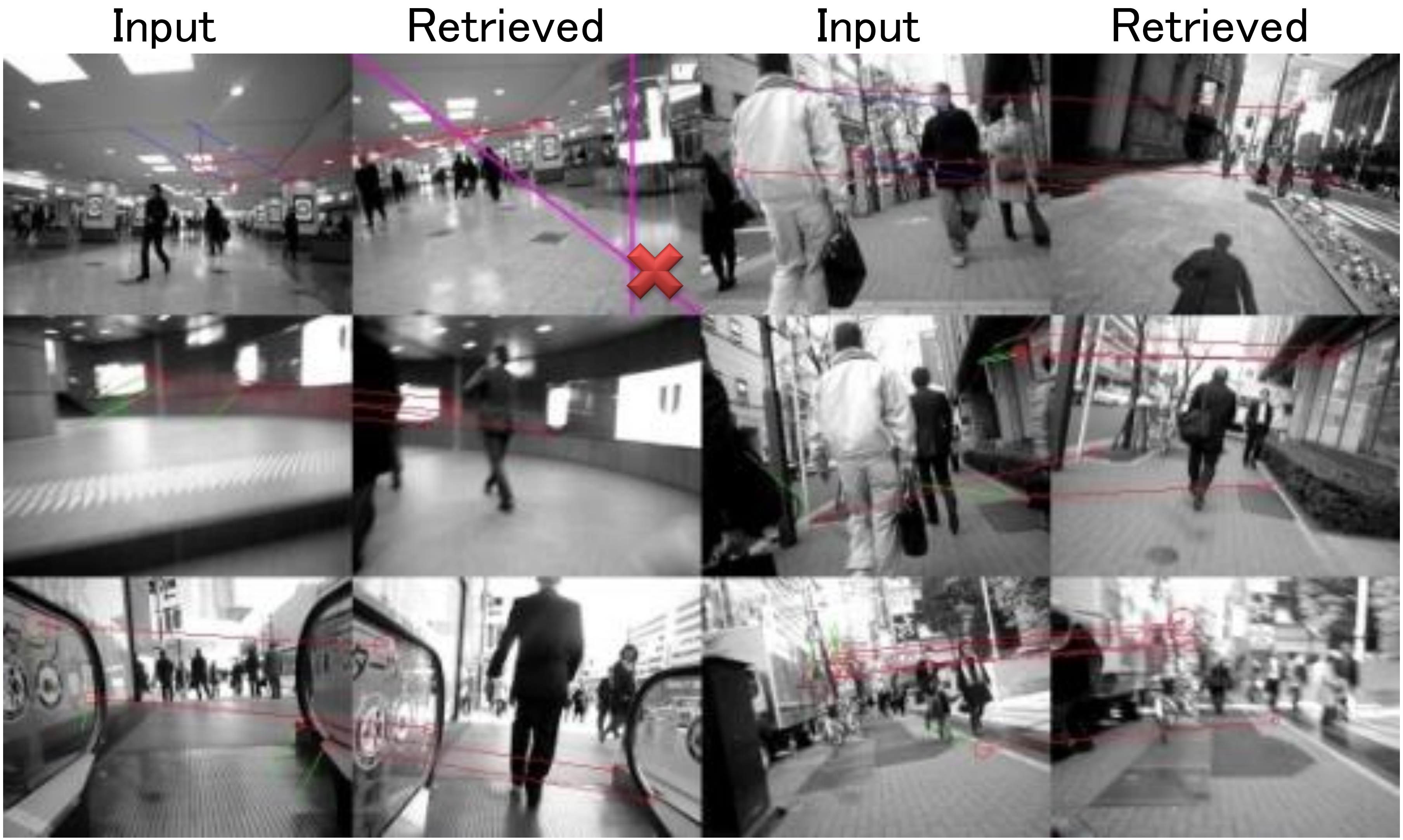
Image retrieval + Verification



Snapshots 1

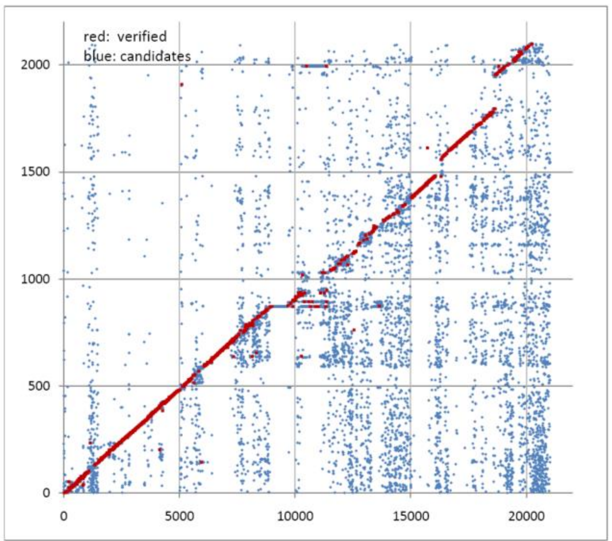
are discarded candidates

