Multi-modal Multi-person Detection and Tracking based on Probabilistic Techniques

By: Hadi Ali Akbarpour

Supervisor: Professor Jorge Manuel Miranda Dias

Electrical and Computer Engineering Department, University of Coimbra

Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

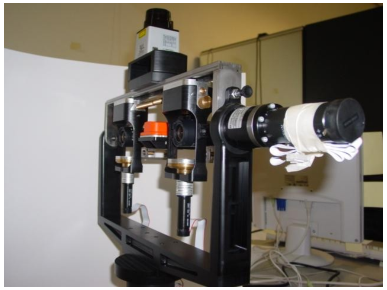
The Outlines

- Using a dispersed network of sensors
- Using several modalities, in addition to use a dispersed network
- Dealing with data registration
- Using two kinds of sensors arrangements: *Static* and *Moving* sensors)
  (By using a mobile robot in addition to the fixed sensors)
- Human Presence Detection & Tracking
- Using Bayesian techniques

The Signal Acquisition Devices

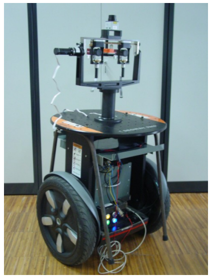
Video camera

Laser Range Finder



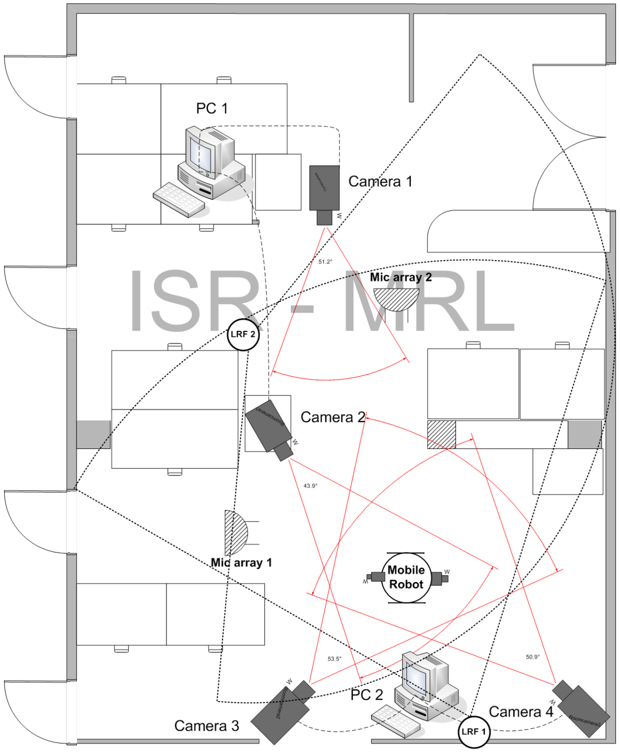
A pan-tilt head

Microphone

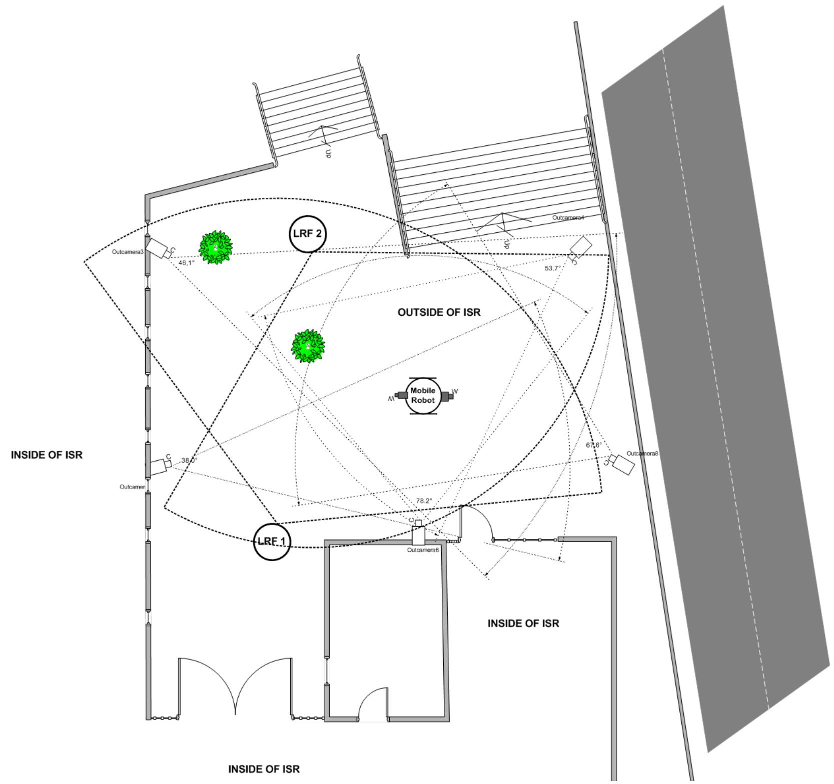


Mobile Robot

Some proposed scenes and their configurations



Indoor



Outdoor

Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

The motivation of using a probabilistic framework

- For inference:

Likelihood — previous knowledge

$$p(state_i|sensoryinput, I) = \frac{\boxed{p(sensoryinput|state_i, I)}\,\boxed{p(state_i|I)}}{\boxed{\int_S p(sensoryinput|\theta, I).p(\theta|I).d\theta}}$$

Normalizing factor

- For filtering:

model of system

$$p(x_t|o_{0..t}) = \underline{\alpha}\, \underline{p(o_t|x_t)} \int \underline{p(x_t|x_{t-1})}\, \underline{p(x_{t-1}|o_{0..t-1})} dx_{t-1}$$

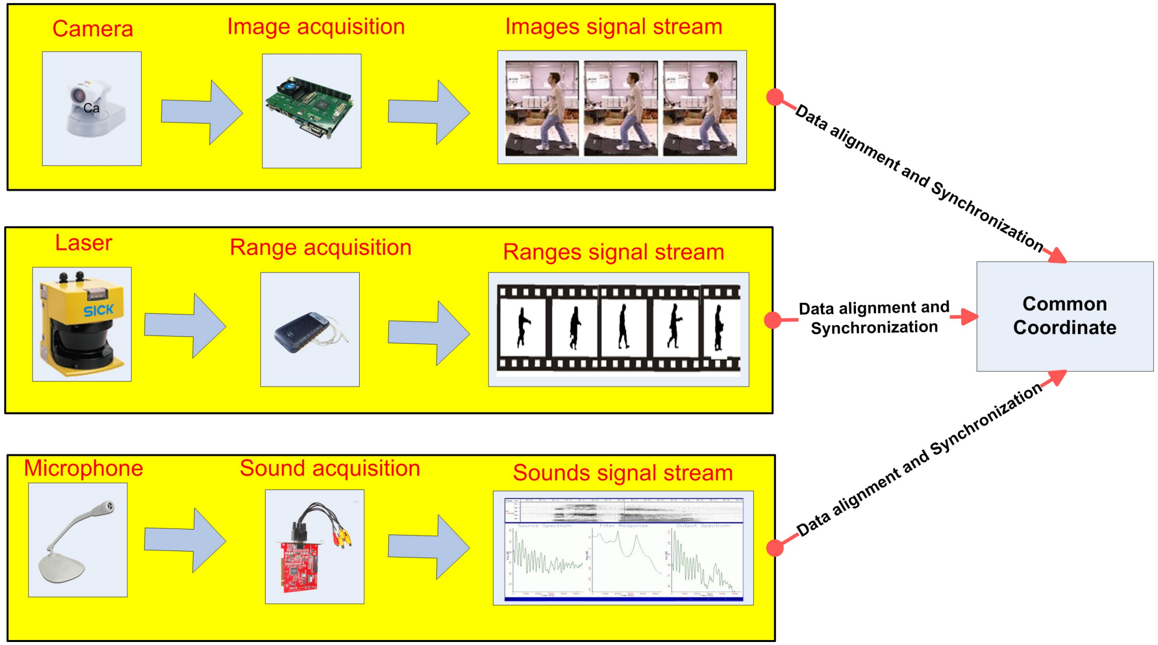
Normalizing factor

Likelihood
(sensor model)

prior distribution

Data registration

- Spatial Data Registration
- temporal Data Registration



Spatial Data Registration

- Commensurate Sensors
- Non-Commensurate Sensors

Step 1- Calibrating the camera network

Step 2- Calibrating the camera network and LRF

Step 3- Calibrating the camera network and Mic. array

Having all sensors calibrated !

