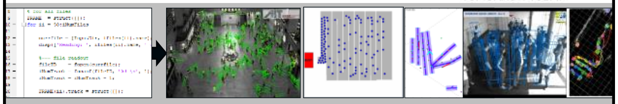


Challenging vision tasks meeting depth sensing: an in-depth look

Csaba Beleznai

Csaba Beleznai
Senior Scientist
Video- and Safety Technology
Safety & Security Department
AIT Austrian Institute of Technology GmbH
Vienna, Austria

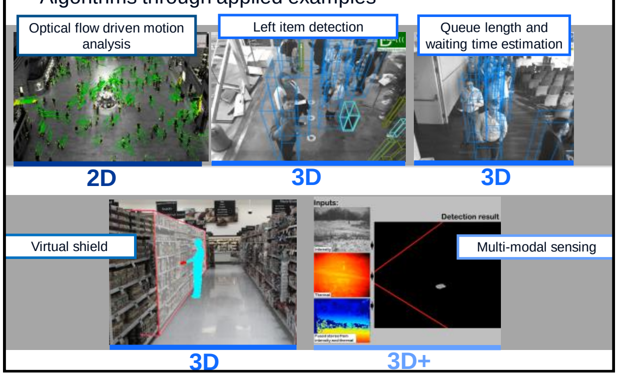
Michael Rauter, Christian Zinner, Andreas Zweng, Andreas Zoufal, Julia Simon, Daniel Steininger, Markus Hofstätter und Andreas Kriechbaum

Research pages: http://ivs.ait.ac.at

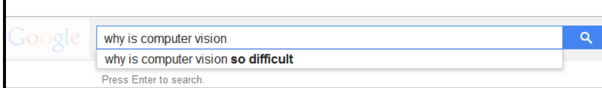
Austrian Institute of Technology

Contents

- Motivate & stimulate
- Algorithms through applied examples



A frequently asked question



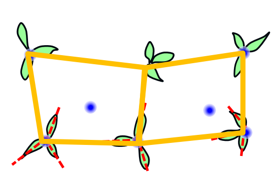
Why is Computer Vision difficult?

(from a Bayesian perspective)

- **Primary challenge in case of Vision Systems (incl. biological ones):**

? uncertainty/ ambiguity ?

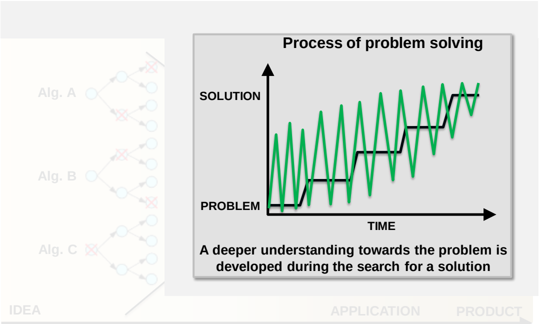
Example: Crop detection



- Radial symmetry
- Near regular structure

Motivation

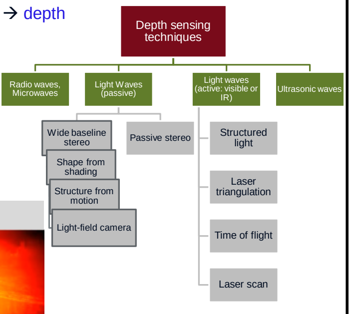
- **Challenges when developing Vision Systems:**
  - **Complexity ← Algorithmic, Systemic, Data**
  - **Non-linear search for a solution**

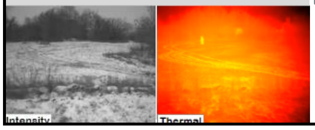


A deeper understanding towards the problem is developed during the search for a solution

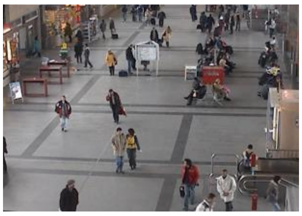
Depth sensing

- **Emerging trends in sensing other modalities**
  - Active (illumination) stereo → depth
    - Kinect (PrimeSense)
    - Time of Flight (TOF)
  - Passive stereo → depth
  - Light-field camera (Lytro, Raytrix) → depth
  - Thermal Infrared





Visual Surveillance - Motivating example



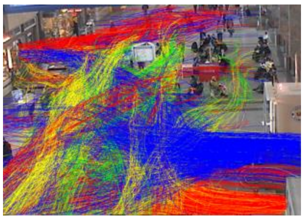
Algorithmic units:

- Object detection and classification
- Tracking
- Activity recognition

Typical surveillance scenario:

Who : people, vehicle, objects, …
Where is their location, movement?
What is the activity?
When does an action occur?