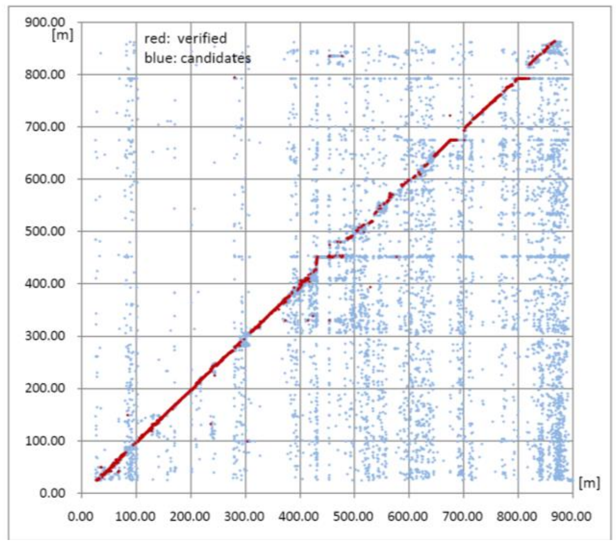
Snapshots 2

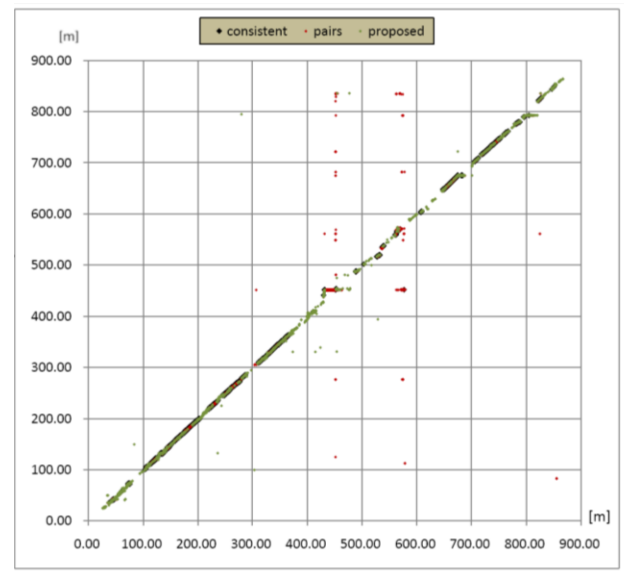


are discarded candidates

Result (Video 1)