Our Setup for Camera Network Calibration Using Svoboda Toolbox

Camera No. 1

Camera No. 2

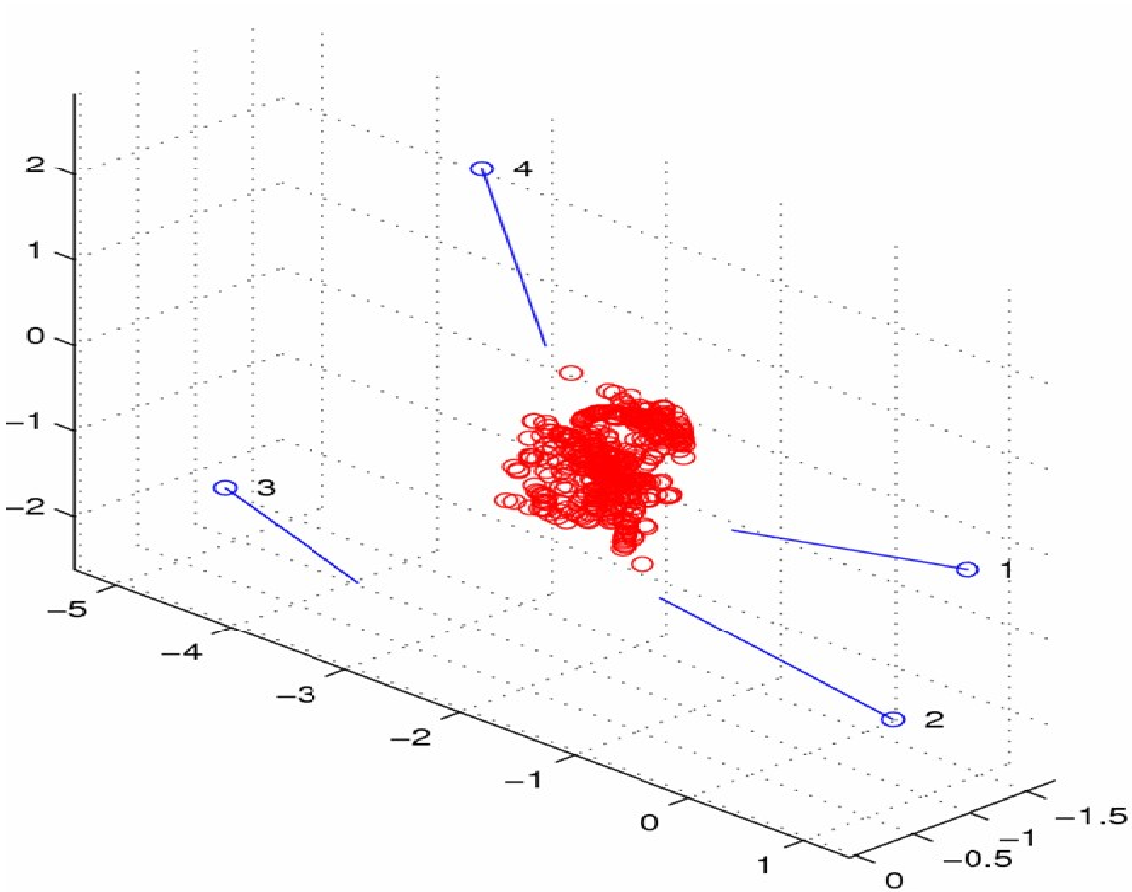
Camera No. 3

Camera No. 4

The scene

Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

Cameras and Reconstructed Points by the Svoboda Calibration Toolbox



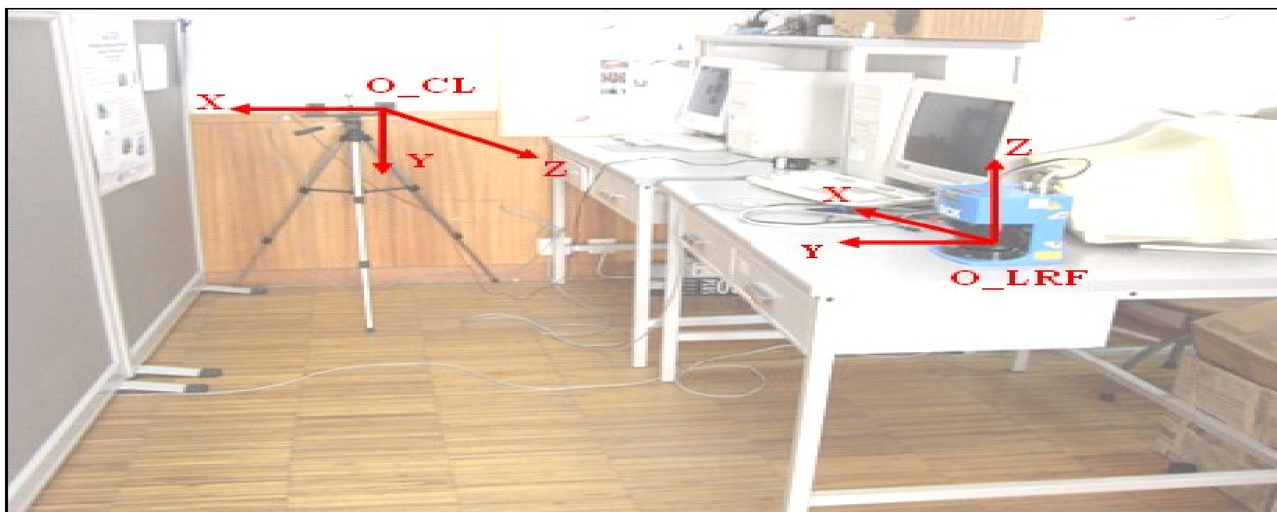
Extrinsic Calibration of Stereo Camera and 2D LRF

Configuration

$$P^{C_L} = R * P^{LRF} + t$$

→ Rigid Coordinate Transformation

Coordinate systems convention



Calibration Steps

- **Step 1**: Placing the laser pointer, somewhere out of the LRF´s FOV which is planar.
- **Step 2**: The LRF starts to capture range signals form the area of its view.
- **Step 3**: The laser pointer has to be moved to hit the sensing plane of the LRF. (The times of these hitting must be registered)
- **Step 4**: The stereo camera has to capture and register the images at the hitting times of the step 3.
- **Step 5**: Repeating the steps 1 to 4, about 20 times, meanwhile that all of sensors are capturing the data.
- **Step 6**: Extracting the laser pointer positions in the recorded frames by the stereo cameras and converting each paired 2D point to a single 3D point (a real world Cartesian coordinate).
- **Step 7**: Converting the data recorded by LRF module (at the hitting moments) into a Cartesian coordinate, the same as step 6.
- **Step 8**: Feeding the outputs of steps 6 and 7 to the GetProcrustes algorithm (which is described further) and get outputs. The outputs will be a rotation matrix and a translation vector between the stereo camera and the LRF coordinates.