Visual Surveillance - Motivating example



Algorithmic units:

- Object detection and classification
  - Counting, Queue length, Density, Overcrowding
  - Abandoned objects
  - Intruders
- Tracking
  - Single objects
  - Video search
  - Flow
- Activity recognition
  - Near-field (articulation)
  - Far-field (motion path)

Typical surveillance scenario:

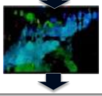
Who : people, vehicle, objects, …
Where is their location, movement?
What is the activity?
When does an action occur?

2D

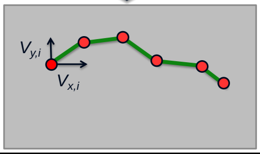
Real-time optical flow based particle advection

Optical flow driven advection

***Advection*: transport mechanism induced by a force field**



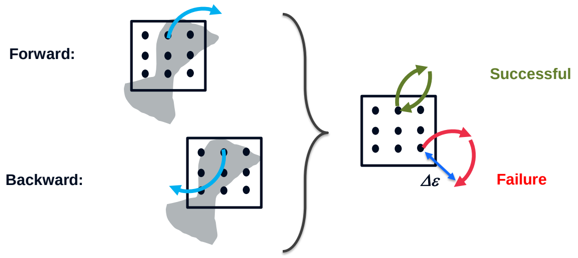
Dense optical flow field



A particle trajectory induced by the OF field

Particle advection with FW-BW consistency

- **A simple but powerful test**



Consistency check: $\Delta\varepsilon < \beta\,\overline{\Delta x}$ $\overline{\Delta x}$: **mean offset**

Pedestrian Flow Analysis

Public dataset: Grand Central Station, NYC: 720x480 pixels, 2000 particles, runs at 35 *fps*

Wide-area Flow Analysis

Other examples: wide area surveillance (small objects, nuisance, clutter)

2D + 3D

Left-item detection using depth and intensity information

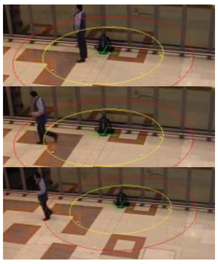
- **Composite task:**
  - **Static object detection**
  - **Human detection and tracking**

What is a static object?

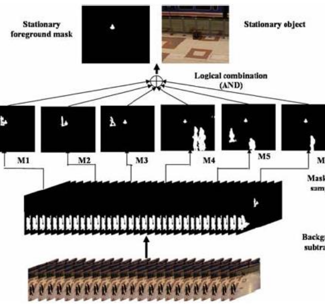
- “non-human” foreground which keeps still over a certain period of time
- Two fundamentally different approaches:

1. Background modeling (foreground regions becoming static)
   - +: simple, pixel-based
   - -: object removal, ghosts
2. Tracking detected foreground regions
   - +: many adequate tracking approaches (blob-based, correlation-based)
   - -: crowd, occlusion → failure

Both techniques experience problems with illumination variations → motivation for depth-based sensing



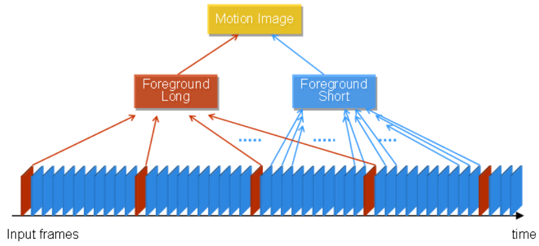
Method 1: single background models + sub-sampling of FG results



Sub-sampling and combination procedure

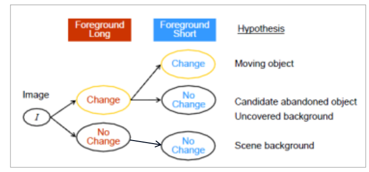
Liao,H-H.; Chang,J-Y.; Chen, L-G. *"A localized Approach to abandoned luggage detection with Foreground–Mask sampling"*, Proc. of AVSS 2008, pp. 132-139.

