Video analysis in Pan-Tilt-Zoom camera networks

Dr. Christian Micheloni

Dept. Computer Science

University of Udine

Materials and Acks

- Material
  - http://users.dimi.uniud.it/~christian.micheloni/PTZ4ICDSC.html
- Acknoledgements
  - Dr. Claudio Piciarelli, University of Udine for the cooperation on PTZ network reconfiguration.
  - Dr. Sanjeev Kumar, University of Roorke for the work on Stero-PTZ object localization
  - European Regional Development Fund, Interreg IV Italia-Austria Program, Project SRSnet ID 4687 2009-2012

SRSnet

Lakeside Labs
SELF-ORGANIZING NETWORKED SYSTEMS

EYE-TECH

Introduction

- Video Sensor Networks provide security by deploying video cameras.

- Greater security requires more cameras
  - Wide Vs. Narrow angle of view (resolution Vs. coverage)
  - Non overlapping FOVs limit processing accuracy
    - Occlusions
    - Localization
  - Overlapping FOVs requires huge amount of cameras

C. Micheloni, B. Rinner, G.L. Foresti, «Video Analysis in Pan-Tilt-Zoom Camera networks», *IEEE Signal Processing Magazine*, vol.27, no.5, pp.78-90, Sept. 2010.

Pan Tilt Zoom Cameras

- Pan-Tilt-Zoom cameras by adapting the FOV can limit such a requirement.
  - High resolution
  - Overlap FOV to solve occlusion, localization and tracking
- PTZs on the other hand introduce new problems:
  - Ego-Motion estimation
    - Low level motion detection techinques cannot be directly exploited
  - Calibration requires new solutions

C. Micheloni, B. Rinner, G.L. Foresti, «Video Analysis in Pan-Tilt-Zoom Camera networks», *IEEE Signal Processing Magazine*, vol.27, no.5, pp.78-90, Sept. 2010.

ACTIVE VISION

- Aloimonos proposed the active vision paradigm to describe the dynamic interaction between the observer and the object observed to actively decide what to see.

With PTZ cameras we want to decide

Where

What

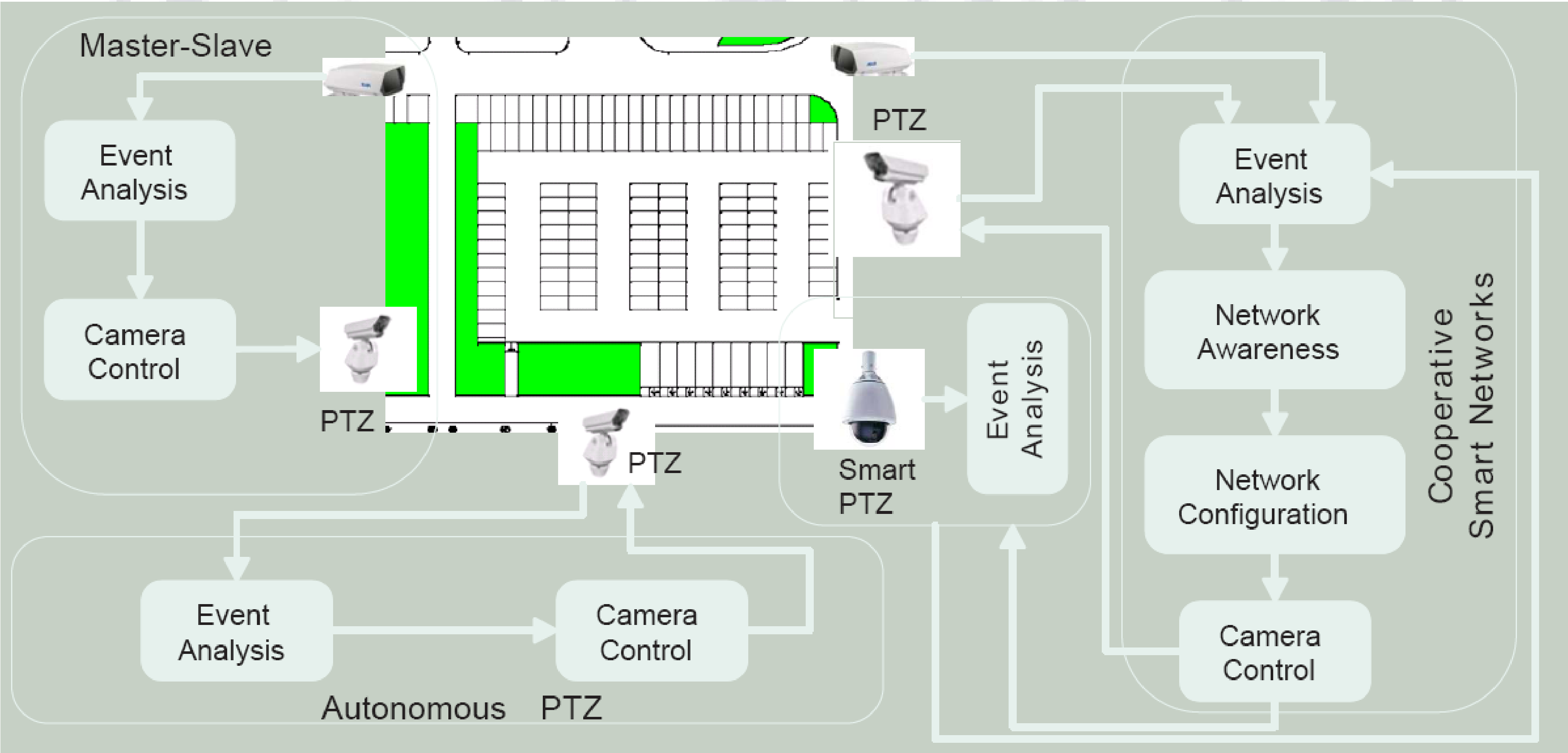
How

looking at the scene

Y. Aloimonos. Active Perception. Lawrence Erlbaum Associates, 1993.

Evolution of PTZ usage

- From master slave to cooperative smart networks



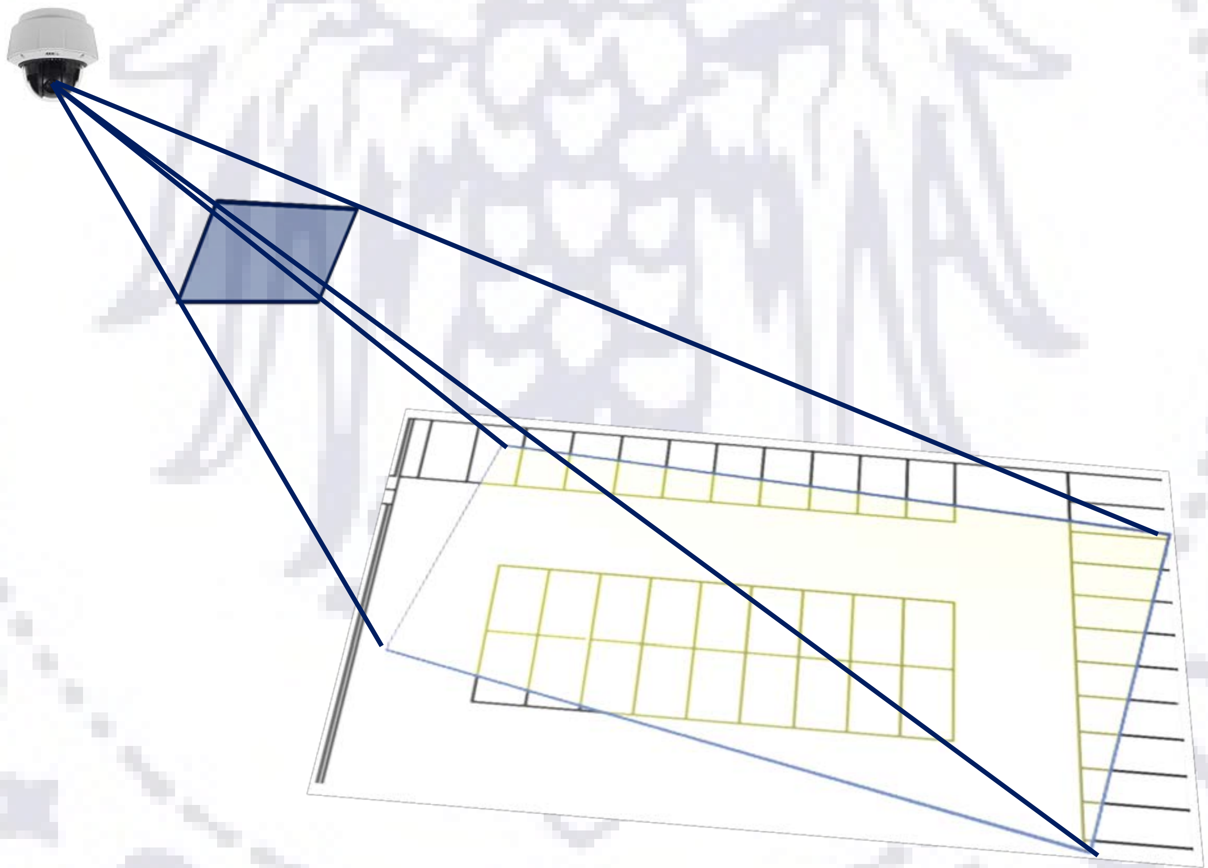
C. Micheloni, B. Rinner, G.L. Foresti, «Video Analysis in Pan-Tilt-Zoom Camera networks», *IEEE Signal Processing Magazine*, vol.27, no.5, pp.78-90, Sept. 2010.

Where is a PTZ looking at?

The estimation of the PTZ field of view while moving

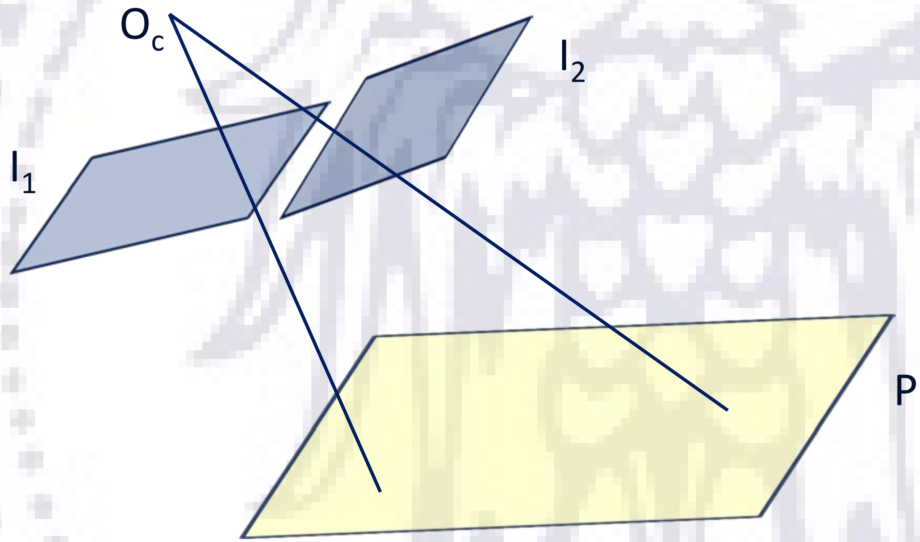
Camera Field of View

- The area covered is the projection of the Image plane to the ground. Usually a convex polygon.



Static camera Homography

- If space points are coplanar then there is a projective transformation H between the world P and image plane I[1].



$$p_{i_1} = H_1 P$$

$$p_{i_2} = H_2 P$$

$$H_{3\times3} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix}$$

- Easy to compute: at least 4 points of P matching in I.