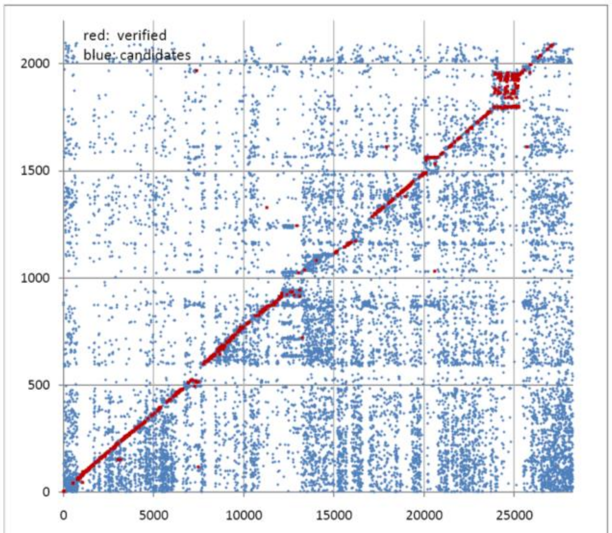


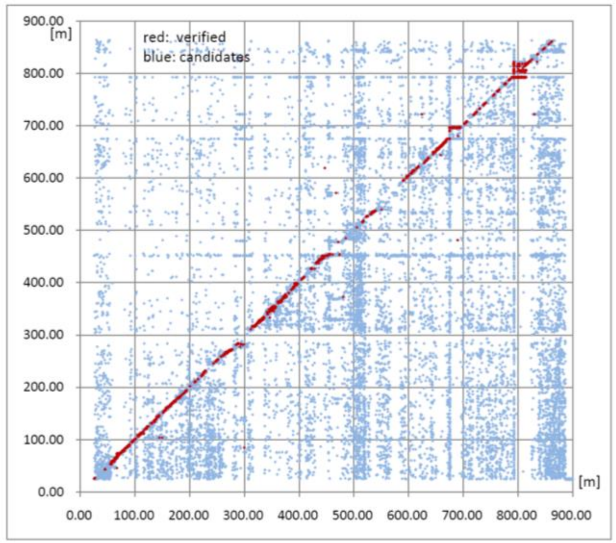


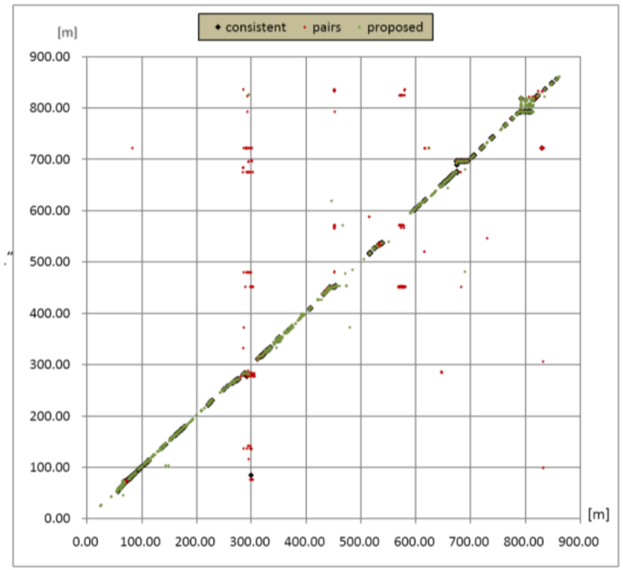


Left : By frame
Middle: By distance
Right: Comparison with normal image retrieval

Result (Video 2)