A sample between 20 (N) recorded samples (synchronized)

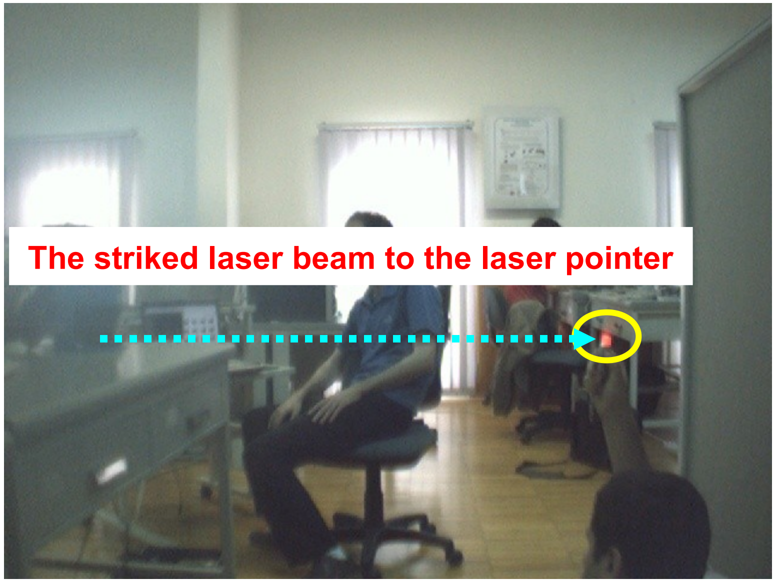


Image captured by the left camera.
The position of the laser pointer (red light) in this frame is (501,274) in pixel.

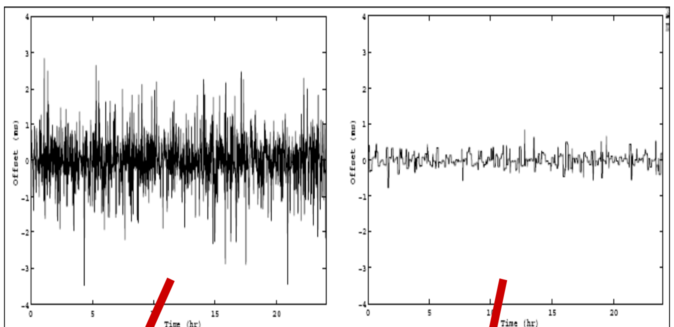
Image captured by the right camera.
The position of the laser pointer (red light) in this frame is (434,281) in pixel.

The LRF reported it as 133 cm and 62 degree (62.44,117.43) cm in its local frame

Temporal Data Registration

1- Synchronizing all PCs based on Network Time Protocol (NTP)

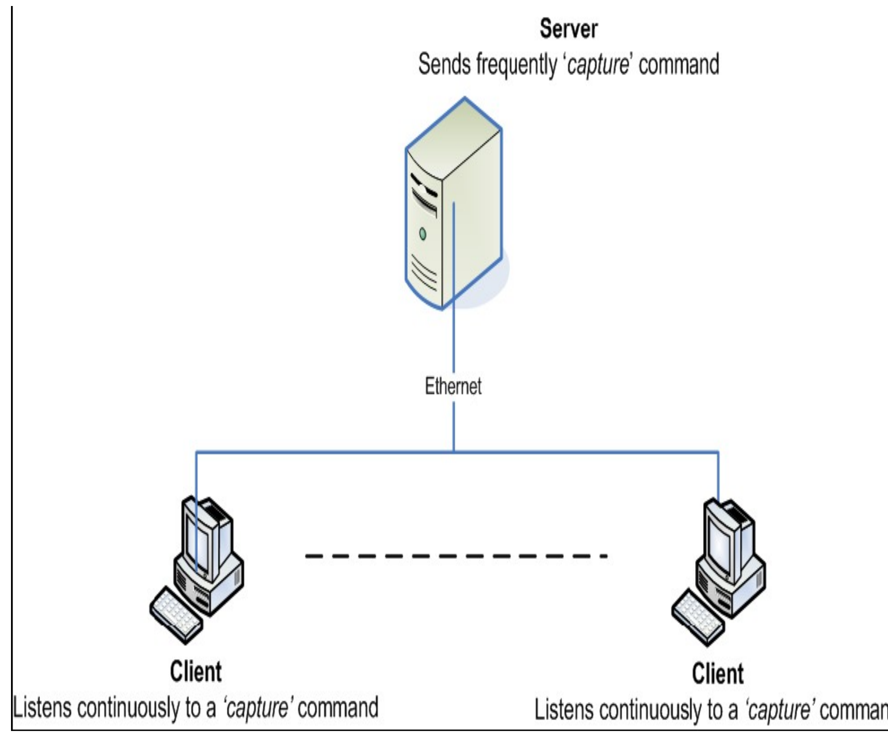
Time offsets between two PCs



Without using NTP

Using NTP

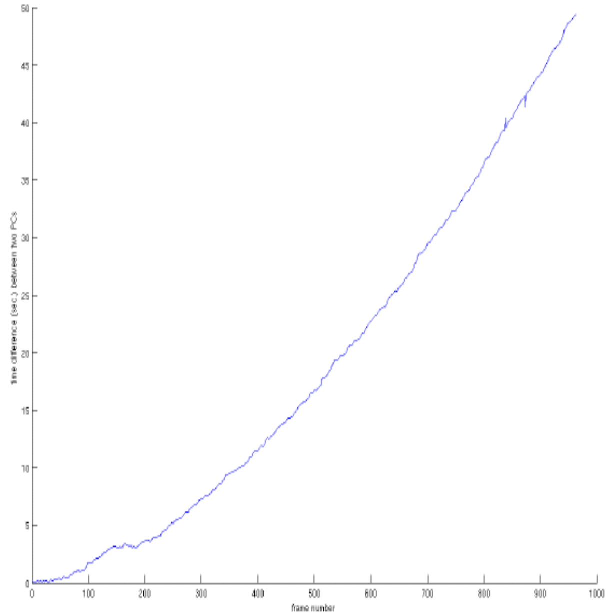
2- Using a Network-Socket based data capturing architecture