Problem

We need a homography for each combination of P,T,Z paramters

[1]R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press

Possible solutions

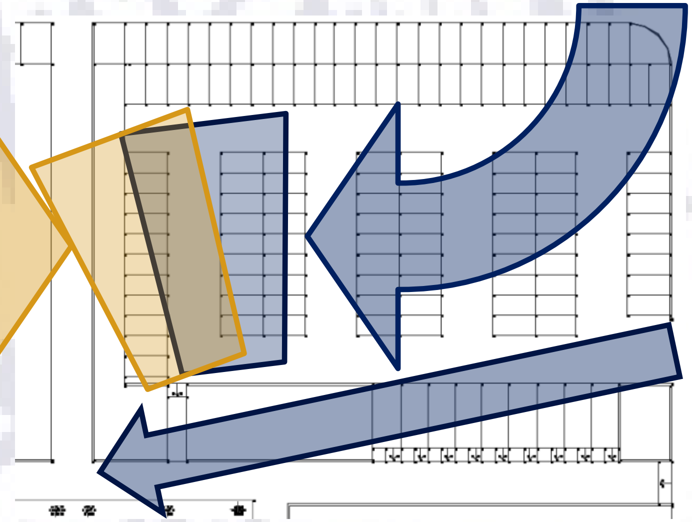
- Full calibration of the PTZ camera
  - S. N. Sinha and M. Pollefeys, ”Pan–tilt–zoom camera calibration and high-resolution mosaic generation”, Computer Vision and Image Understanding, Vol. 103, Issue 3, Sep. 2006, pp. 170–183
- Homography estimation
  - I. N. Junejo and H.Foroosh Optimizing PTZ camera calibration from two images, Machine and Vision Applications, Vol. 23, No 2 (2012), pp.375-389
- Trigonometry

Homography Estimation

$I_1(\varphi_C^1, \theta_c^1, Z_c^1)$

$I_2(\varphi_C^2, \theta_c^2, Z_c^2)$

M



$$M = H_{I_1,M} I_1$$

$$M = H_{I_2,M} I_2$$

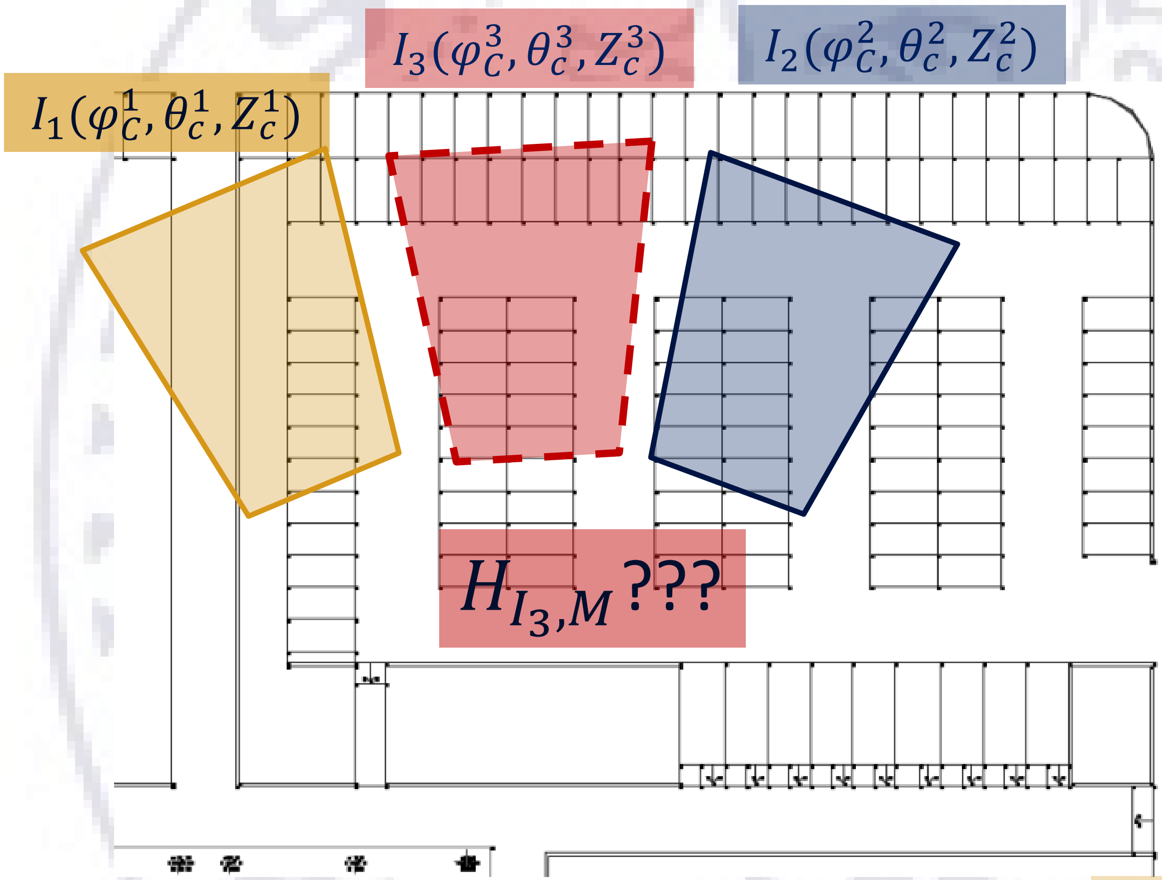
$$H_{I_2,M}^{-1} M = I_2$$

$$H_{I_2,M}^{-1} M = H_{I_2,M}^{-1} H_{I_1,M} I_1 = I_2$$

$$H_{I_1,I_2} = H_{I_2,M}^{-1} H_{I_1,M}$$

Problem: I need to compute H_{I_1,I_2} for all I_2 given $H_{1,M}$

Homography Estimation



We have

$H_{I_2,M}$ $H_{I_1,M}$

- Compute sift matching between I_3 and I_1/I_2
- Compute H_{I_3,I_1} and H_{I_3,I_2}
- Select the one that minimizes the errors

$$H_{I_3,M} = H_{I_1,M} * H_{I_3,I_1} * I_3$$

$$H_{I_3,M} = H_{I_2,M} * H_{I_3,I_2} * I_3$$

Problem: sift/surf are not fast

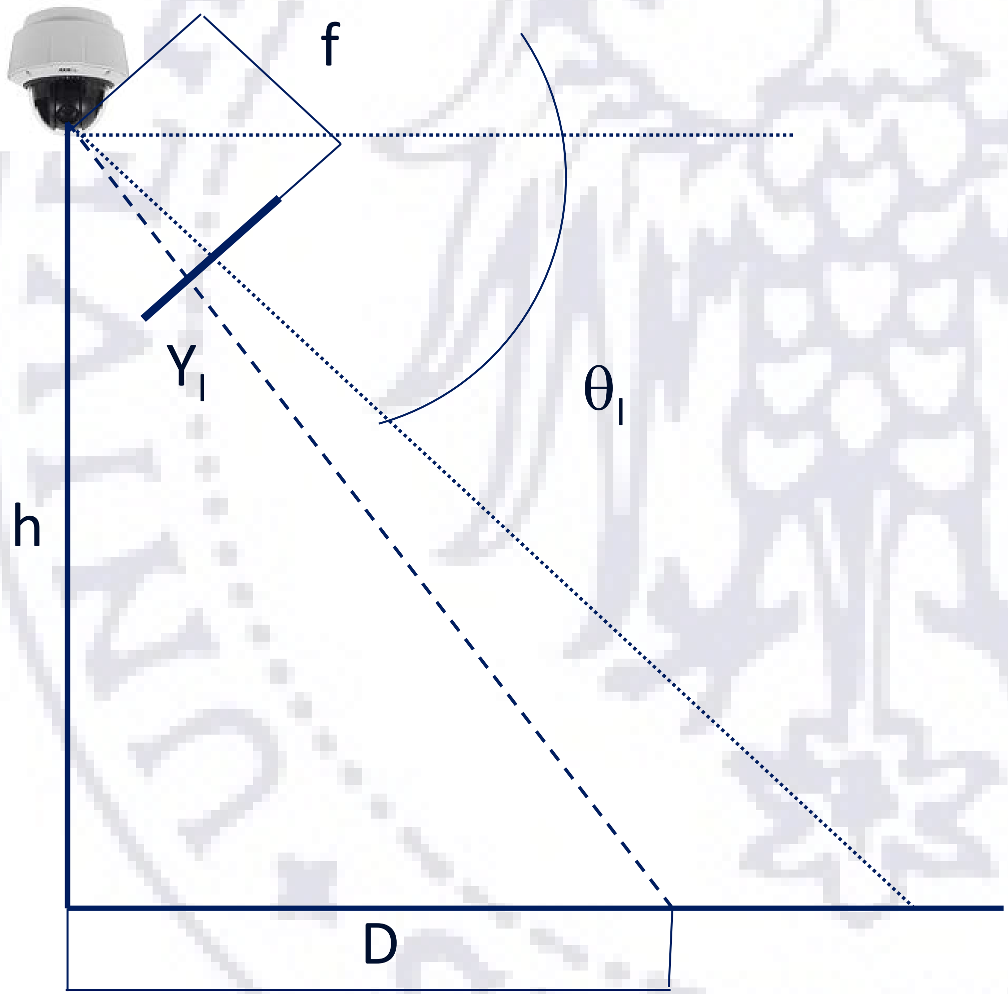
Trigonometry

- Useful when the environment is not well structured then does not allow full vision approaches.

- It is required:
  - To be able to read camera parameters
  - To know the position of the camera reference system with respect to the world reference system
  - The camera should incorporate an algorithm to center a camera point into the center of the image by panning and tilting.

From camera to world

- An image point (x_I, y_I) can be projected into the world by using geometric properties