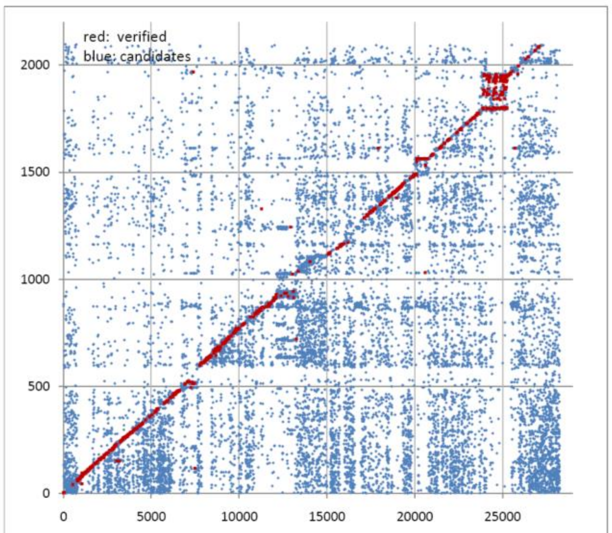


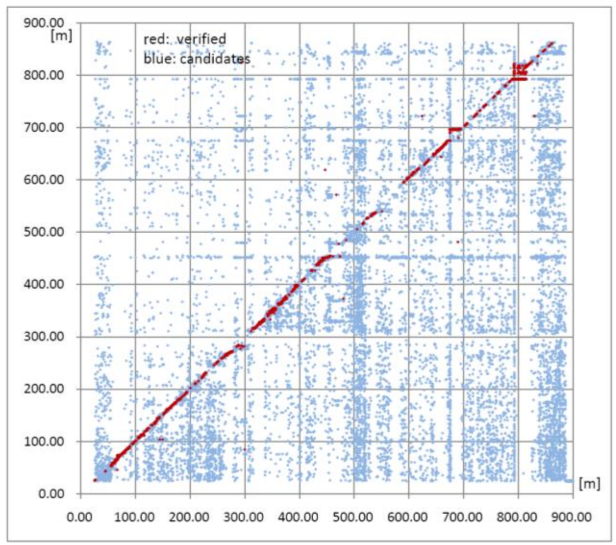


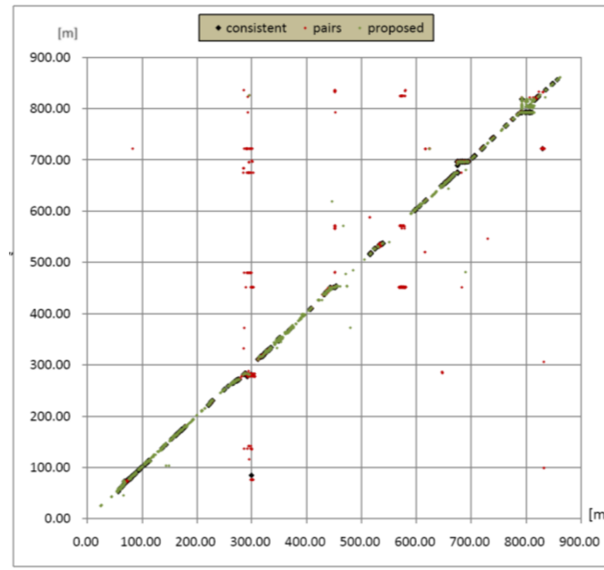


Left :	By frame
Middle:	By distance
Right:	Comparison with normal image retrieval

Result (Video 3)







Left : By frame
Middle: By distance
Right: Comparison with normal image retrieval

Conclusion

- Pedestrian localization method in Indoor and outdoor scene by computer vision
  - Designed for support of visually impaired people
  - Image retrieval in reference image dataset
  - Verification step with less SURF matched keys
- Evaluation in practical scenario

A Study on Micro-Expression Detection and Multi-Sensor Data-Set Acquisition

for Psychophysiological Analysis during Interview Scenario

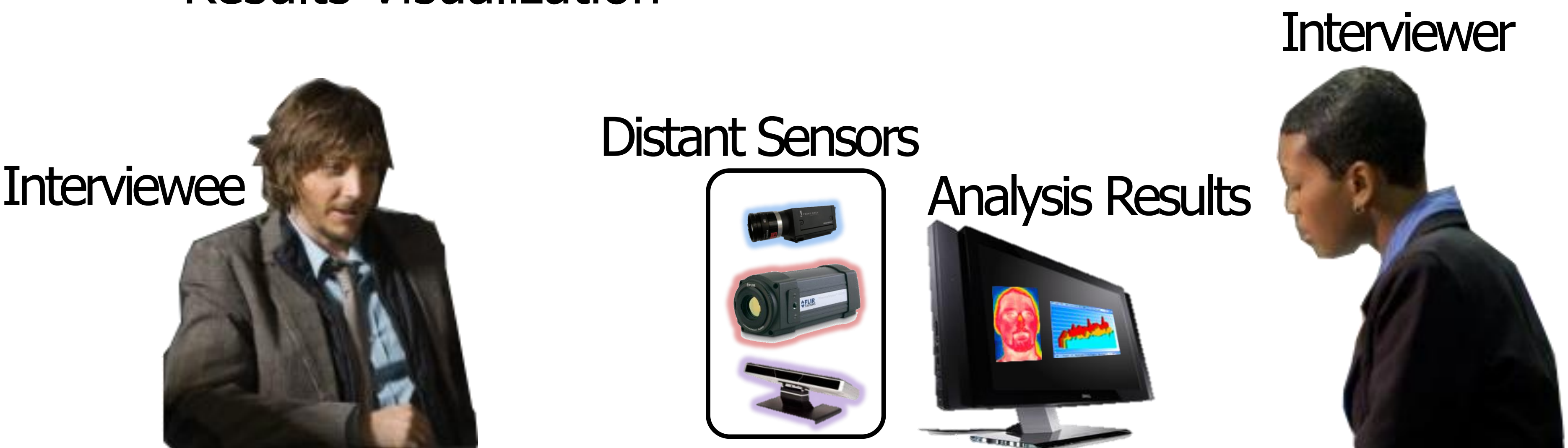
Computer Vision and Image Media Lab.