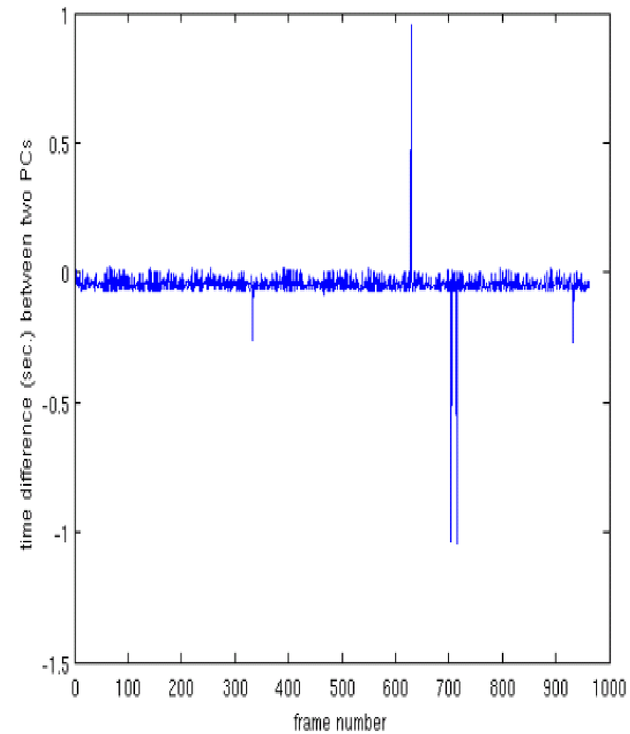


Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

Temporal Data Registration

Time delays between recorded images in a camera network

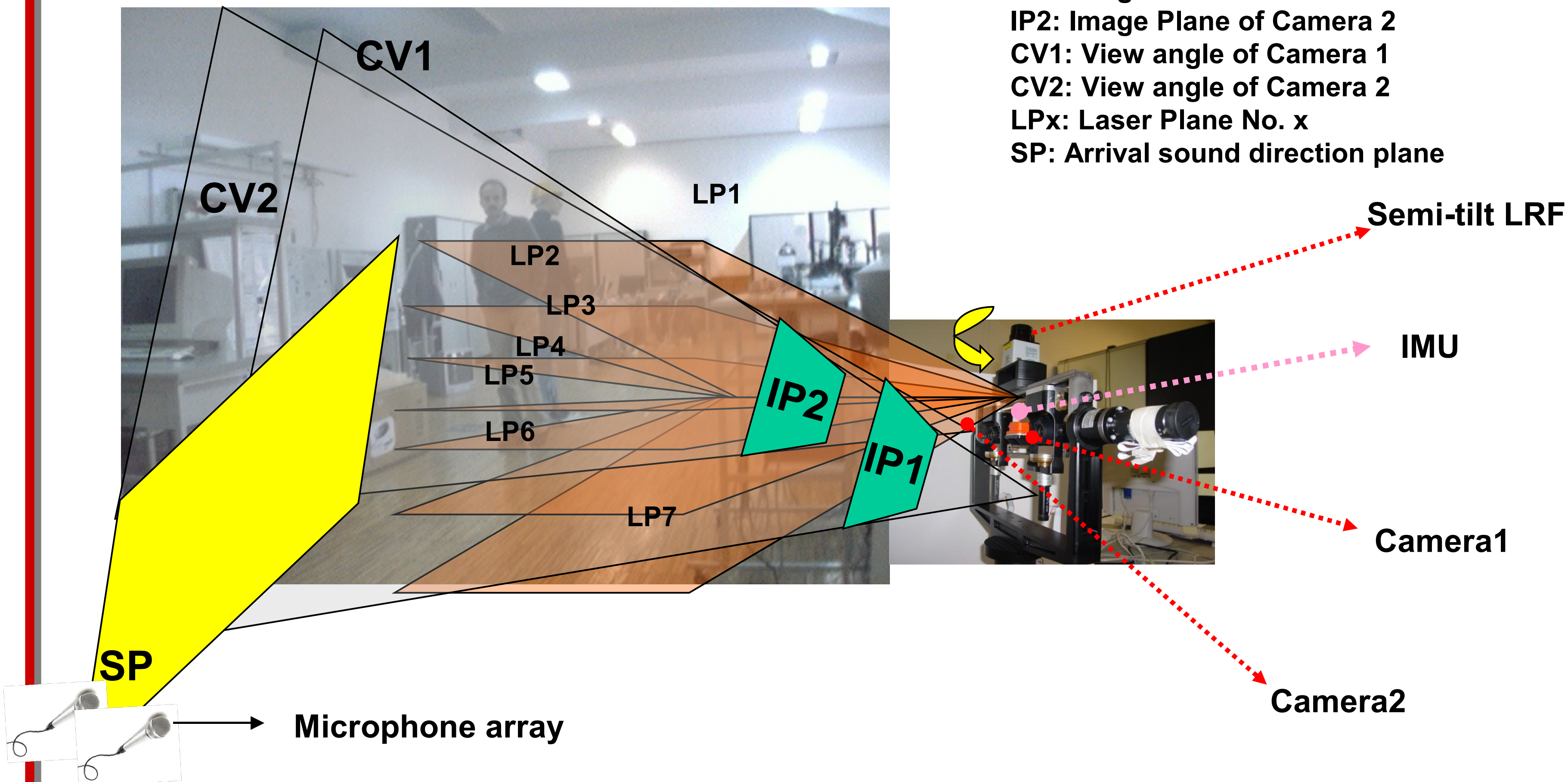




The left figure shows the delays between recorded frames when just the starting time is the same. The right one shows is for the same subject when the described architecture is used. The values are in ms.

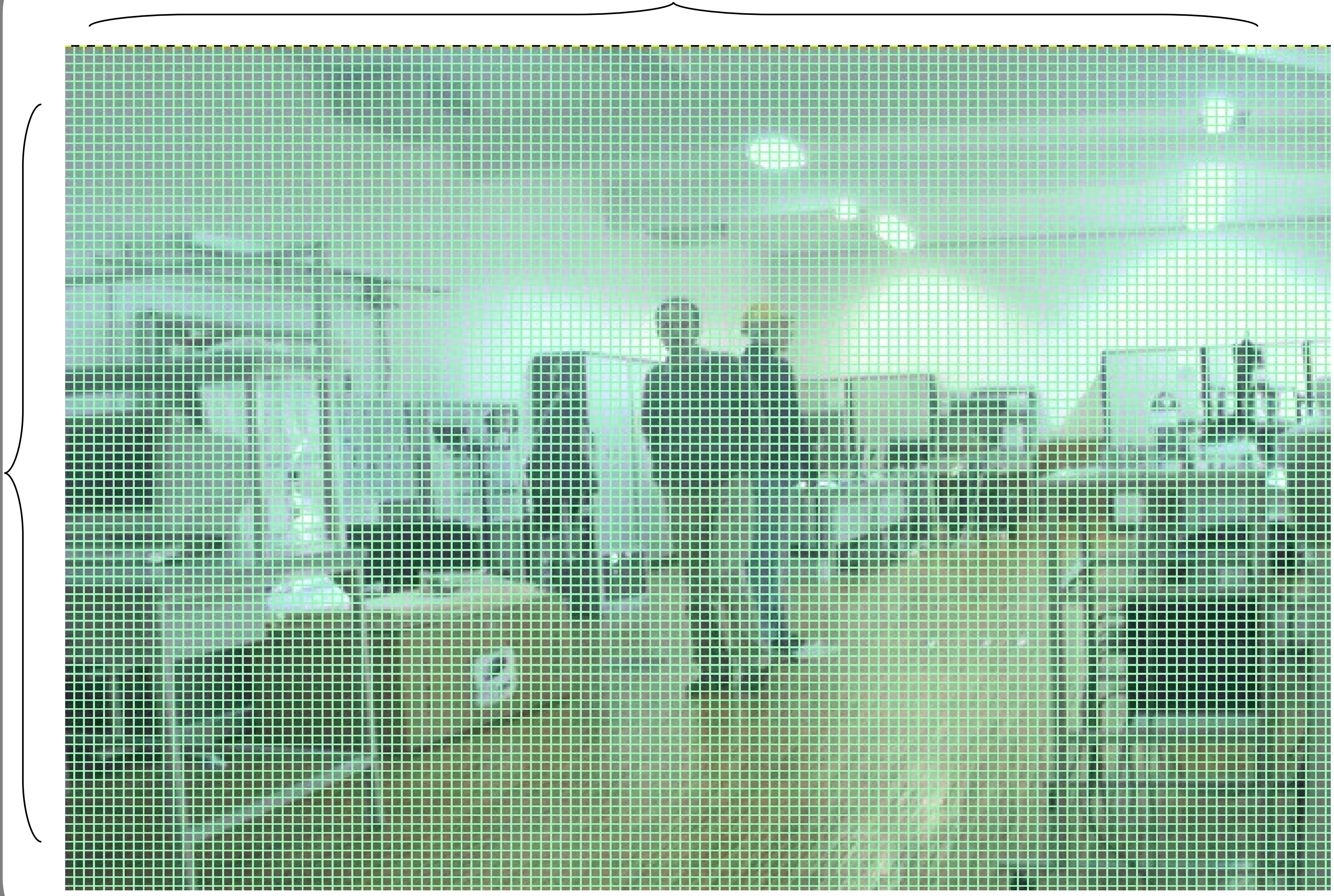
Constructing a Probabilistic Distance-Color Grid

IP1: Image Plane of Camera 1
IP2: Image Plane of Camera 2
CV1: View angle of Camera 1
CV2: View angle of Camera 2
LPx: Laser Plane No. x
SP: Arrival sound direction plane

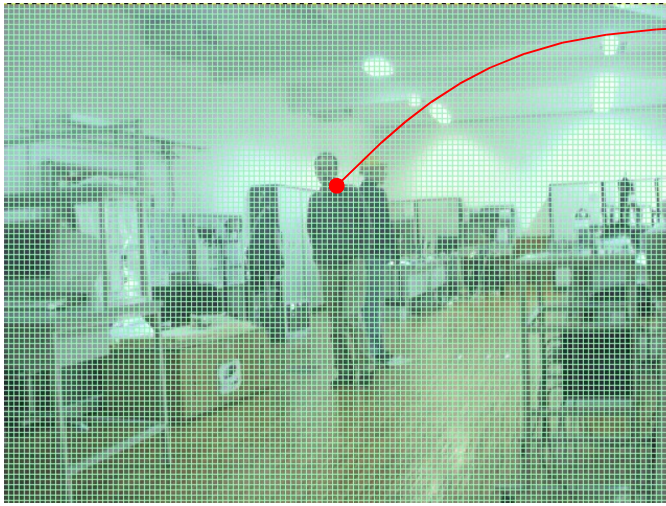


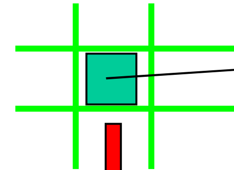
Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