$\theta_{\ I} = func(\mathrm{f}, \mathrm{y_I}, \theta_c)$

$D = h * ctan(\theta_I)$

φ_C Pan angle of the camera

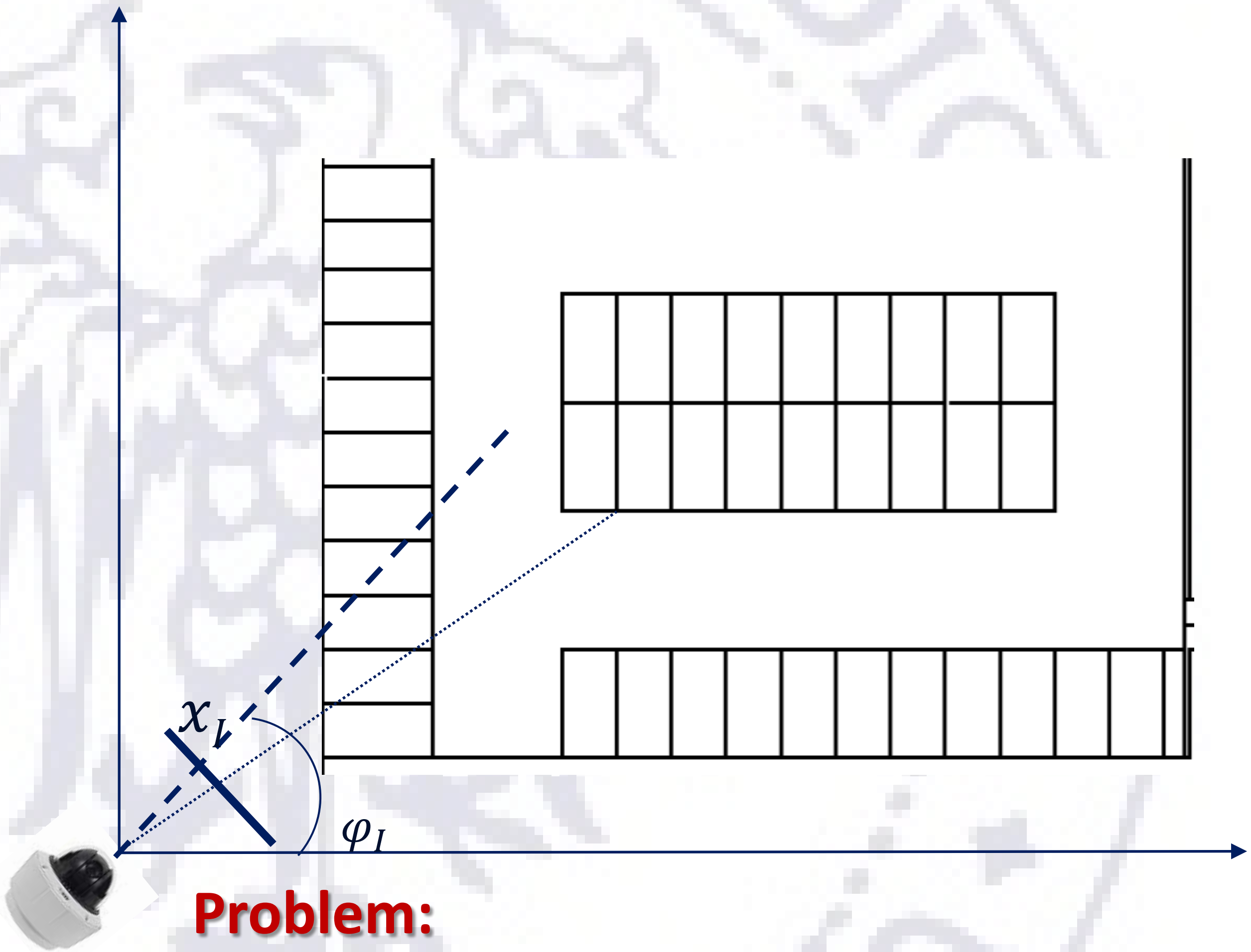
θ_C Tilt angle of the camera

From camera to world

$$X_m = D * \cos(\varphi_I)$$

$$Y_m = D * \sin(\varphi_I)$$

$$\varphi_I = func(f, x_I, \varphi_c)$$



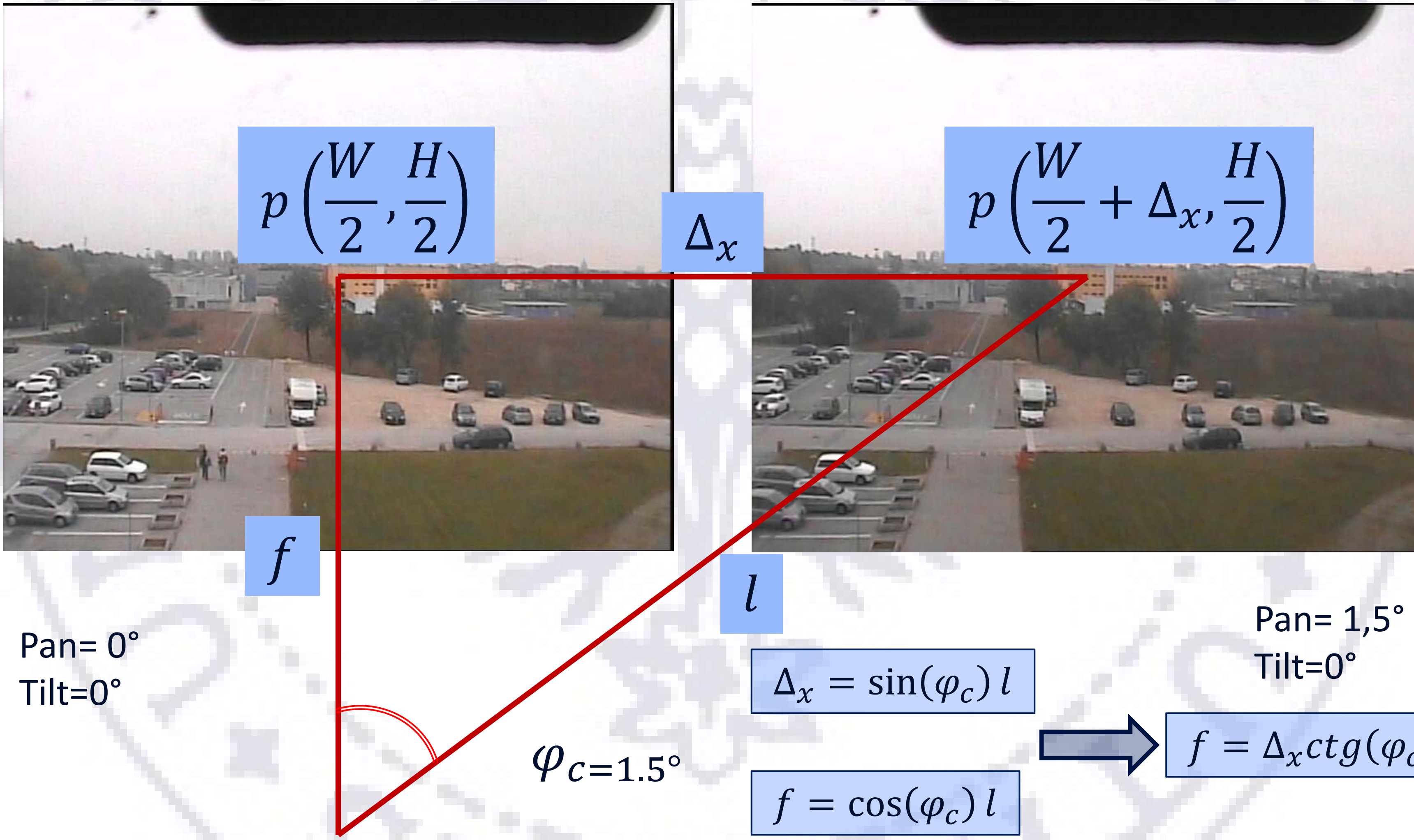
Problem:

We do not read the focal length from the camera. We read the zoom step Z

We can estimate $f = func(Z)$

Computing f

- Center the camera on two different points lying on the central line



Pan= 0°
Tilt=0°

Pan= 1,5°
Tilt=0°

$$\Delta_x = \sin(\varphi_c)\, l$$

$$f = \cos(\varphi_c)\, l$$

$$\Rightarrow \quad f = \Delta_x ctg(\varphi_c)$$

Computing f

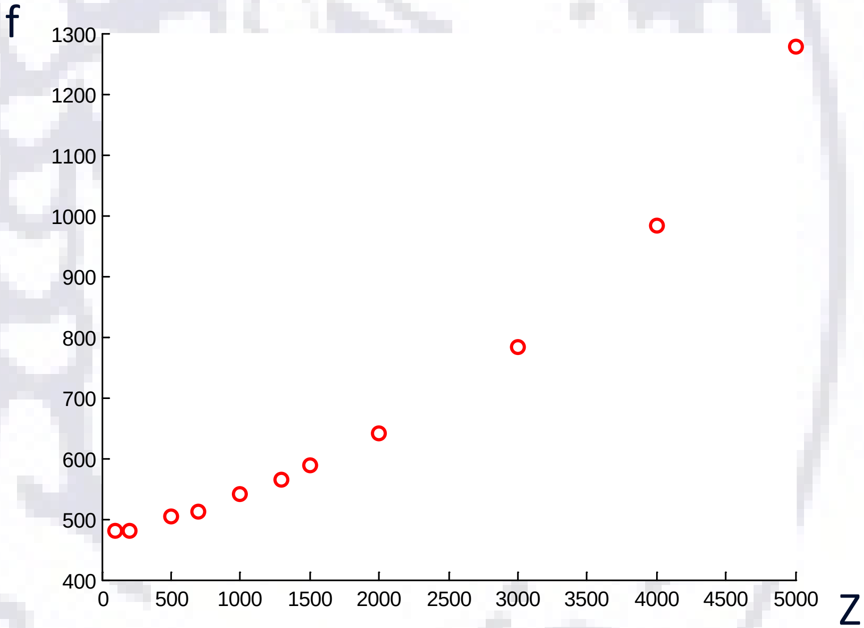
Zoom 1

Zoom 100

⋮

Zoom 600

⋮

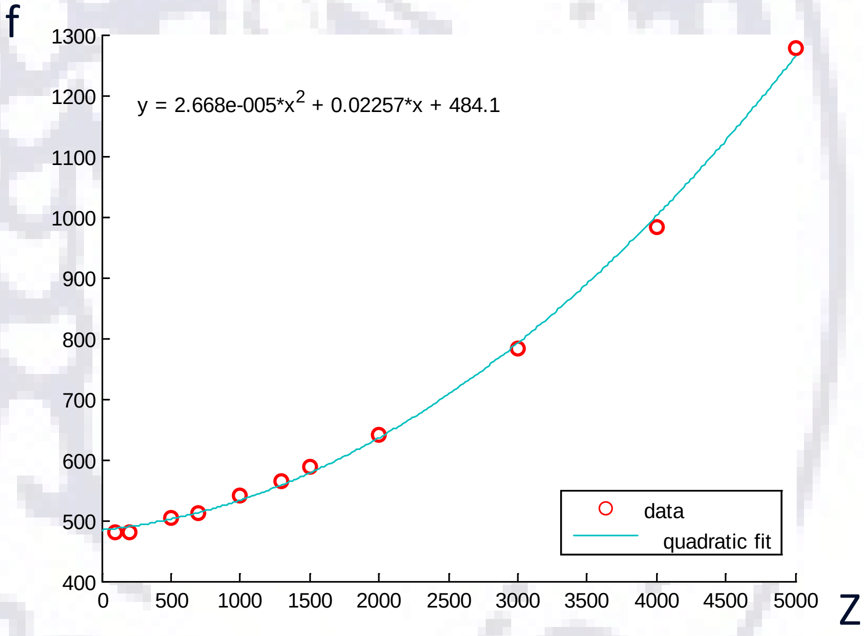


Computing f

Zoom 1

Zoom 100

Zoom 600

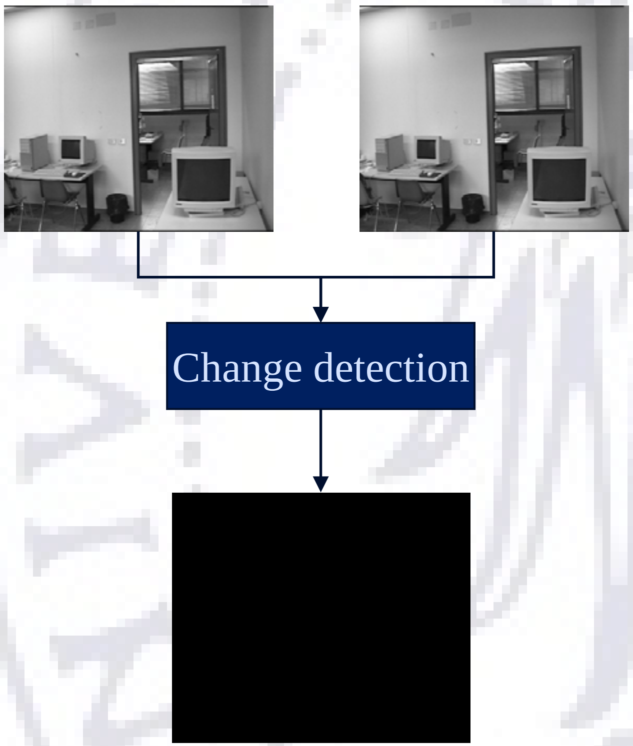


What is a PTZ looking?

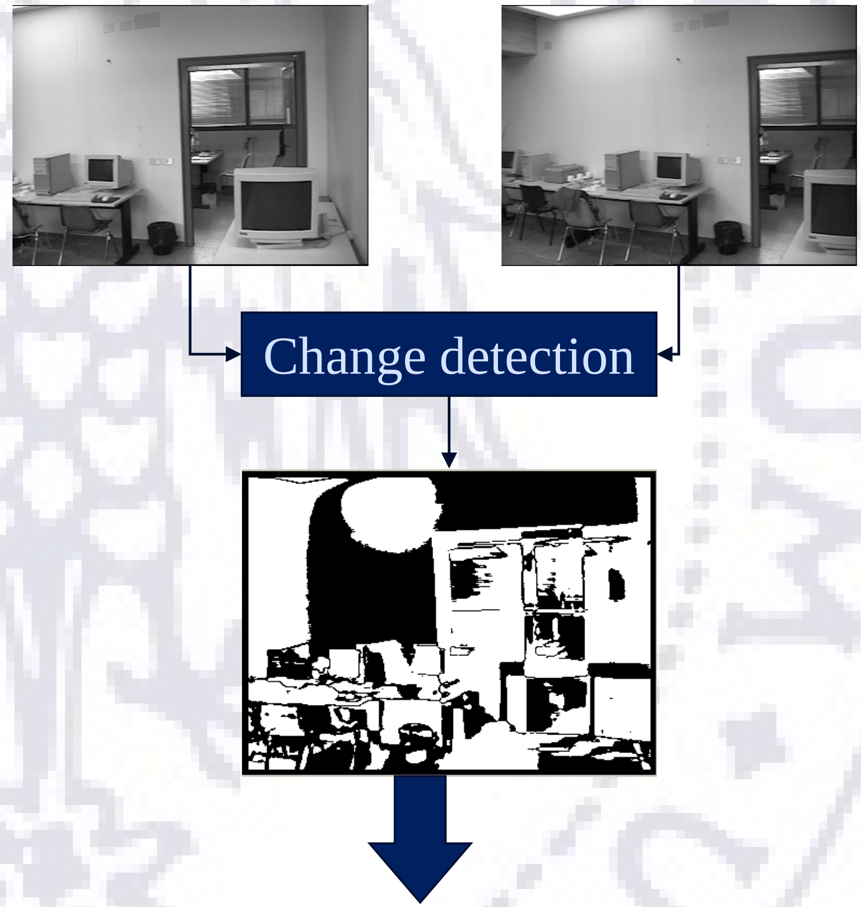
Detecting and tracking moving objects

Change Detection based

Static camera



PTZ camera



Compensation of the motion induced by the camera

Tracking in Pan-Tilt

- 9 Areas of tracking
- Area Alignment
- Best 3 used for Affine Computation

Tracking in Pan-Tilt

- Computation of clusters

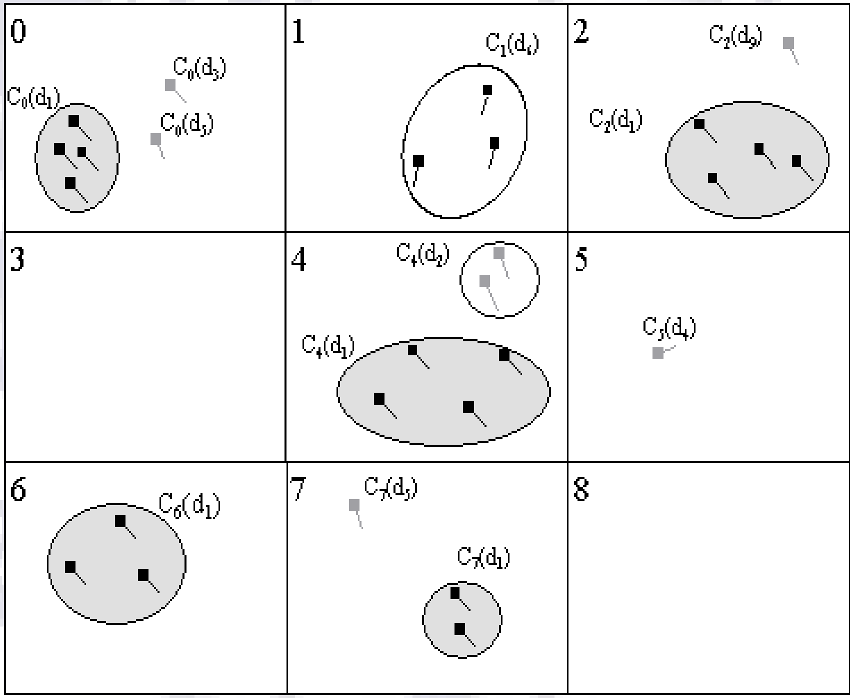
$$C_w(\mathbf{d}) = \{f_i \in TFS_w \mid \mathbf{d_i} = \mathbf{d}\}$$

- Selection of the best cluster for each area

$$RF(C_w(\mathbf{d}_k)) = \frac{\sum_{f_j \in C_w(\mathbf{d}_k)} E_{f_j}}{|C_w(\mathbf{d}_k)|^2}$$