Senya Polikovsky

Our Vision…

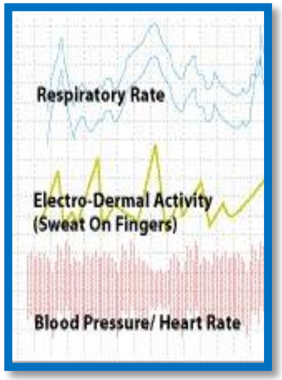
Emotion Sensing Support System (ESSS), is a human assistive technology for decision-making process during interview scenarios.

- Development and integration of sensors technology
- Development of analysis algorithms
- Establishment of psychophysiological models
- Results Visualization



Interpersonal deception theory

1. Respiration Rate
2. Electro-Dermal
3. Hear Rate



Physiology in Polygraph

Motivation
(event / context / personality)

↓

Truth / Lie

↓

Emotion ⟷ Cognition

↓

Behavior / Change of baseline

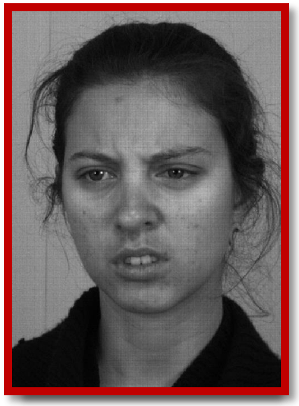
↓

"Leakage" / Cues

[1] Burgoon 1996

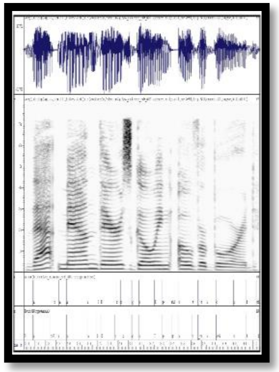
Content

Gesture / Touch

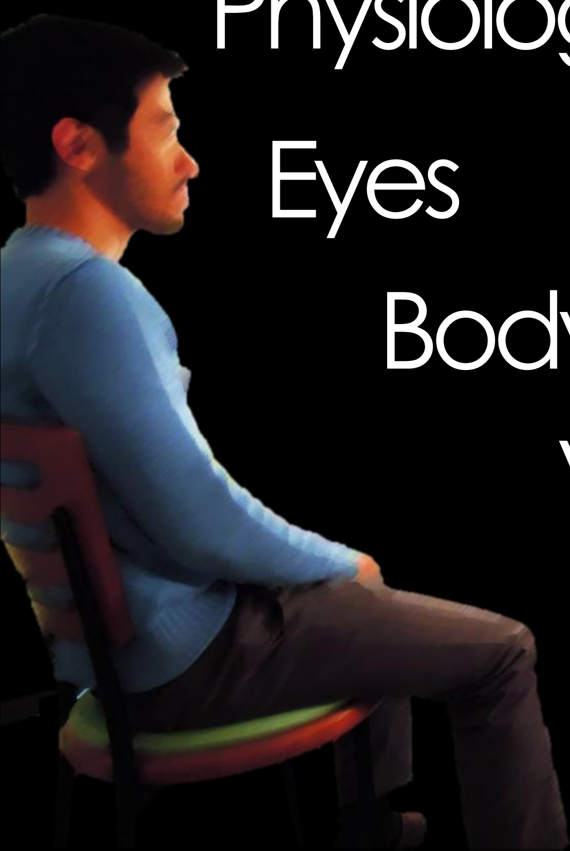


Micro-expressions

Gaze/Blinking/Dilation



Voice



Face

Physiology

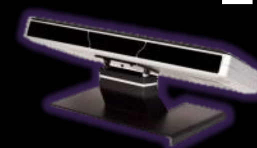
Eyes

Body

Voice

High Speed Camera
180fps , 640X480

Thermal Camera
60fps , 320X280