A 2D grid with m x n cells



Probabilistic Parameters of the Grid





Each cell of the grid

Parameters and random variables

$\{i, j\}$ → **Position of the cell: (row and column)**

$\{\rho_{SC}, \rho_{LRF}, \rho_{M}\}$ → **Distance to the camera2 image center, in a real world scale (e.g. mm), measured by Stereo Camera, LRF and Microphones, respectively.**

$\{p(\rho_{SC}), p(\rho_{LRF}), p(\rho_{M})\}$ → **Probability (confidence degree) of the distances**

c → **Color of the cell**

$p(c \in \text{skin color})$ → **Probability of skin-color being of the color**

$p(c \in background)$ → **Probability of being background**

md → **Movement direction**

$p(md)$ → **Probability (confidence degree) of the md**

Domains of the parameters and random variables

$$i \in \{1..m\}$$

$$j \in \{1..n\}$$

$$\rho_{SC}, \rho_{LRF}, \rho_{M} \in \Re^{+}$$
$$p(\rho_{SC}), p(\rho_{M}) \in [0,1]$$

$$p(\rho_{LRF}) \in \{0,1\}$$

$$c \in \text{ color space}$$
$$p(c \in \text{ skin\_color}) \in [0,1]$$

$$p(c \in \text{ background\_being}) \in [0,1]$$
$$\text{md} \in \{left, left-top, left-down, top, down, right, right-down, right-up\}$$

$$p(md) \in [0,1]$$

Sources of information to fill each cell of the grid

Parameters of each cell is determined by fusing data from either all or some of the following sensors:

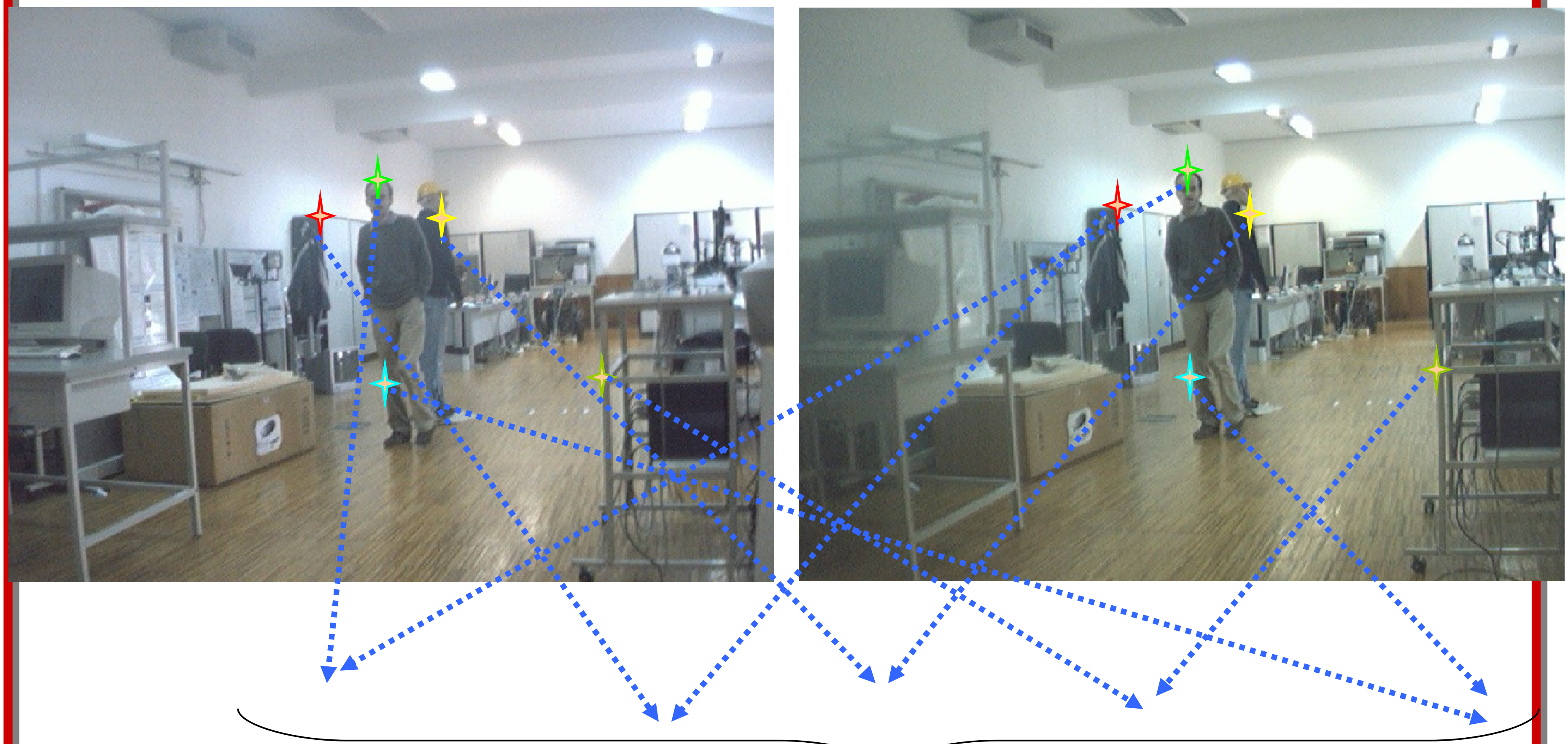
1- Stereo camera system

2- Semi-tilt Laser Range Finder

3- Microphones

1st Source: Using Stereo camera system

a) Finding some strong features using e.g. SIFT
b) Finding the distances (in real world scale e.g. mm) by using matching and then triangulation (considering that the stereo camera is calibrated)



Result of this phase: **Distances of matched points**

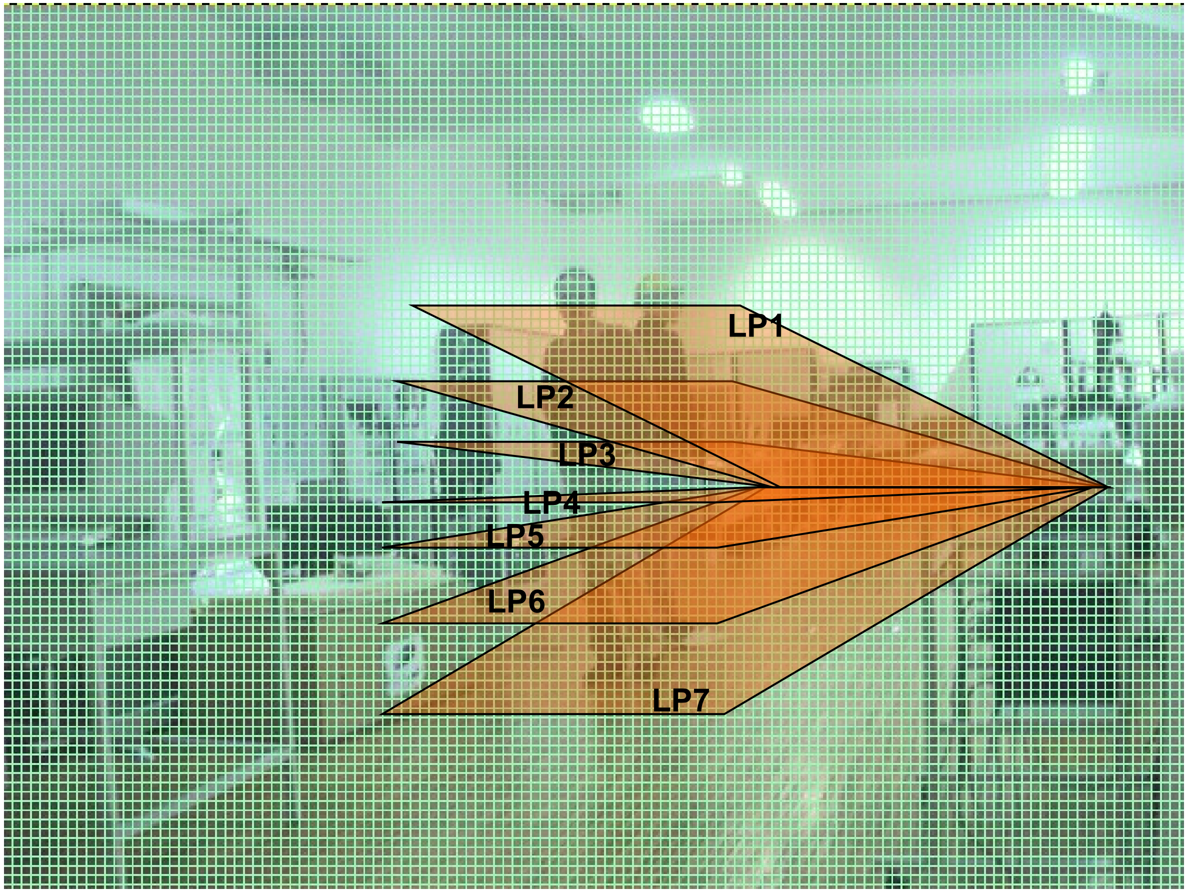
2nd Source: *Using tilt Laser Range Finder*