- Selection of the best feature for each of the 3 best clusters

$$f'_i = \underbrace{\begin{pmatrix} a_1 & a_2 & a_5 \\ a_3 & a_4 & a_6 \\ 0 & 0 & 1 \end{pmatrix}}_{A} f_i$$



E_{f_i} is the residual or the error in tracking the feature i

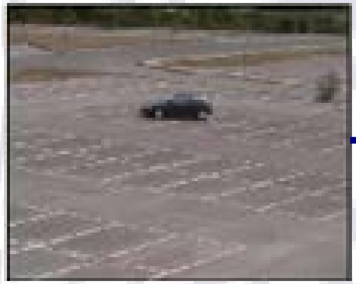
Pan & Tilt to track

Registration is performed on the basis of an affine transform

Registration when zooming

Frame t-1

Frame t



Object's features extraction

Feature Tracking

Fixation Computation

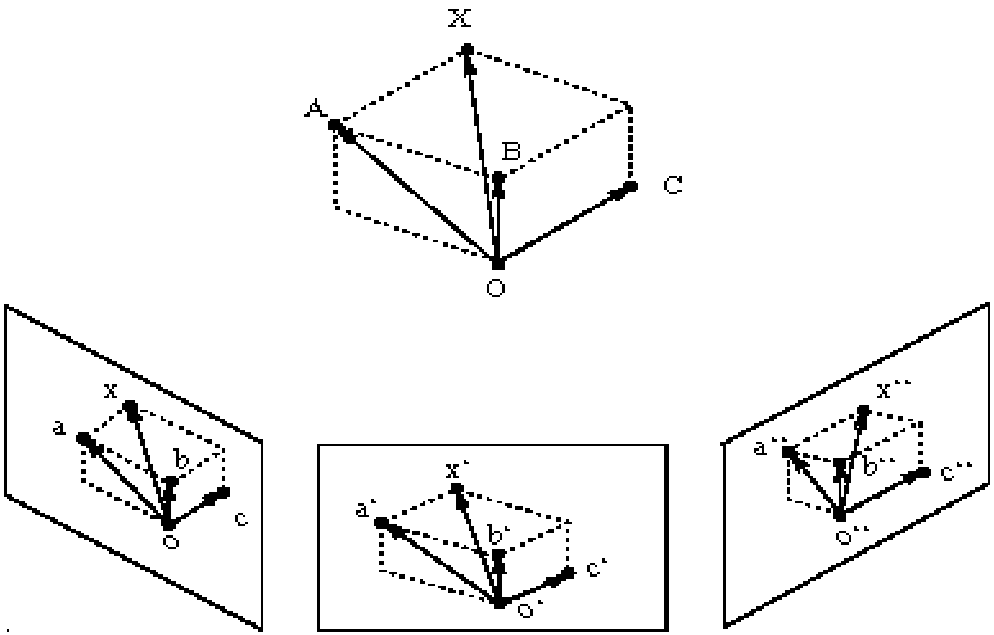
Fixation Point Esteemation

A Fixation Point

Affine transfer

Planar Points $g' = Ng + r$

$$N = \begin{pmatrix} n_1 & n_2 \\ n_3 & n_4 \end{pmatrix} \qquad r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

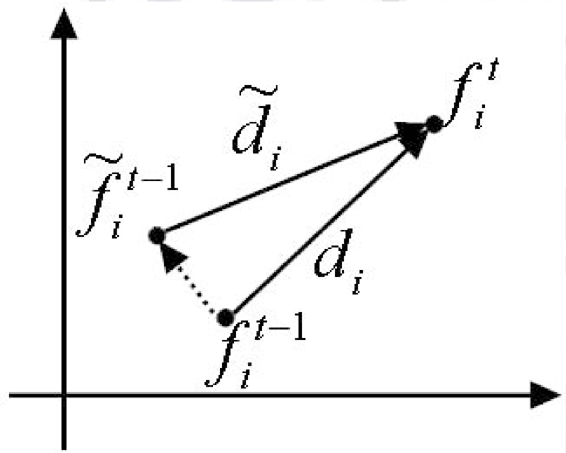


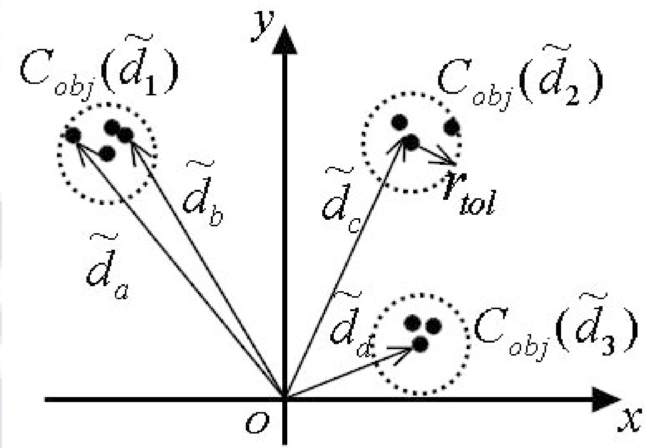
$$\begin{pmatrix} x_1' & y_1' & 1 \\ x_2' & y_2' & 1 \\ \vdots & \vdots & \vdots \\ x_n' & y_n' & 1 \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \\ r_1 \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}$$

Tordoff- Murray – PAMI, January 2004

- Least square
- SVD

Clustering Feature Points





$$\tilde{d}_i = f_i^t - \tilde{f}_i^{t-1} = f_i^t - \tilde{A} f_i^{t-1}$$

Algorithm 1 Clustering

repeat
  Feature extraction and tracking
  Clusters computation
  Background cluster deletion
  Computation of the centre of mass for each cluster
until zooming

Example

Localization in PTZs network

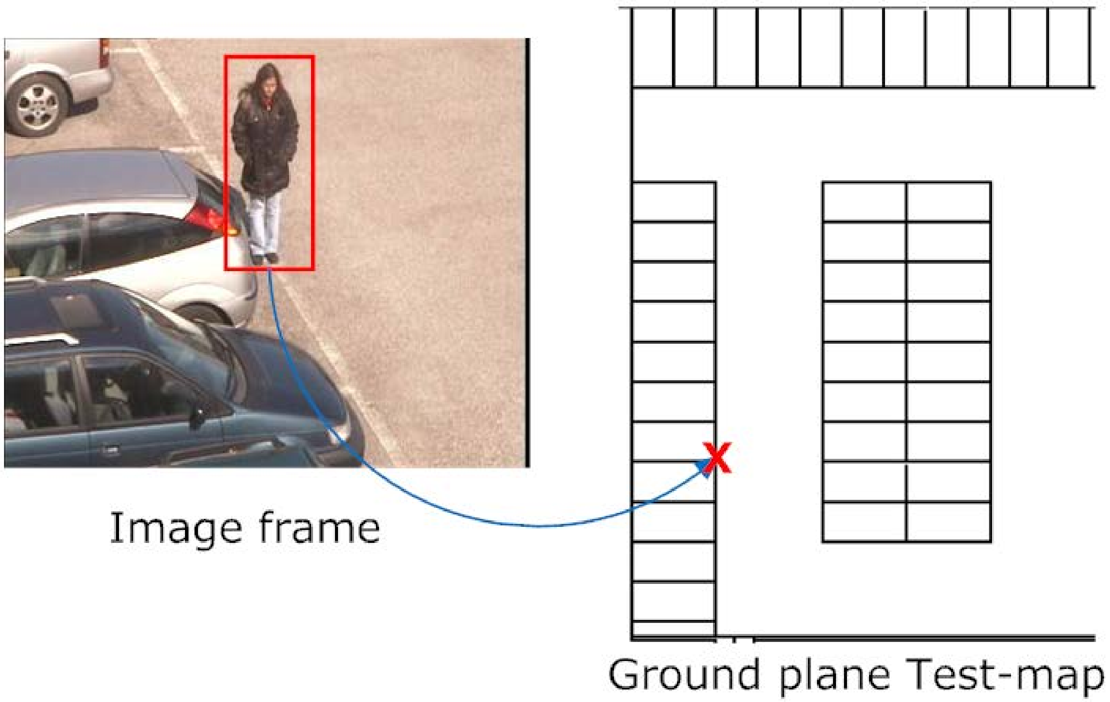
From static-PTZ to PTZ-pair cooperation exploiting stereo vision

Localization

What is localization?

The task of localization is to localize a object/target (moving, static) on a given test-map.

Generally, localization is made by taking the intersection of optical axis with the ground position of object/target in respective frames.

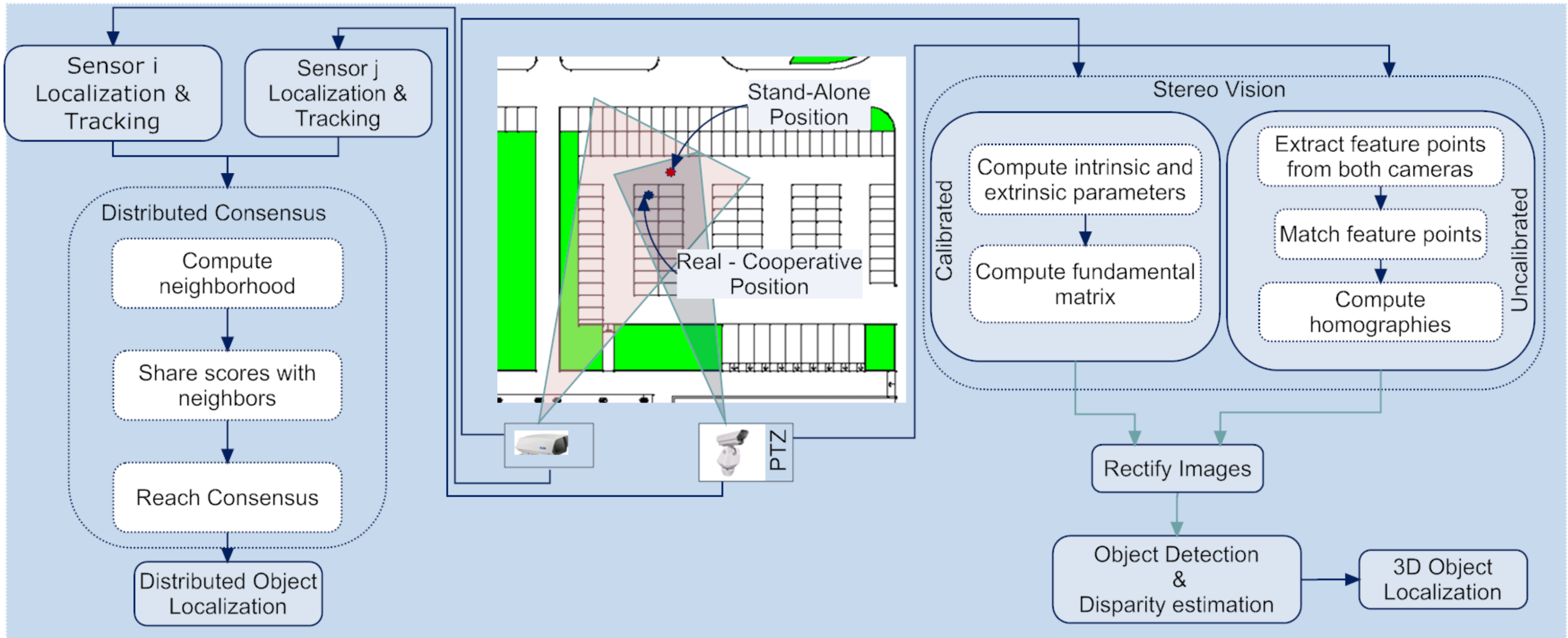


Occlusions in localization

What happens in the case of partially occluded object? Means, where ground position of object/target is occluded by some other objects.

Erroneous localization using monocular camera based existing techniques.

Cooperative camera localization



Comparison

Technique	Advantage	Disadvantage
Centralised	Complete network information Pixel level fusion	High bandwidth usage Affected by occlusions
Distributed	Dynamic network topology Distributed consensus Low bandwidth usage Ad-hoc network communication	Accuracy depending on single camera estimations Affected by occlusions
Stereo	Precise 3D localisation Occlusions free	Constrained points of view Base line constrained

R. Olfati-Saber, J. Fax and R. Murray. Consensus and cooperation in networked multi-agent systems. Proceedings of the IEEE, 95(1):215-233, 2007

C. Soto, B. Song and A. Roy-Chowdury. Distributed multi-target tracking in a self-configuring camera network. In IEEE Conference on Computer Vision and Patter Recognition, pp 1486-1493, 2009

Bi Song, D. Chong, A.T. Kamal , J.A. Farrell and A. Roy-Chowdury, Distributed Camera Networks, IEEE Signal Processing Megazine, Vol. 28, No. 3, pp. 20-31, 2011

Stereo Vision:

- Stereo vision: To extract 3-D information of a real world scene/object from its two or more 2-D images [2, 5].

- Steps:
  1. Calibration,
  2. Rectification,
  3. Stereo Matching

- Remarks: Generally, a pair of cameras having equal internal image parameters and fixed field of views are used in classical stereo systems.

Nonhomgeneous Stereo Vision

- Use of a pair static-PTZ cameras as a stereo system.