Method 2: computing two background models at two different framerates

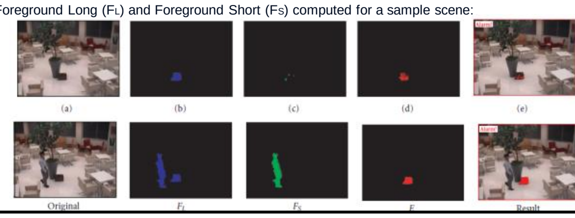


Porikli, F.; Ivanov, Y.; Haga, T. "Robust Abandoned Object Detection Using Dual Foregrounds", Journal on Advances in Signal Processing, art. 30, 11 pp., 2008.

Method 2: computing two background models at two different framerates



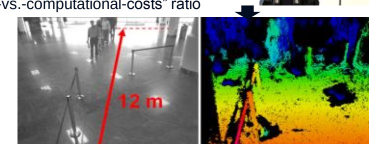
Foreground Long (F_L) and Foreground Short (F_S) computed for a sample scene:



Obtaining stereo depth information

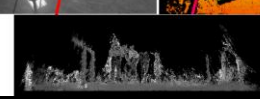
Passive stereo based depth measurement

- 3D stereo-camera system developed by AIT
  - Area-based, local-optimizing, correlation-based stereo matching algorithm
  - Specialized variant of the Census Transform
  - Resolution: typically ~1 Mpixel
  - Run-time: ~ 14 fps (Core-i7, multithreaded, SSE-optimized)
  - Excellent "depth-quality-vs.-computational-costs" ratio
  - USB 2 interface



Advantage:

- Depth ordering of people
- Robustness against illumination, shadows,
- Enables scene analysis



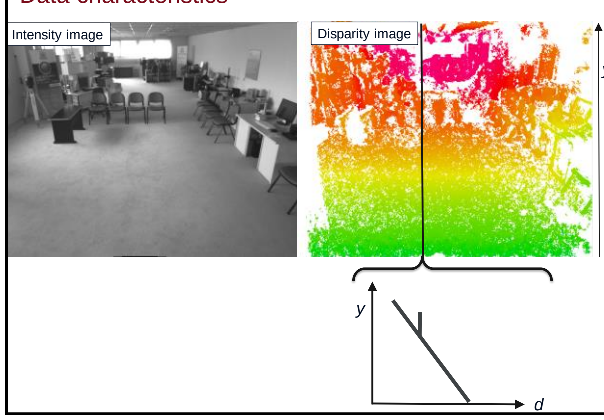
Stereo camera characteristics

Trinocular setup:

- 3 baselines possible
- 3 stereo computations with results fused into one disparity image

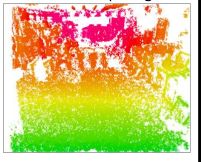
small medium

Data characteristics



2.5D vs. 3D algorithmic approaches

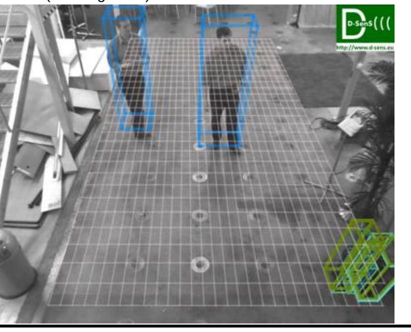
2.5D == using disparity as an intensity image



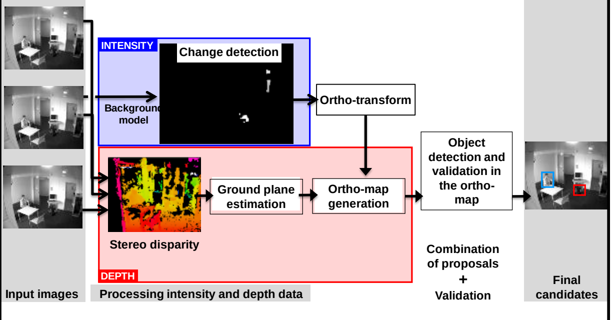
Left Item Detection

Additional knowledge (compared to existing video analytics solutions):