It can be done in two phases:

a) Finding intersection between laser planes and camera2's plane.

a) Transforming each seen point by LRF into camera2's plane.

a) Finding intersection between laser planes and camera2 plane



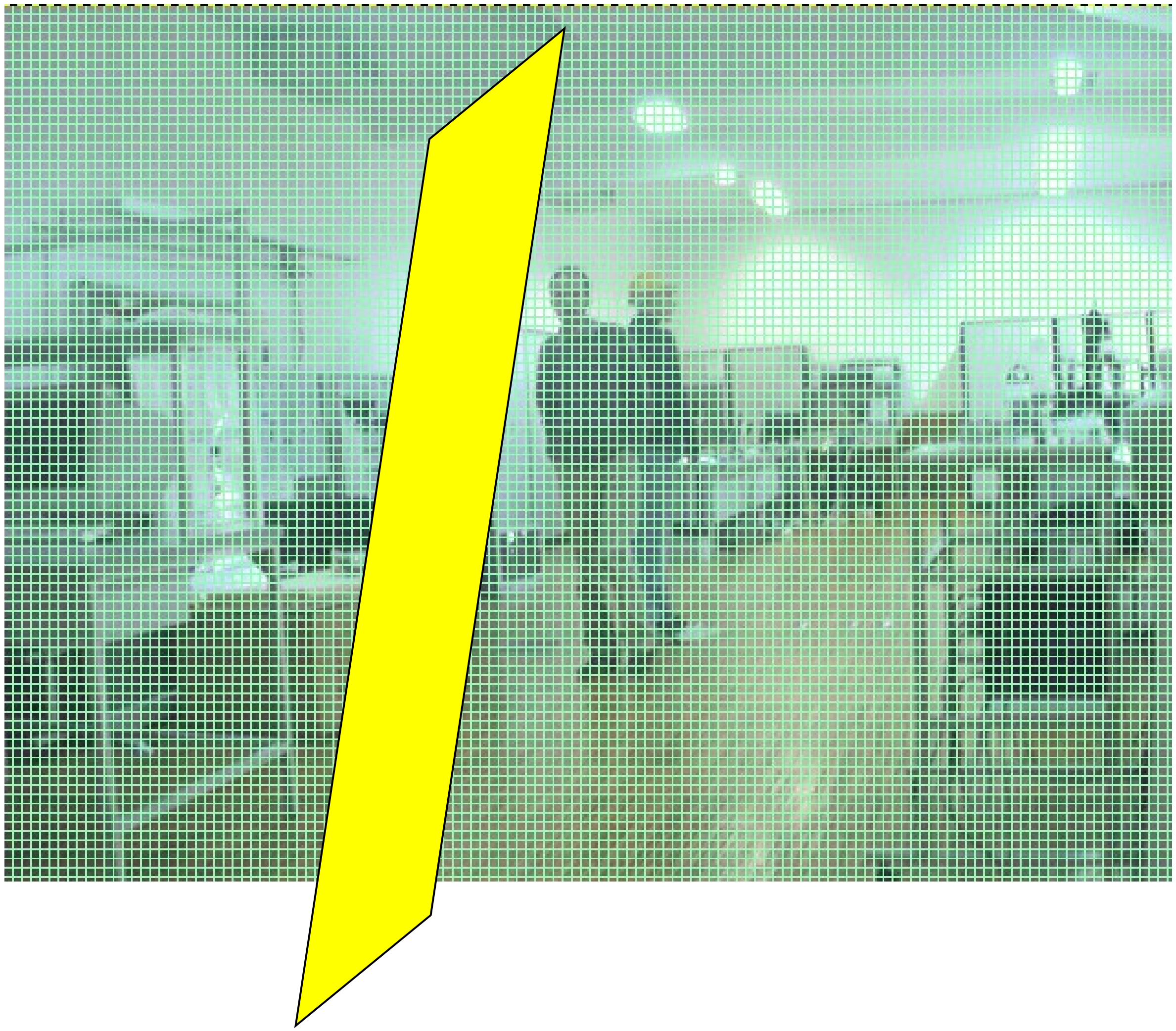
intersection between laser planes and camera2 plane (lines)

b) Transforming the range data in intersection lines into the probabilistic grad

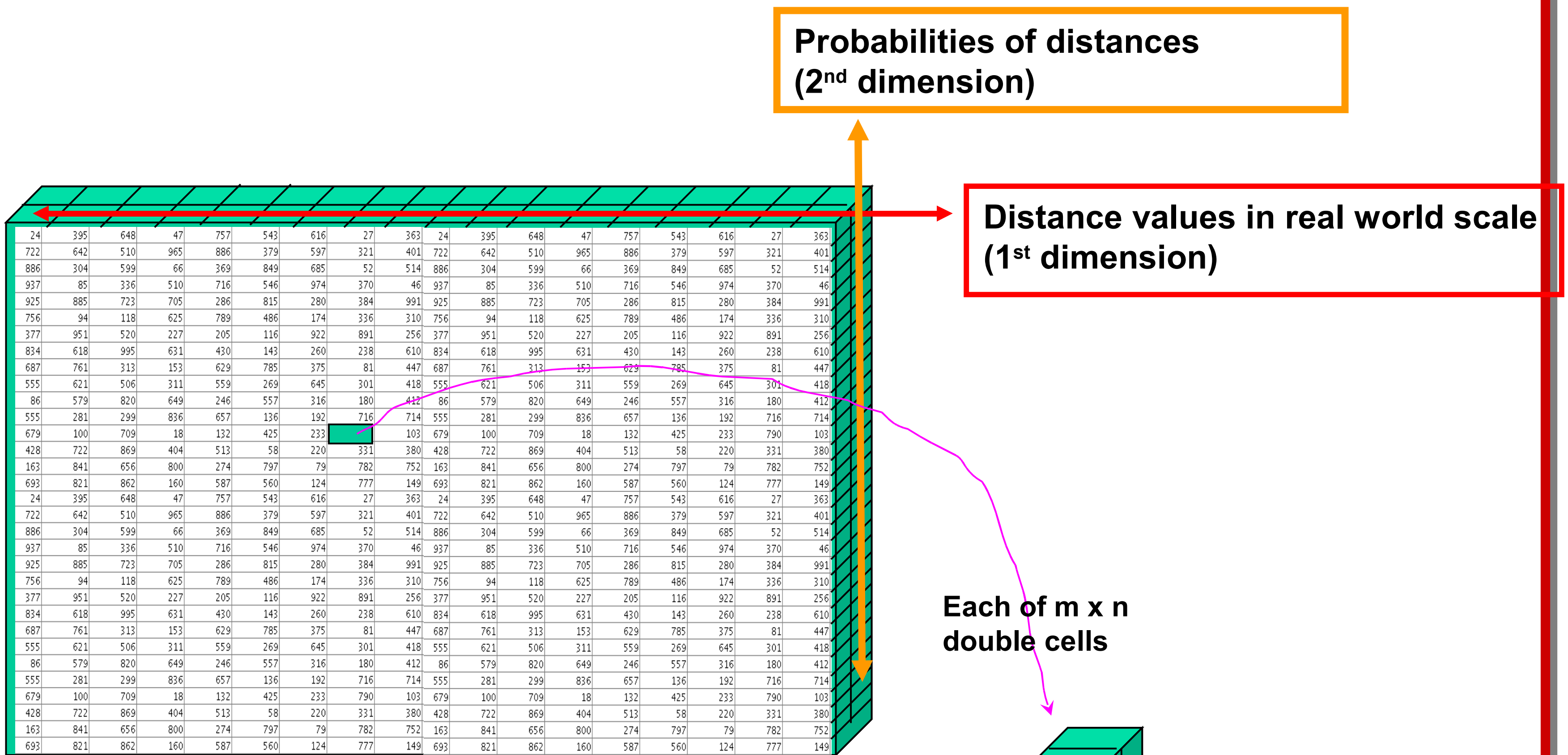
Note: These points have a confidence probability near unit