  Abidi, B., Koschan, A., Kang, S., Mitckes, M., Abidi, M., 2003. Automatic Target Acquisition and Tracking with Cooperative Static and PTZ Video Cameras. Kluwer Academic, Ch. Multisensors Surveillance Systems: The Fusion, pp. 43–59.

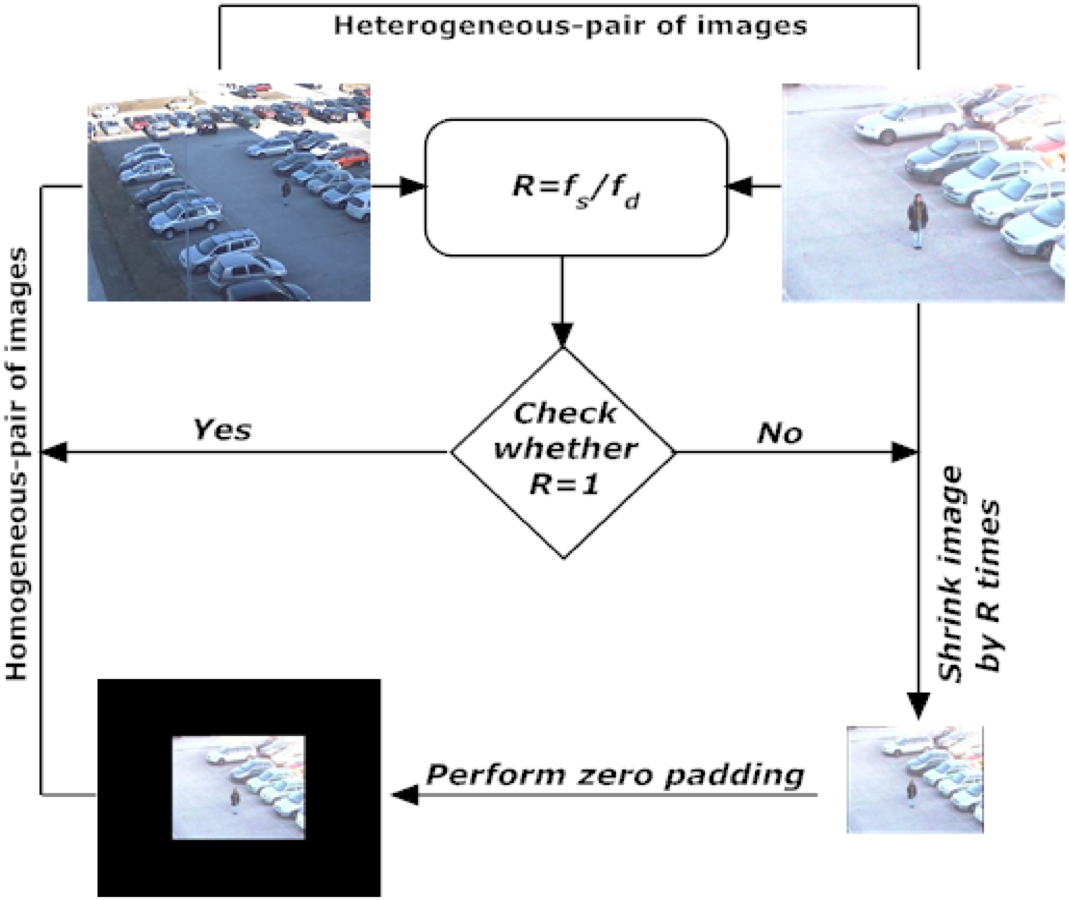
  Chen, C. H., Yao, Y., Page, D., Abidi, B., Koschan, A., Abidi, M., 2008. Heterogeneous fusion of Omni-directional and PTZ cameras for multiple object tracking. IEEE Transaction on Circuit and Systems for Video Technology 18(8),, 1052–1063

  Kumar, S., Micheloni, C., Piciarelli, C., Foresti, G. L., 2009. Stereo localization based on network's uncalibrated camera pairs. In: Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 502–507.

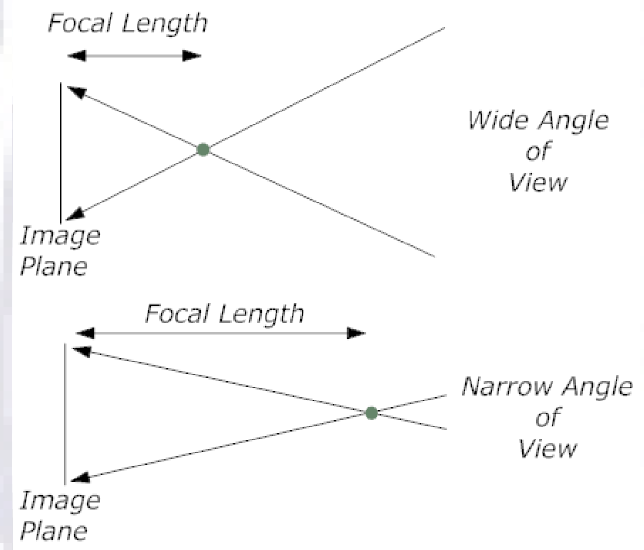
- Advantage: More degree of freedoms in terms of
  - Angle of Views
  - Field of views
  - Zoom (close or wide focus according to the requirement)

- Difficulties:
  - To calibrate PTZ camera is very difficult for each position in real time applications.
  - Heterogeneity involved in internal camera parameters.

Compensation of the heterogeneity

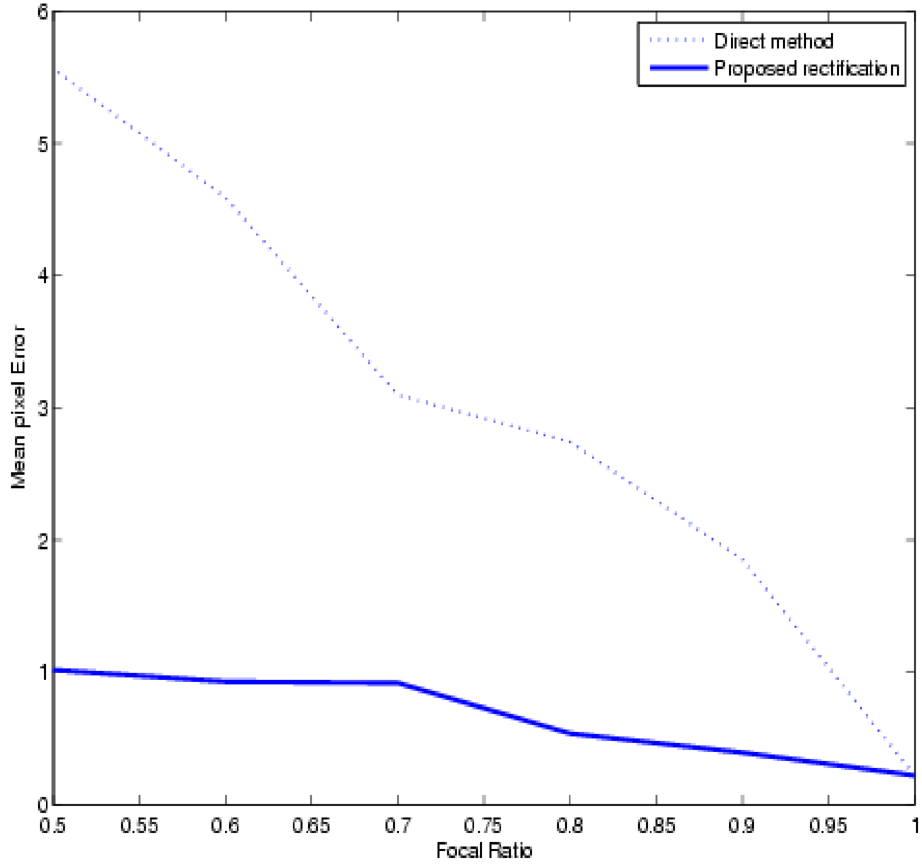


Idea: Focal length is inversely proportional to field of view.



Rectification Results

Mean-Error between epipolar lines in rectified pairs of images (Kumar et al Vs. Hart et al.)



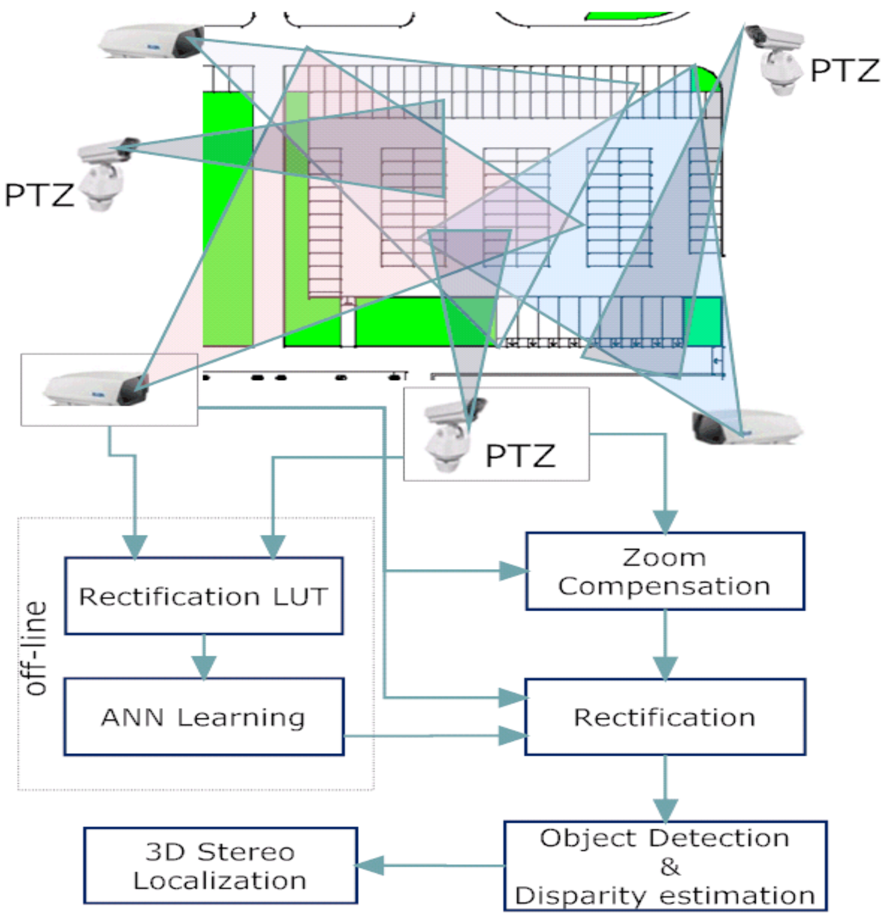
Direct method[2]: Rectification using image pairs without compensating the effect of heterogeneity in internal image parameters.

Homogeneous rectification[1]: Rectification made after making image pairs as homogeneous in terms of internal image parameters.

Kumar, S., Micheloni, C., Piciarelli, C., Foresti, G. L., 2009. Stereo localization based on network's uncalibrated camera pairs. In: Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 502–507.

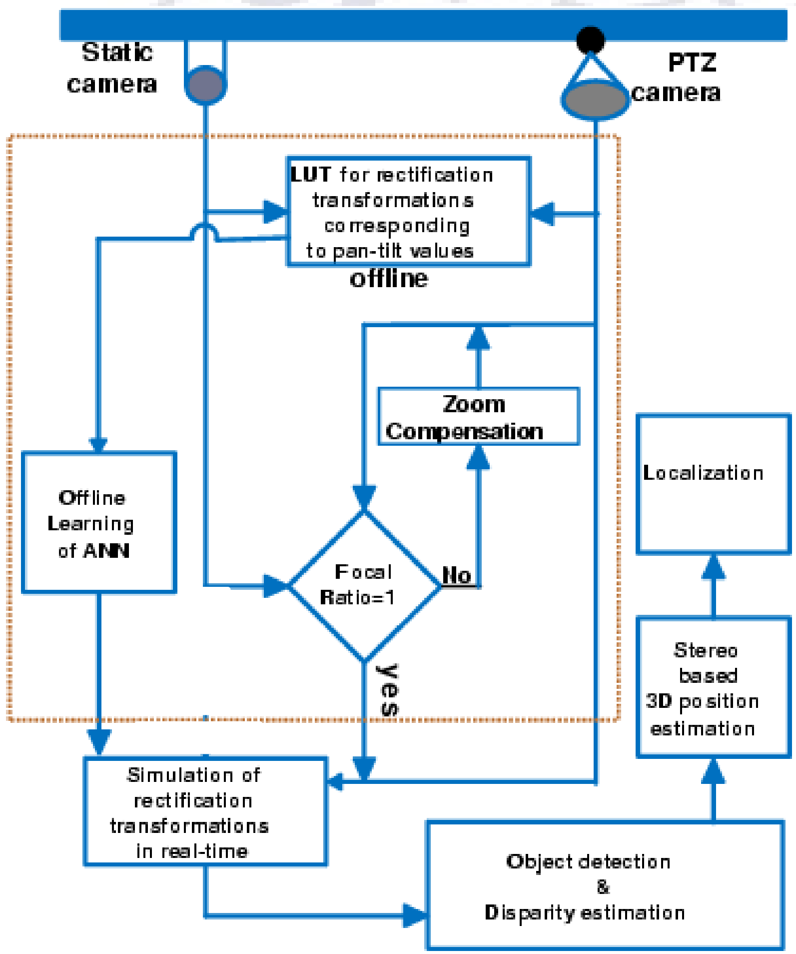
Hart, J., Scassellati, B., Zucker, S., 2008. Epipolar geometry for humanoid robotic heads. In: International Cognitive Vision Workshop. pp. 24–36.