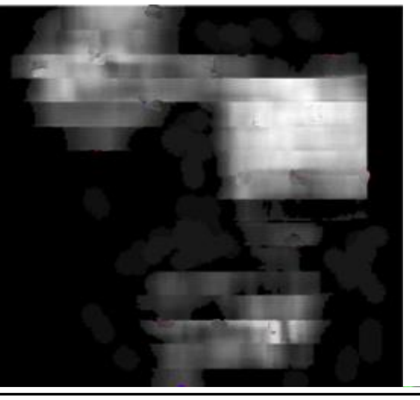
- Stationary object (Geometry introduced to a scene)
- Object geometric properties (Volume, Size)
- Spatial location (on the ground)



Methodology

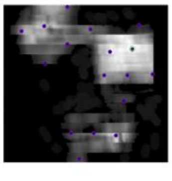


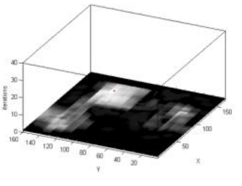
Human detection as clustering



Scale-adaptive clustering using a stability criterion (1)

- Evolution of cluster centroid coordinates during CamShift

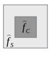




Algorithmic steps:

(1) Computation of integral images: $ii_{00}(x,y) \quad ii_{02}(x,y)$

(2) Locating the sample set: $\hat{f}_s / \hat{f}_c < T$



center surround feature

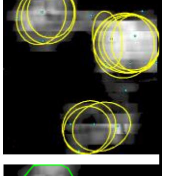
Scale-adaptive clustering using a stability criterion (2)

3. CamShift clustering iterations

- Stopping criterion:

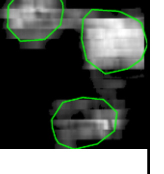
Area of elliptic region: $A_j^i = L_1 \, L_2$

Rate of area change: $\gamma(i) = \left| A_j^{i-\delta} - A_j^{i+\delta} \right| \,/\, A_j^i$



4. Spatial grouping of elliptic clusters

- Subsampling ellipse contours ➜ piecewise linear boundaries
- Pairwise check for overlap ➜ convex hull approximation (fast BFP algorithm)



Left Item Detection – Demos

Left Item Detection – Interesting cases

Object form Transferred objects

Quantitative evaluation

Detection results | Ground truth | Depth-based proposals | Motion-based proposals

2D + 3D

Queue length detection using depth and intensity information

Queue Length + Waiting Time estimation

What is waiting time in a queue?

Time measurement relating to last passenger in the queue

Why interesting?

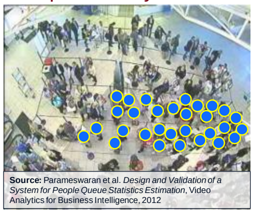
Example: Announcement of waiting times (App) → customer satisfaction
Example: Infrastructure operator → load balancing

Queue analysis

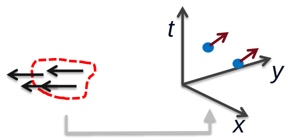
- Challenging problem

Waiting time = $\frac{Length}{Velocity}$ → 1. What is the shape and extent of the queue? → 2. What is the velocity of the propagation?

Visual queue analysis - Overview



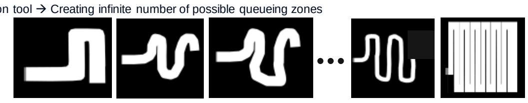
How can we detect (weak) correlation?



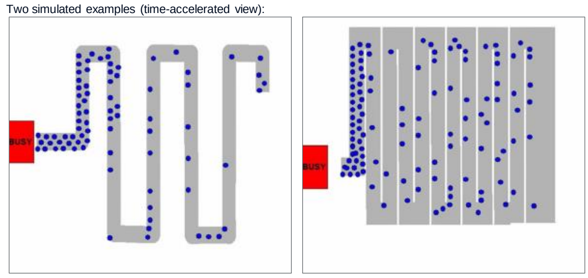
- Much data is necessary → Simulating crowding phenomena in Matlab
  - *Social force model* (Helbing 1998)