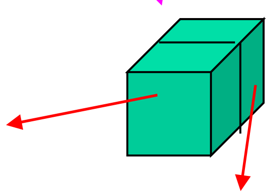
Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

3nd Source: Microphones



Probabilistic Information Fusion on Range, Image *and* Sound *Data*



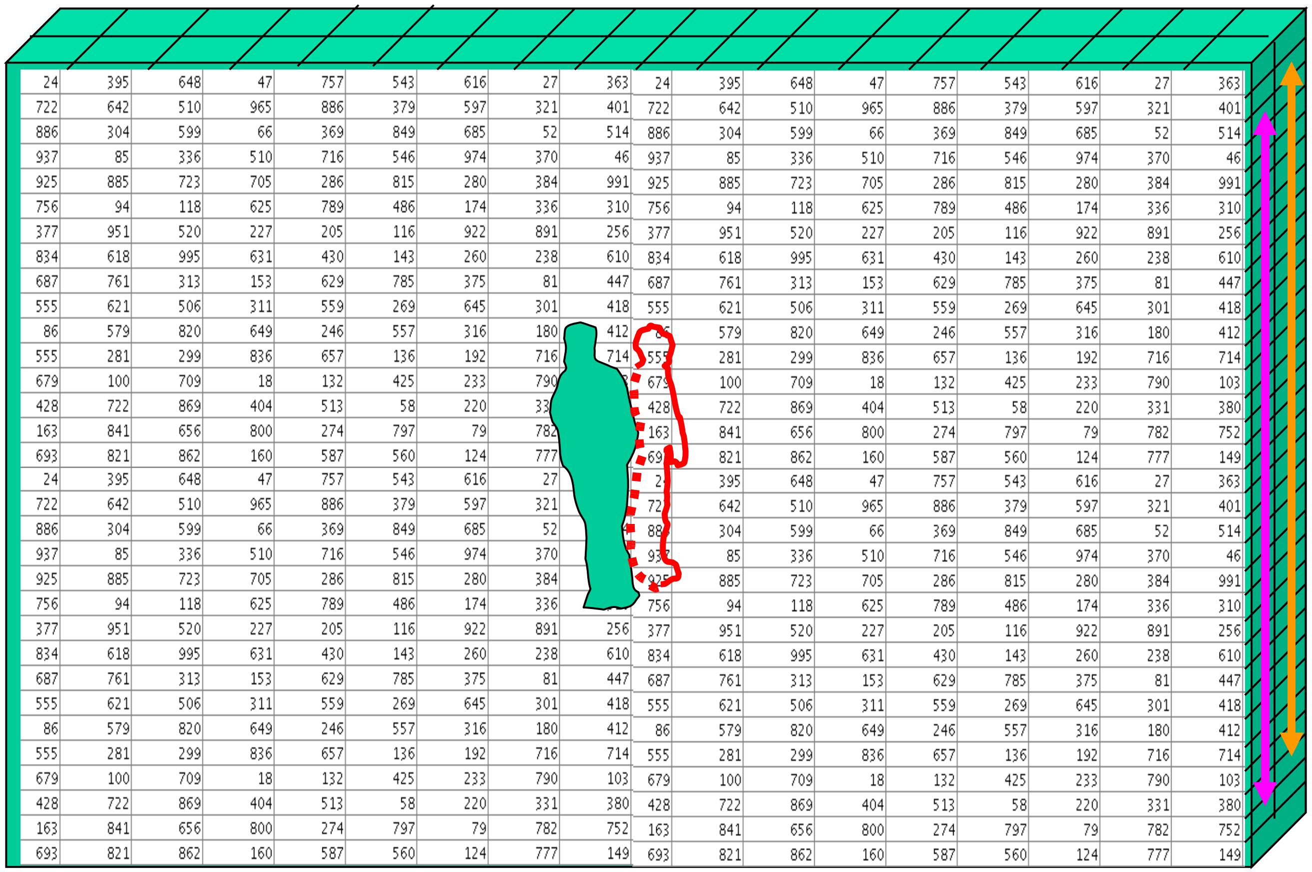


Final distance of the cell: $\{\rho_{SC}, \rho_{LRF}, \rho_{M}\} \xrightarrow{fusion} \rho_f$

Final probability of the cell: $\{p(\rho_{SC}), p(\rho_{LRF}), p(\rho_{M})\} \xrightarrow{fusion} \Rightarrow p(\rho_f)$

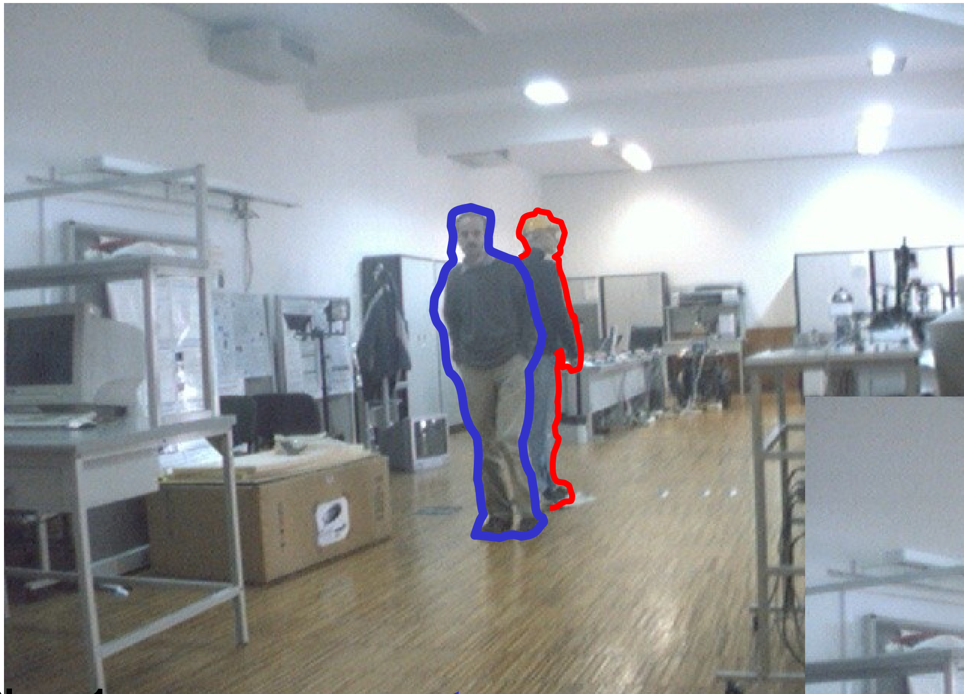
Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