System architecture:



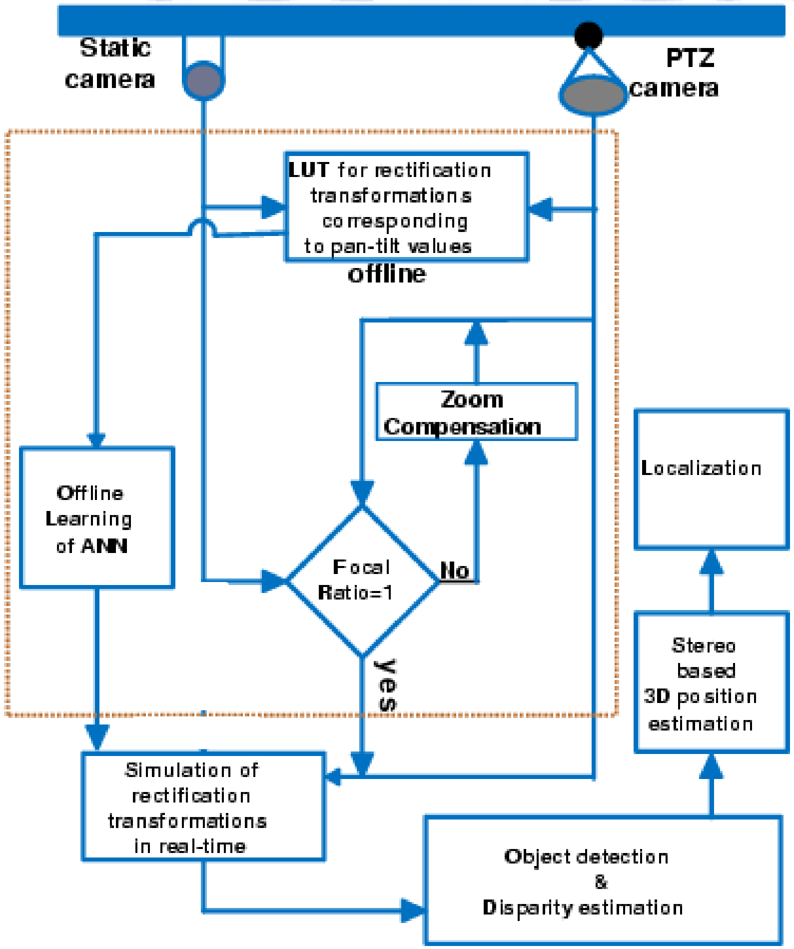
In a network containing various PTZ and static cameras, cooperativeness between various cameras involved for considering following two points

- **Once the target is selected in any static camera then selection of nearest PTZ camera for stereo vision task.**
- **Once the target move away from the field of view of respective static camera then the selection of next static camera based on the trajectory created by PTZ camera based on the motion of target.**



Main steps: Offline

1. Sampling of pan-tilt angles for the PTZ camera.
2. Construction of LUT for rectification transformation for each possible pairs of static-PTZ cameras.
3. Train neural network

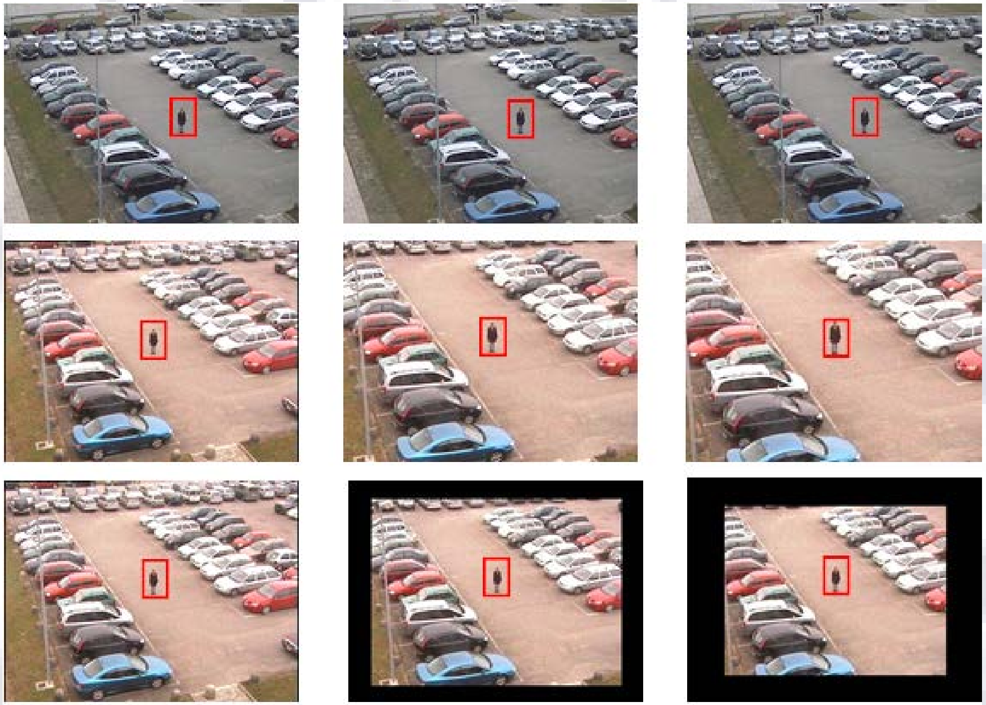


Main steps: Online

1. **Select target with a static camera**
2. **Redirect the nearest PTZ camera**
3. **Grab image sequences from both cameras.**
4. **Acquire Pan and Tilt Information for a grabbe image**
5. **Interpolate rectification transformation using NN-based interpolation.**
6. **Rectify stereo sequences and compute disparity for target position**
7. **Computer stereo based 3D position of object**
8. **Make localization**

Heterogeneous stereo vision for object’s localization

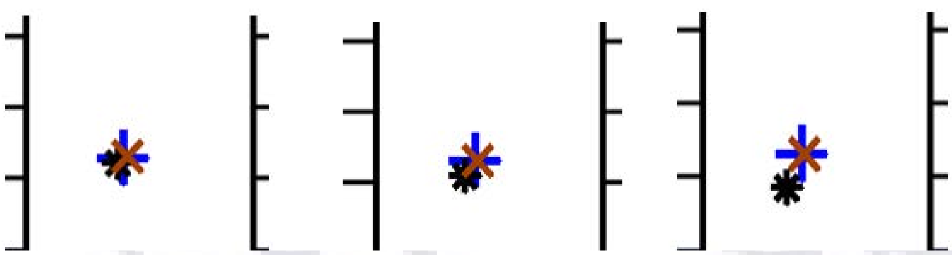
Results in unequal zoom case



Static camera frames

PTZ camera frames

PTZ camera frames (after zoom compensation)



Localization

+ Proposed stereo localization

* Kinematics chain app. (Hart et al., 2008)

x Monocular localization

*J. Hart, B. Scassellati and S.W. Zucker. Epipolar Geometry for Humanoid Robotic Heads, Proc. Of *4th* International Cognitive Vision Workshop, 24–36,2008.

Heterogeneous stereo vision for object's localization

Results in case of partially occluded target

Static camera frames

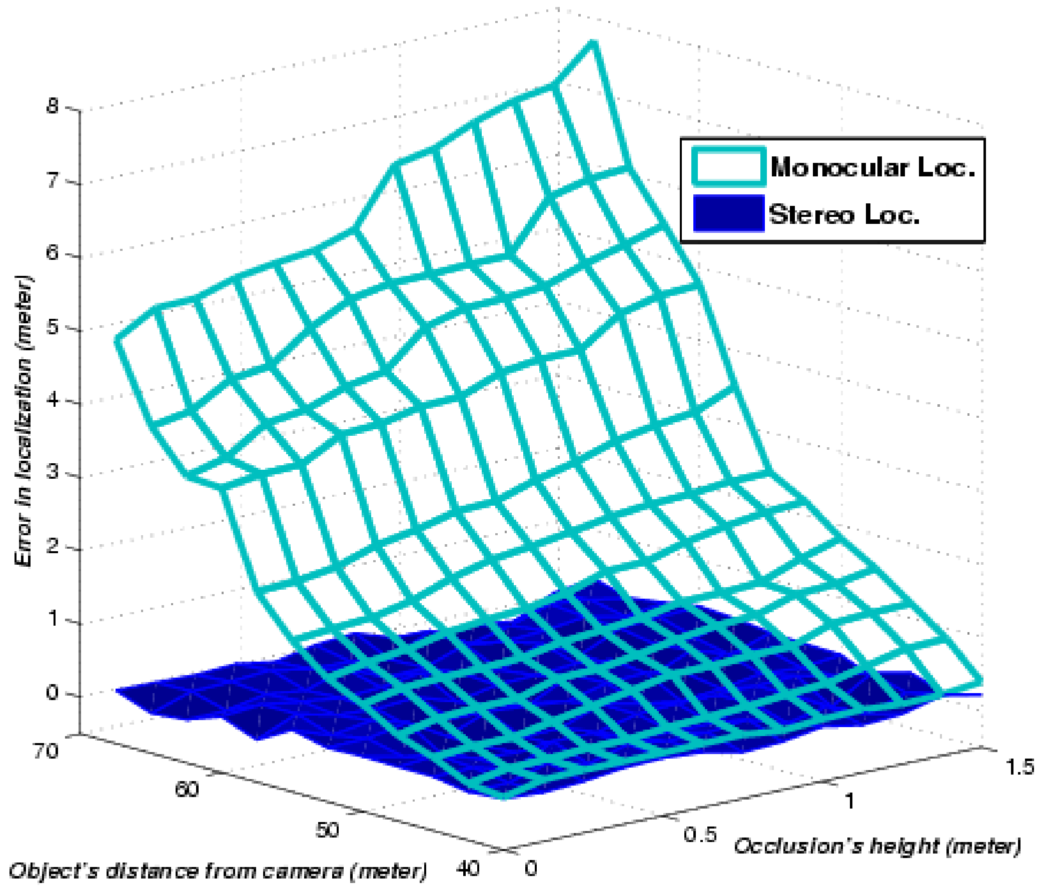
PTZ camera frames

Localization

+ Proposed stereo localization

x Monocular localization

Localization Error



Using Two PTZ for stereo localisation

- Rectification is usually performed by using features:
  - Corners
  - Sift
  - Surf

- These features work well only under a sligthly change in the point of view

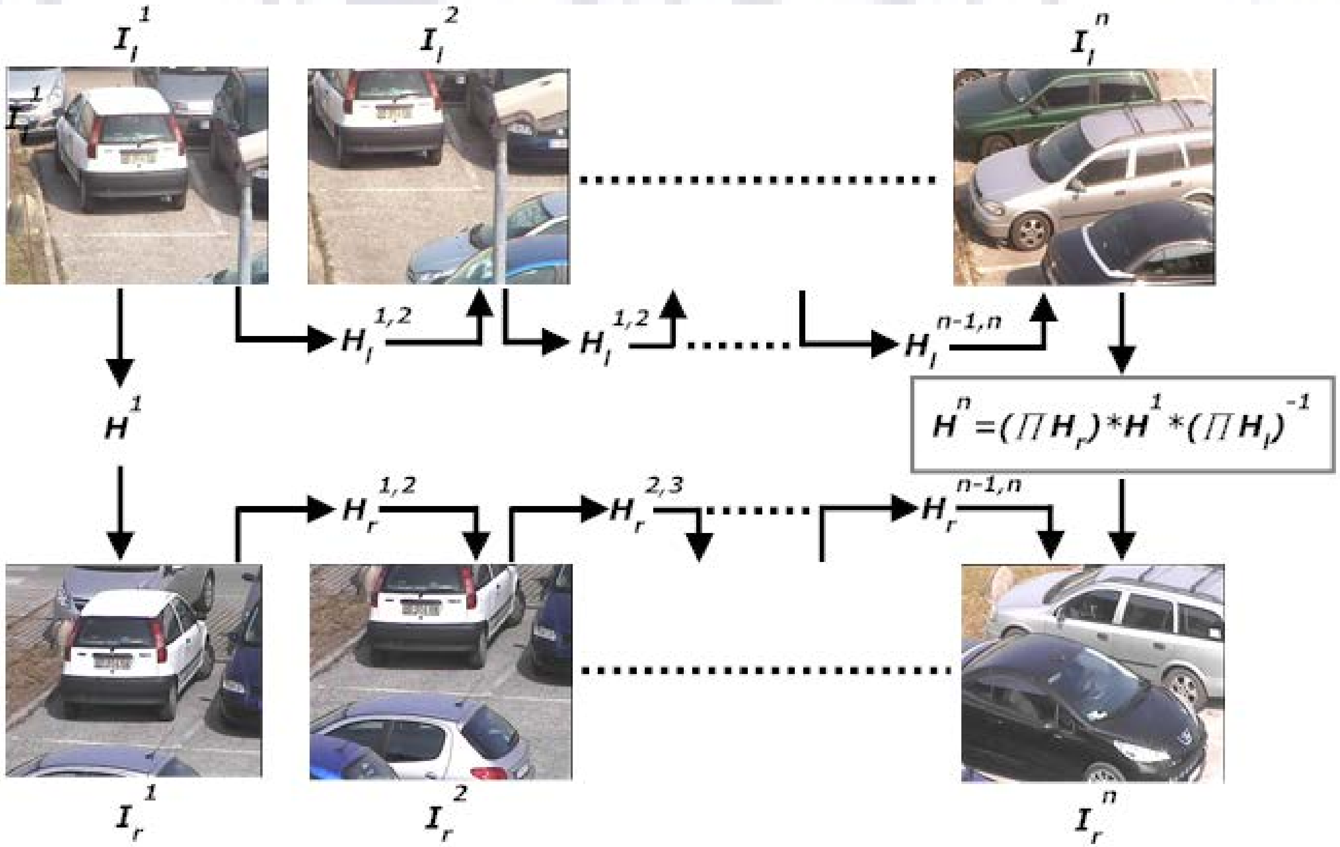
- In stereo vision this means a short baseline

Problem

PTZs in a video network are deployed far away so they have a wide baseline

Solution for wide baseline matching

- Using the chain of homographies:



OFFLINE OPERATIONS

- An offline LUT is constructed for the possible combinations of pan and tilt angles of both the cameras based on the sampling.
- A NN is trained to estimated homography parameters given the pan and tilt position.