Queue analysis

Simulation tool → Creating infinite number of possible queueing zones



Two simulated examples (time-accelerated view):



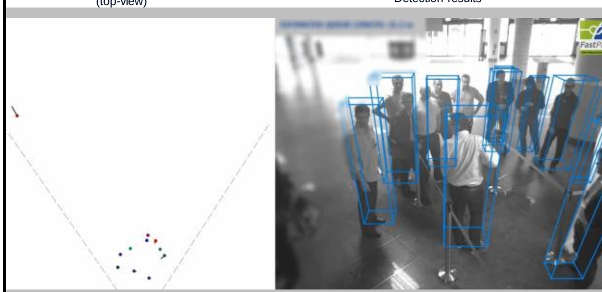
Queue analysis (length, dynamics)

Straight line | Meander style

Staged scenarios, 1280x1024 pixels, computational speed: 6 *fps*

Adaptive estimation of the spatial extent of the queueing zone

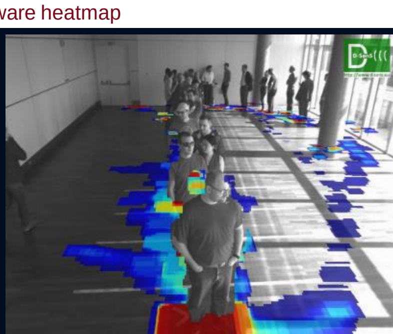
Estimated configuration (top-view) | Detection results



Left part of the image is intentionally blurred for protecting the privacy of by-standers, who were not part of the experimental setup.