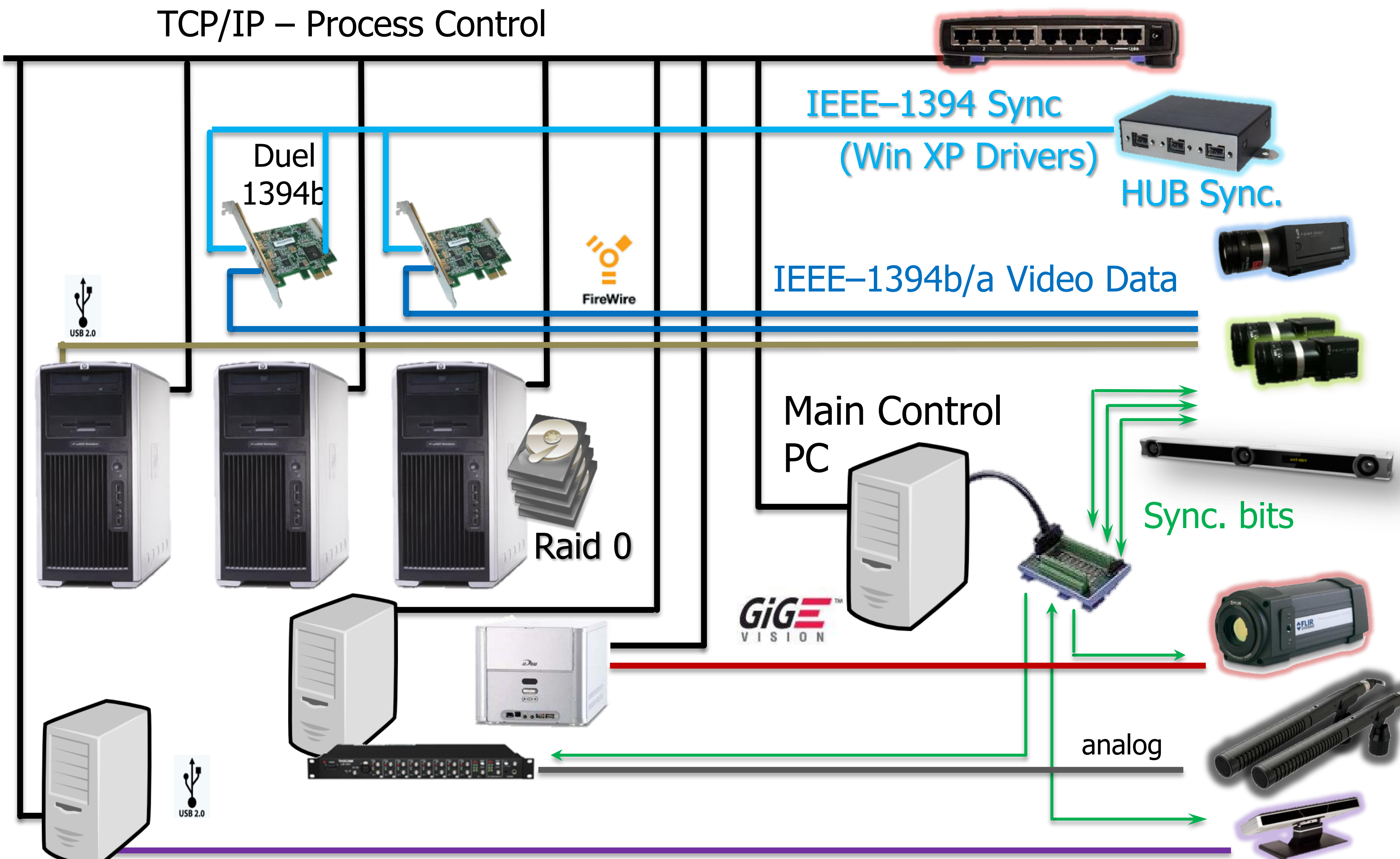


Eyes Tracking Sys.
300fps

High Resolution Camera
30fps, 1024x764

Directional Mic.
22050 Hz

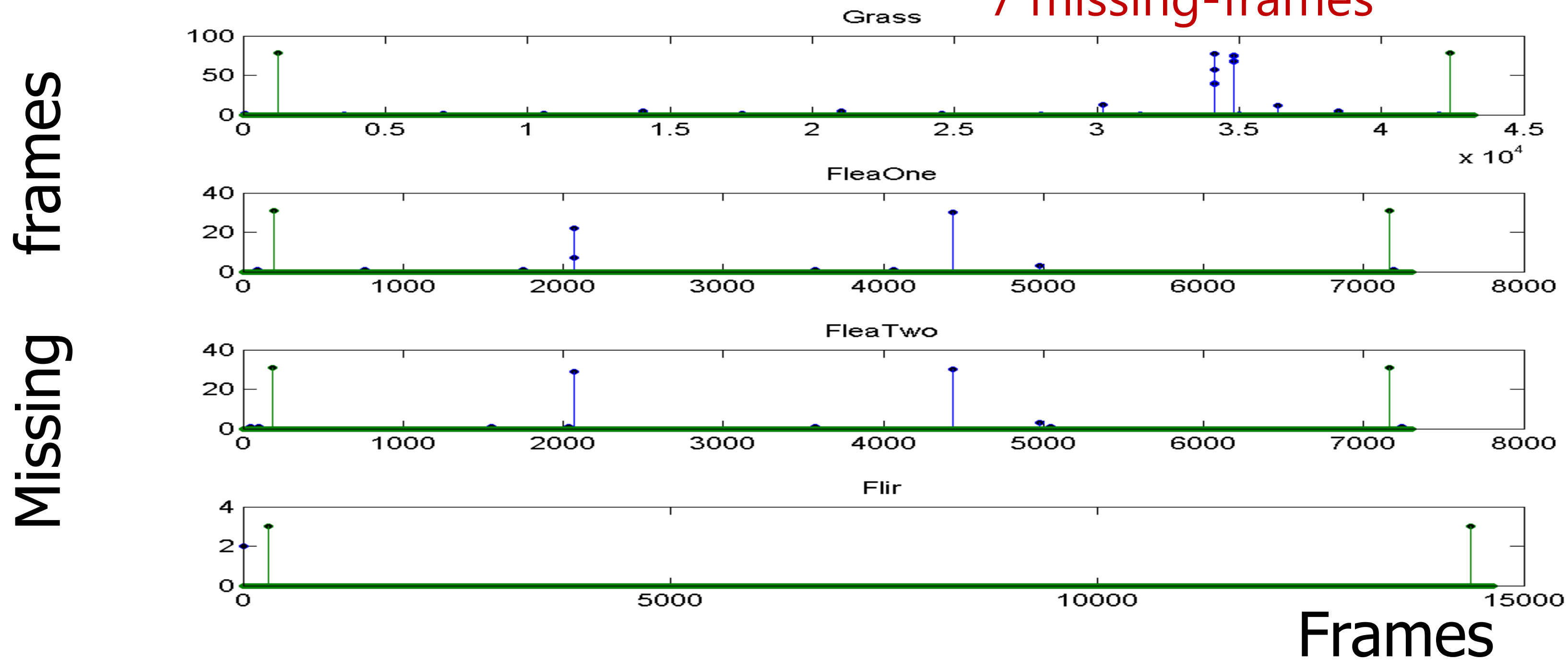
Capturing System Design



Missing Frames Detection based on Time-stamp and Frame ID

ID 100 ID 101 ID 102 ID 110

7 missing-frames



Excremental Setups

Setup 1

- Interview
- Posed expressions
- 10% side view

Setup 2

- Large screen
- During ICPR
- Concealed Emotion

Setup3

- External location

Layout of Experimental Setup 1