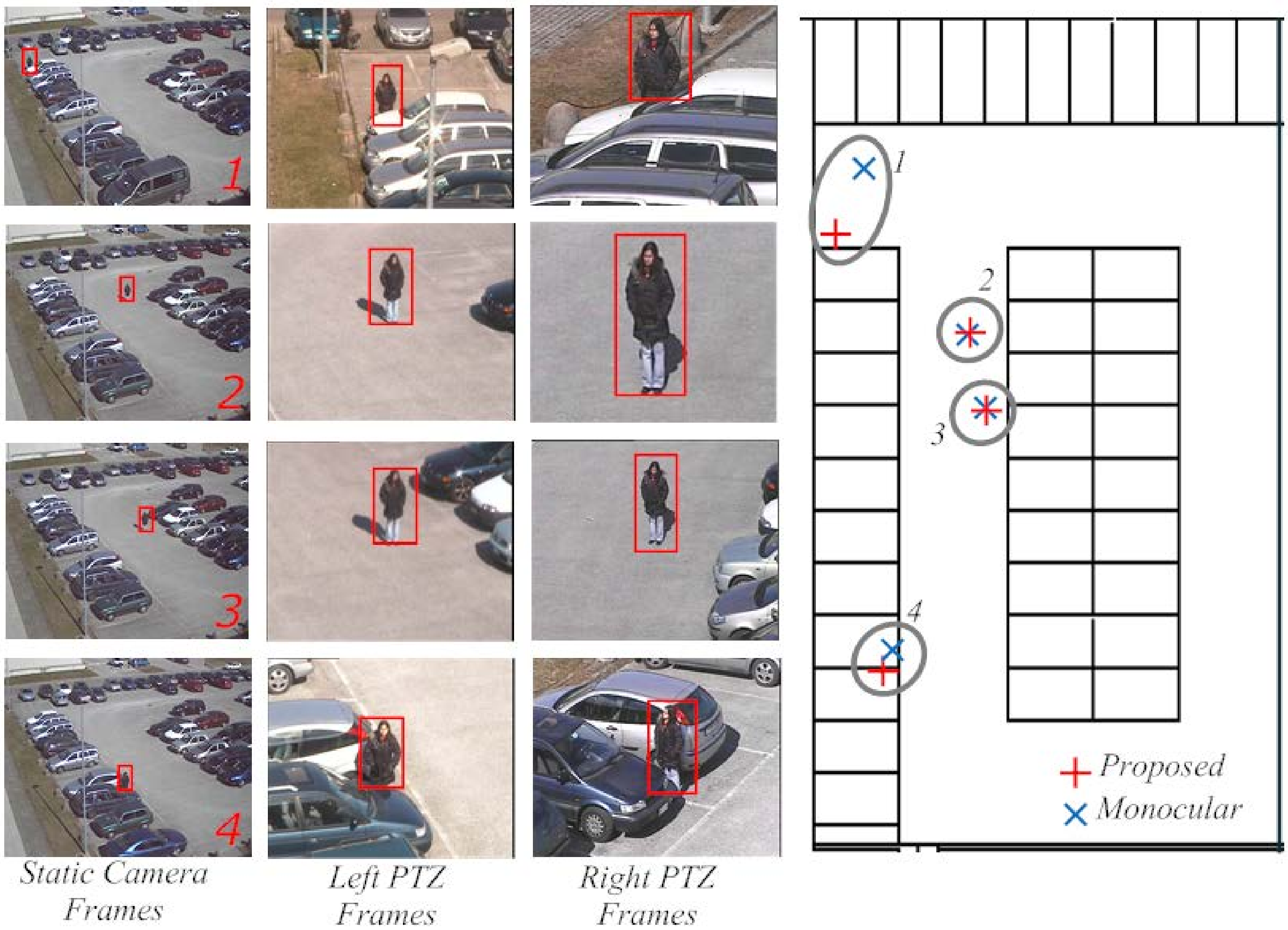
ONLINE OPERATIONS

- Cameras redirected on the target
- Homography matrices are computed using NN interpolation
- Finally the target is localized by using its 3D position obtained by stereo vision.

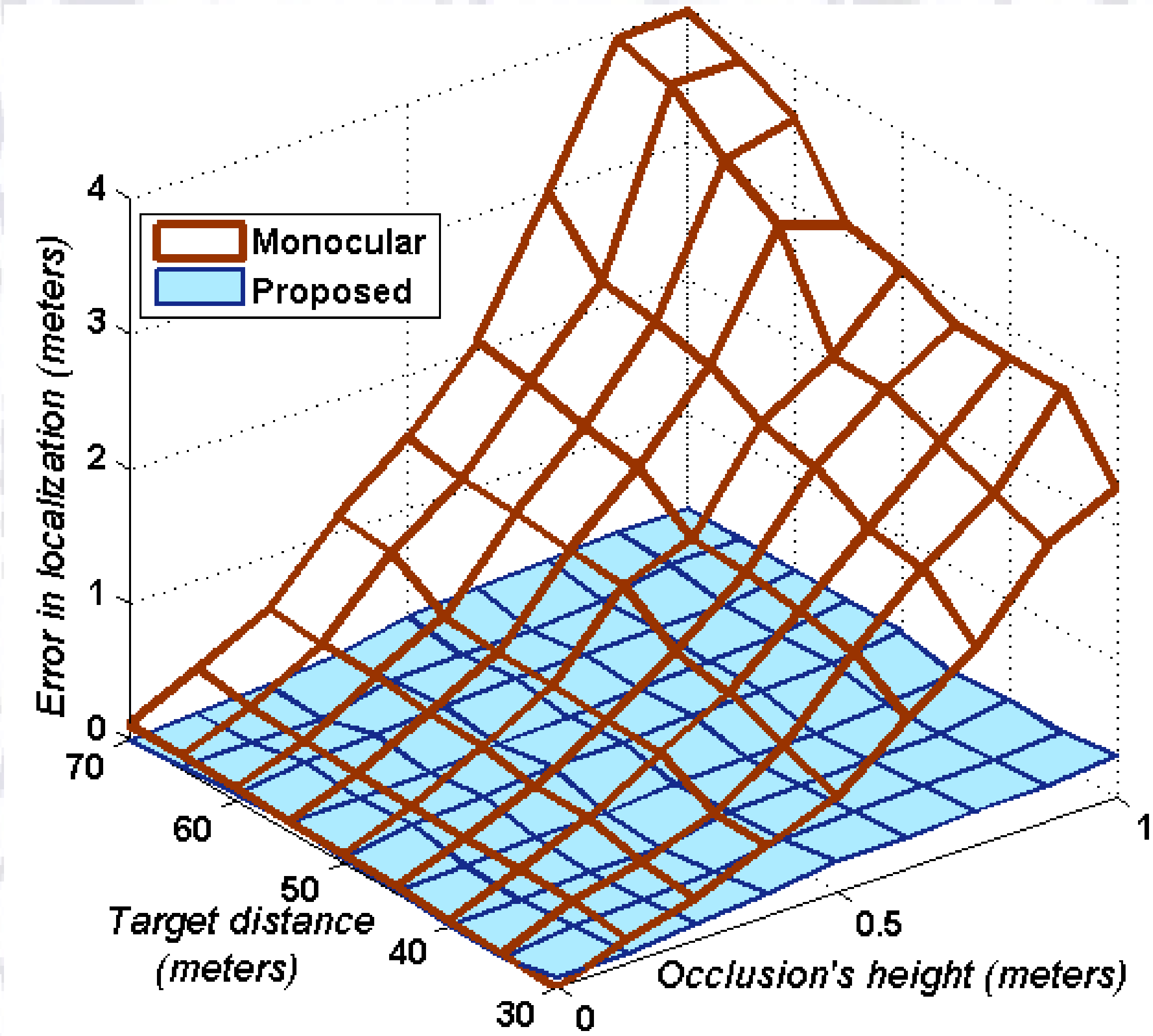
Construction of LUT

- Sample the different pan and tilt angles for the whole pan and tilt ranges for the left and right PTZ camera into equal intervals.
- Compute the possible rectification transformations pairs for these image pairs of stereo images.
- Store all these pairs in a LUT where the indices are the Pan and Tilt angles of the two cameras and the related information is the rectification.

results for localization



Dual-PTZ camera based stereo vision for object's localization

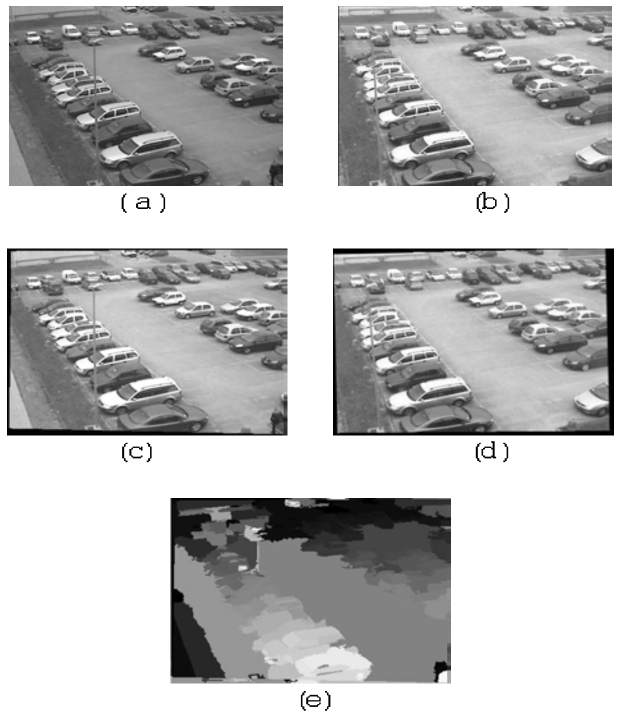


Steps:

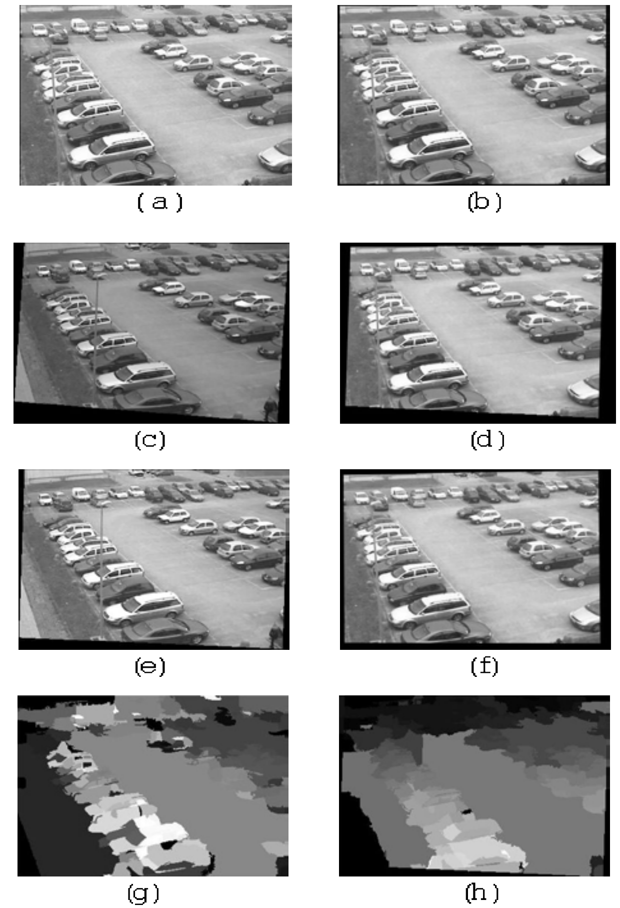
- Grab the sequences from both (static and PTZ) cameras.
- Rectify the corresponding stereo frames.
- Normalize the intensities in stereo frames using linear regression technique.
- Perform SSD measure for disparity estimation.