stereo sensor

Scene-aware heatmap

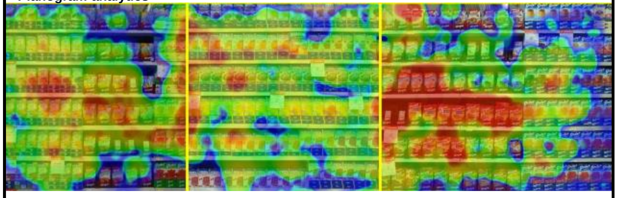


Virtual Shield

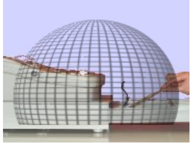
Virtual shield

Virtual shield

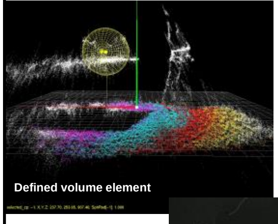
Planogram analytics



Protecting assets



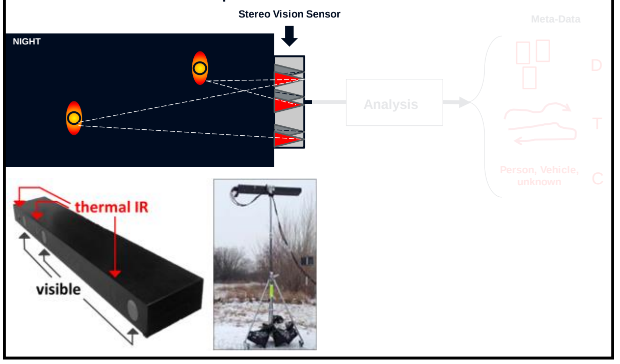
Virtual shield



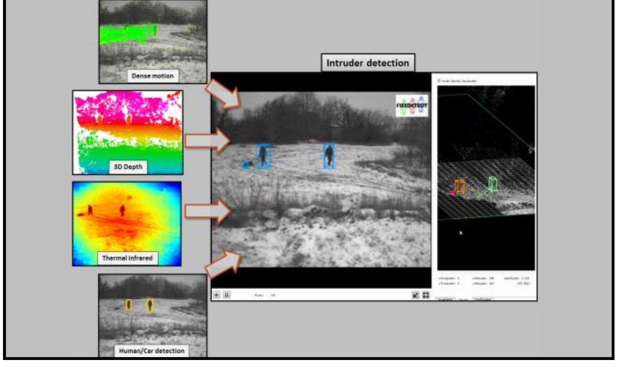
Multi-modal sensing

Objective

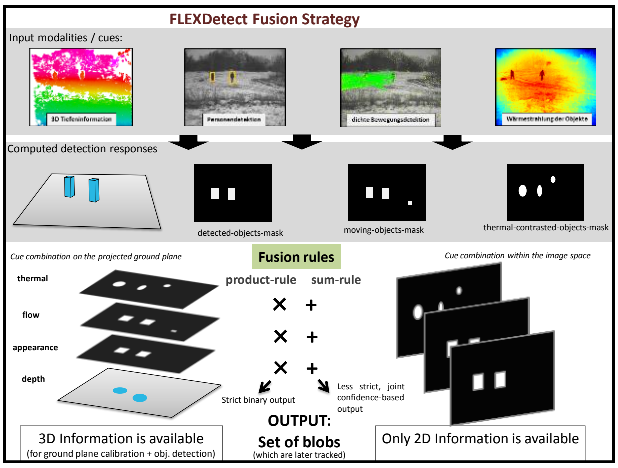
- **Robust detection independent from environmental conditions**



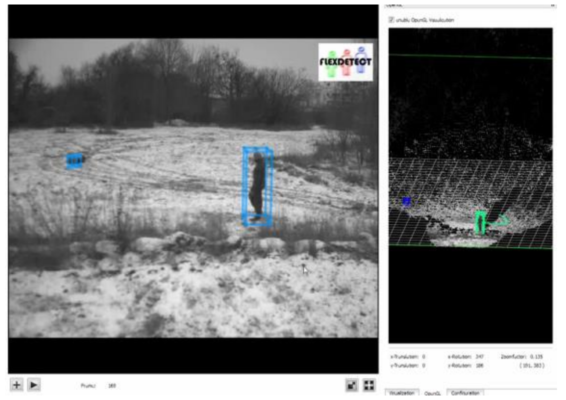
Multi-modal object detection



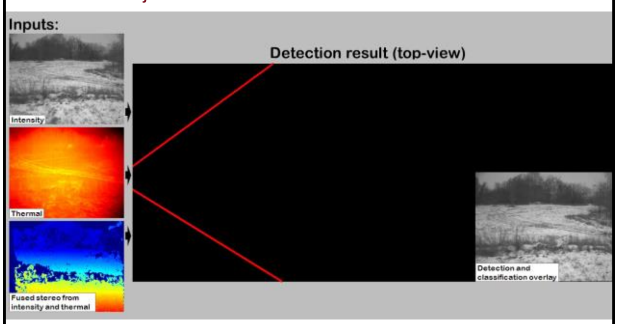
FLEXDetect Fusion Strategy



Result



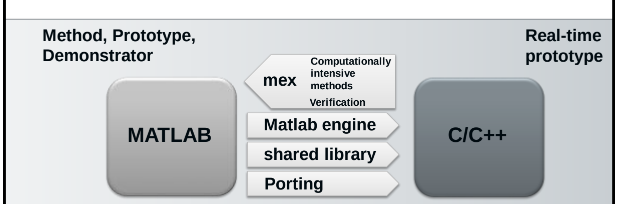
Multi-modal object detection



Implementation details and strategy

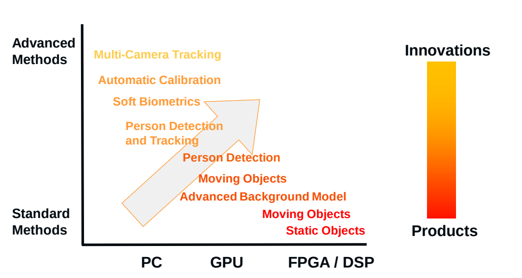
Implementation details

Our development concept



- MATLAB:
  - Broad spectrum of algorithmic libraries,
  - Well-suited for image analysis,
  - Visualisation, debugging,
  - Rapid development → Method, Prototype, Demonstrator
- C/C++
  - Real-time capability

Our development concept



Thank you for your attention!

CSABA BELEZNAI
Senior Scientist
Safety & Security Department
Video- and Security Technology

AIT Austrian Institute of Technology GmbH
Donau-City-Straße 1 | 1220 Vienna | Austria
T +43(0) 664 825 1257 | F +43(0) 50550-4170
csaba.beleznai@ait.ac.at | http://www.ait.ac.at

http://www.d-sens.eu/