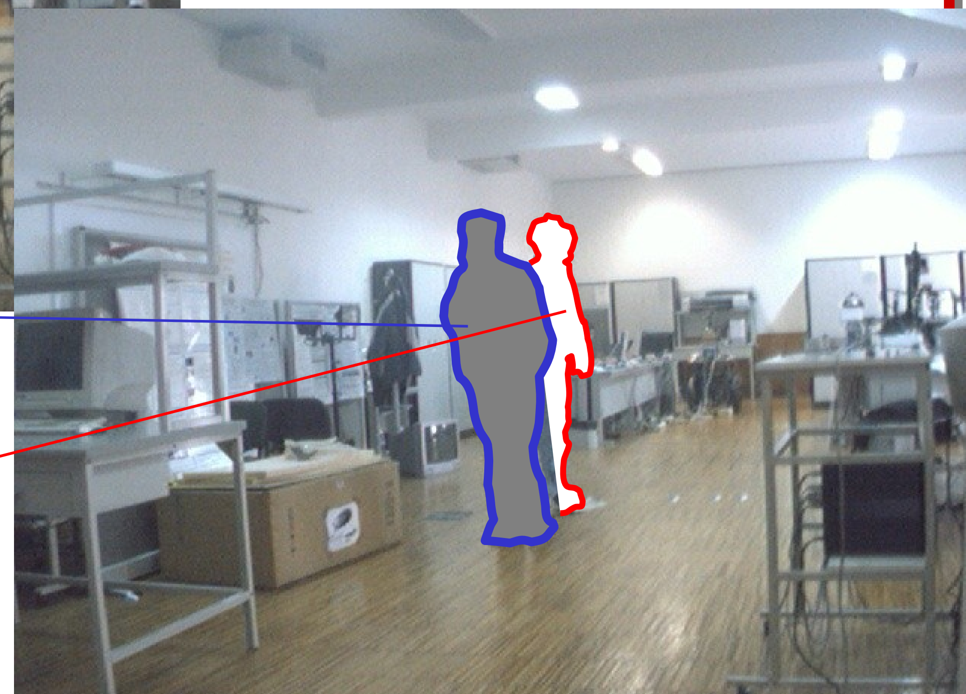
Classification of the distances inside the grid based on close values



24	395	648	47	757	543	616	27	363	24	395	648	47	757	543	616	27	363
722	642	510	965	886	379	597	321	401	722	642	510	965	886	379	597	321	401
886	304	599	66	369	849	685	52	514	886	304	599	66	369	849	685	52	514
937	85	336	510	716	546	974	370	46	937	85	336	510	716	546	974	370	46
925	885	723	705	286	815	280	384	991	925	885	723	705	286	815	280	384	991
756	94	118	625	789	486	174	336	310	756	94	118	625	789	486	174	336	310
377	951	520	227	205	116	922	891	256	377	951	520	227	205	116	922	891	256
834	618	995	631	430	143	260	238	610	834	618	995	631	430	143	260	238	610
687	761	313	153	629	785	375	81	447	687	761	313	153	629	785	375	81	447
555	621	506	311	559	269	645	301	418	555	621	506	311	559	269	645	301	418
86	579	820	649	246	557	316	180	412	[illegible]	579	820	649	246	557	316	180	412
555	281	299	836	657	136	192	716	714	555	281	299	836	657	136	192	716	714
679	100	709	18	132	425	233	790	[illegible]	679	100	709	18	132	425	233	790	103
428	722	869	404	513	58	220	[illegible]	[illegible]	428	722	869	404	513	58	220	331	380
163	841	656	800	274	797	79	782	[illegible]	163	841	656	800	274	797	79	782	752
693	821	862	160	587	560	124	777	[illegible]	[illegible]	821	862	160	587	560	124	777	149
24	395	648	47	757	543	616	27	[illegible]	[illegible]	395	648	47	757	543	616	27	363
722	642	510	965	886	379	597	321	[illegible]	[illegible]	642	510	965	886	379	597	321	401
886	304	599	66	369	849	685	52	[illegible]	[illegible]	304	599	66	369	849	685	52	514
937	85	336	510	716	546	974	370	[illegible]	[illegible]	85	336	510	716	546	974	370	46
925	885	723	705	286	815	280	384	[illegible]	[illegible]	885	723	705	286	815	280	384	991
756	94	118	625	789	486	174	336	[illegible]	756	94	118	625	789	486	174	336	310
377	951	520	227	205	116	922	891	256	377	951	520	227	205	116	922	891	256
834	618	995	631	430	143	260	238	610	834	618	995	631	430	143	260	238	610
687	761	313	153	629	785	375	81	447	687	761	313	153	629	785	375	81	447
555	621	506	311	559	269	645	301	418	555	621	506	311	559	269	645	301	418
86	579	820	649	246	557	316	180	412	86	579	820	649	246	557	316	180	412
555	281	299	836	657	136	192	716	714	555	281	299	836	657	136	192	716	714
679	100	709	18	132	425	233	790	103	679	100	709	18	132	425	233	790	103
428	722	869	404	513	58	220	331	380	428	722	869	404	513	58	220	331	380
163	841	656	800	274	797	79	782	752	163	841	656	800	274	797	79	782	752
693	821	862	160	587	560	124	777	149	693	821	862	160	587	560	124	777	149

Classification of the distances inside the grid based on close values



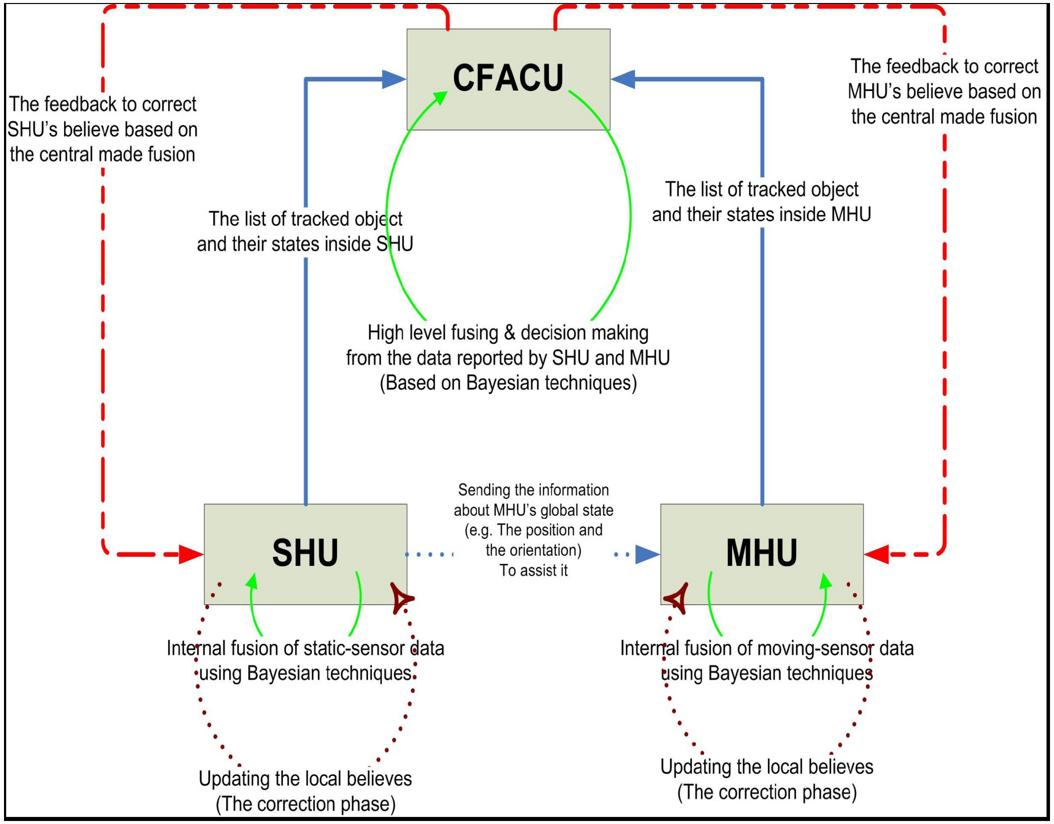


Class 1:
Front and totally visible

Class 2:
Back and partially visible

Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra

Cooperation between static and mobile sensors



SHU: Static Heterogeneous Sensors Unit.
MHU: Mobile Heterogeneous Sensors Unit.

CFACU: Central Fusing And Correcting Unit.

Thanks for your attention !