Range-image estimation using heterogeneous stereo vision

(Focal ratio: 1.0 (left) & 0.97 (right))

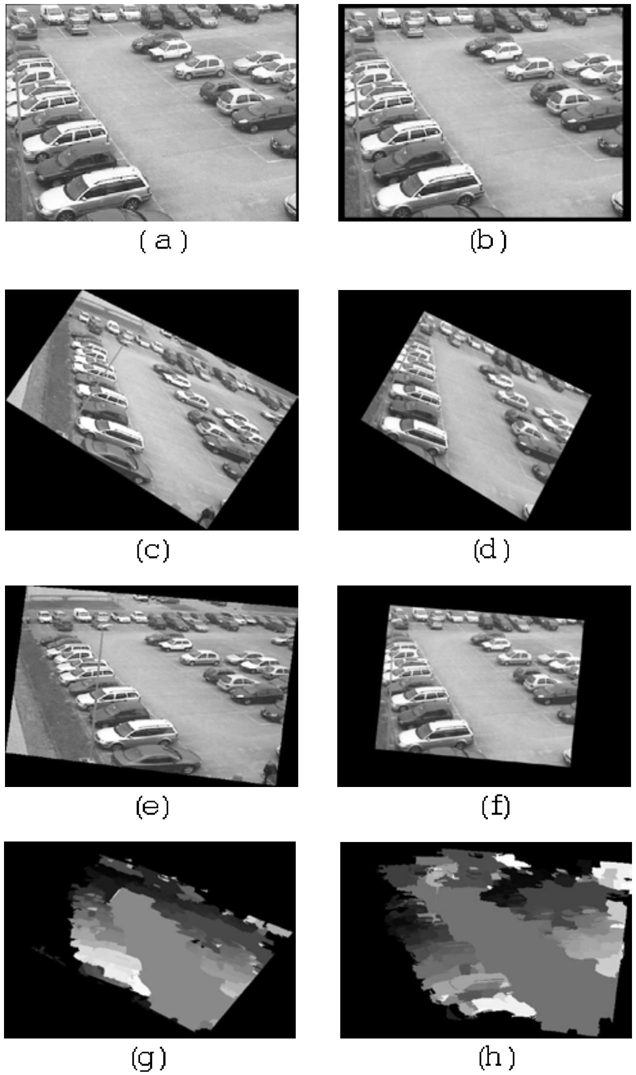


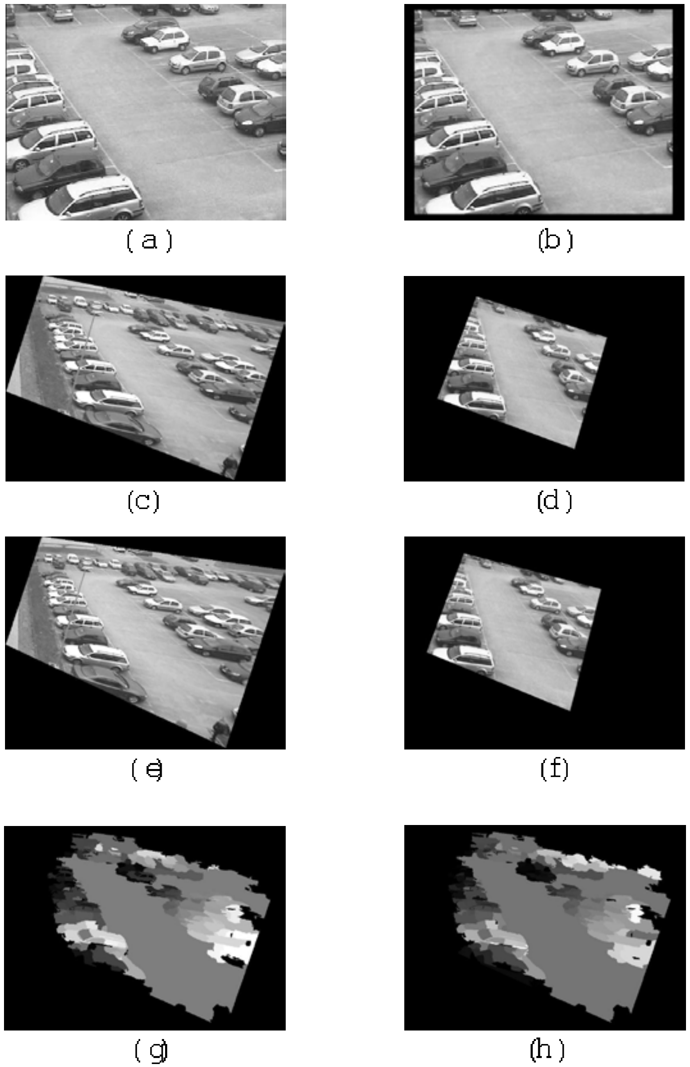
Stereo pairs (a & b), rectified pairs (c & d)
Range-image (e)



Results cont.:

Focal Ratio- 0.94 (left), 0.90 (right)





Stereo Vision Bibliography

1. Abidi, B., Koschan, A., Kang, S., Mitckes, M., Abidi, M., 2003. Automatic Target Acquisition and Tracking with Cooperative Static and PTZ Video Cameras. Kluwer Academic, Ch. Multisensors Surveillance Systems: The Fusion, pp. 43–59.
2. Brown, M., Burschka, D., Hager, G., 2003. Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(8), 993–1008.
3. Chen, C. H., Yao, Y., Page, D., Abidi, B., Koschan, A., Abidi, M., 2008. Heterogeneous fusion of Omni-directional and PTZ cameras for multiple object tracking. IEEE Transaction on Circuit and Systems for Video Technology 18(8),, 1052–1063.
4. Foresti, G., Micheloni, C., Piciarelli, C., 2005. Detecting moving people in video streams. Pattern Recognition Letters 26, 2232–2243.
5. Fusiello, A., Israra, L., 2008. Quasi epipolar unclaibrated rectification. In: IEEE Int. Conf. on Image processing (ICPR). pp. 1–4.
6. Hart, J., Scassellati, B., Zucker, S., 2008. Epipolar geometry for humanoid robotic heads. In: International Cognitive Vision Workshop. pp. 24–36.
7. Kumar, S., Micheloni, C., Piciarelli, C., 2009. Stereo localization using dual PTZ cameras. Computer Analysis of Images and Patterns (LNCS-Springer, 1061–1069.

Stereo Vision Bibliography

8. Kumar, S., Micheloni, C., Piciarelli, C., Foresti, G. L., 2009. Stereo localization based on network's uncalibrated camera pairs. In: Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 502–507.
9. Kumar, S., Micheloni, C., Piciarelli, C., Foresti, G. L., 2010. Stereo rectification of uncalibrated and heterogeneous images. Pattern Recognition Letters 31, 1445–1452.
10. Micheloni, C., Foresti, G., Snidaro, L., 2005. A network of co-operative cameras for visual surveillance. In: IEE-proc. Vis. Image Signal Process. Vol. 152(2). pp. 205–212.
11. Morris, B., Trivedi, M., 2008. A survey of vision-based trajectory learning and analysis for surveillance. IEEE Transactions on Circuits and Systems for Video Technology 18(8), 1114–1127.
12. Wan, D., Zhaou, J., 2008. Stereo vision using two PTZ cameras. Computer Vision and Image Understanding 112(2), 184–194.
13. Wan, D., Zhou, J., 2009. Multiresolution and wide-scope depth estimation using a dual-PTZ camera system. IEEE Transaction on Image Processing 18(3), 677–682.

How can the PTZs monitor better?

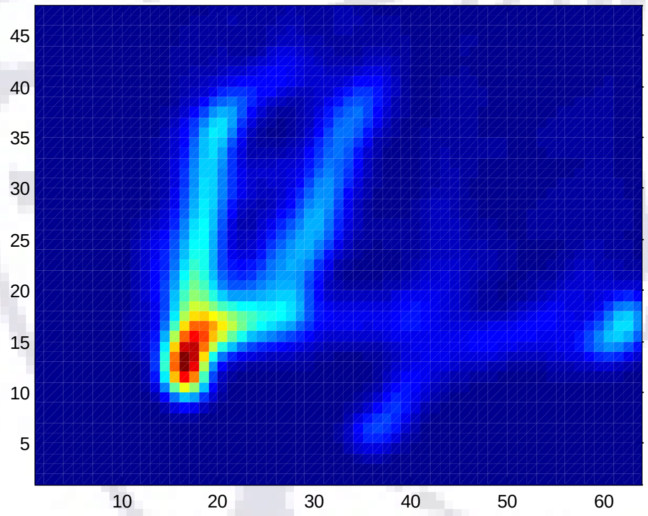
Optimization of the area coverage by
selecting the PTZs fields of view

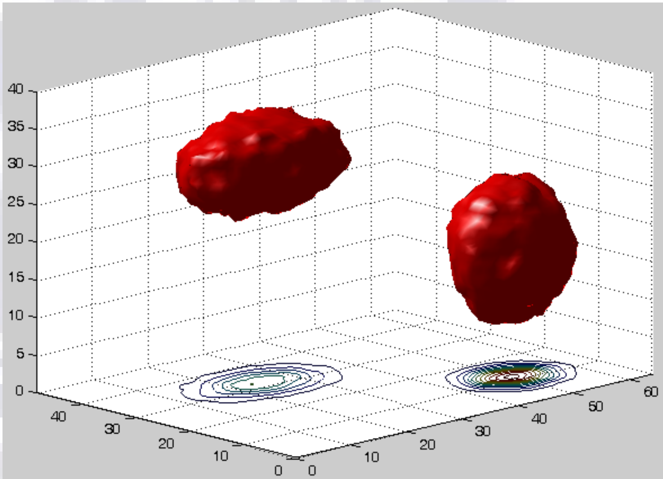
Motivations

- Nowadays large camera networks are becoming more and more popular, e.g. for surveillance systems, traffic monitoring, etc.
- However, the performance of the system will be suboptimal if the cameras are observing the wrong zones
- How can we achieve an optimal coverage of the observed environment?

- A method for optimizing the configuration (pan, tilt and zoom values) of a network of PTZ cameras could be interesting

- Optimize the coverage of the observed area according to a *relevance map*, denoting which zones are more important than others
- The exact definition of the relevance maps depends on the specific application. Relevance maps can either be 2D or 3D

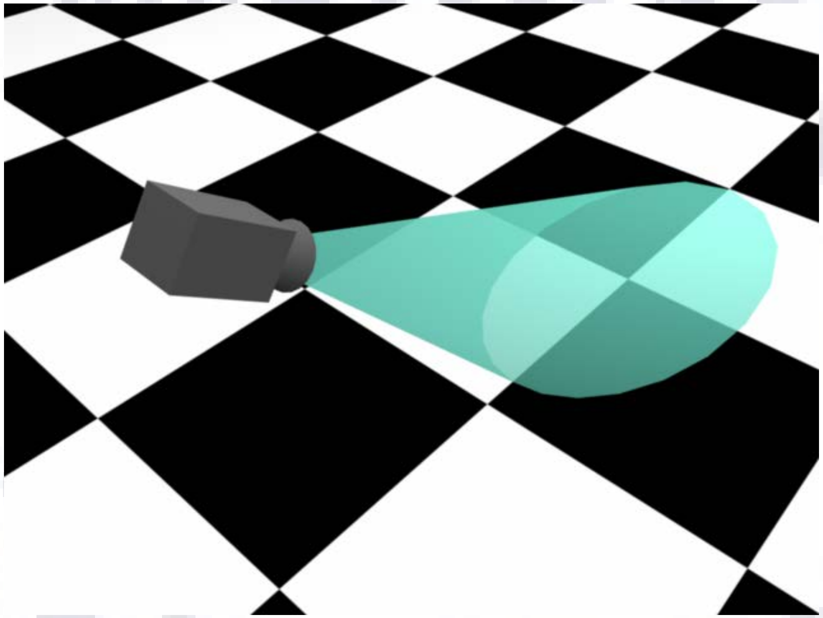




Problem formulation

- Approximate the zone observed by a camera with a *cone of view.* The *observation function* denotes if a point x falls within the cone of view of a camera with configuration parameters Θ

$$\gamma_k(\mathbf{x};\Theta) = \begin{cases} 1 & \text{if } \mathbf{x} \in \text{ cone-of-view of camera } k \\ 0 & \textit{otherwise} \end{cases}$$



Problem formulation / 2

- The score function measures the total coverage of a point x, giving higher scores to points observed by more than one camera and to points with higher relevance

$$f(\mathbf{x}; \Theta, C) = \left(\sum_{k=1}^{K} c_k \gamma_k (\mathbf{x}; \Theta) \right)^{\omega(\mathbf{x})}$$

- The relevance of a point is given by a relevance map

$$w(\mathbf{x}) \geq 0 \; \forall \mathbf{x} \in \mathrm{H}$$

- Each camera receives an importance score

$$c_k \in C$$

Problem formulation / 3

- The global score function in case H is a discrete set of points:

$$\Lambda(\Theta, C) = \prod_{\mathbf{x} \in \mathcal{H}} f(\mathbf{x}; \Theta, C)$$

- The goal of the proposed method is to compute the unknowns Θ (pan ϕ, tilt θ, cone of view width ζ) and C (camera weights) such that the global score function is maximized

EM algorithm / 1

- The Expectation-Maximization algorithm finds a maximum likelihood estimate of the unobserved latent variables of a statistical model.

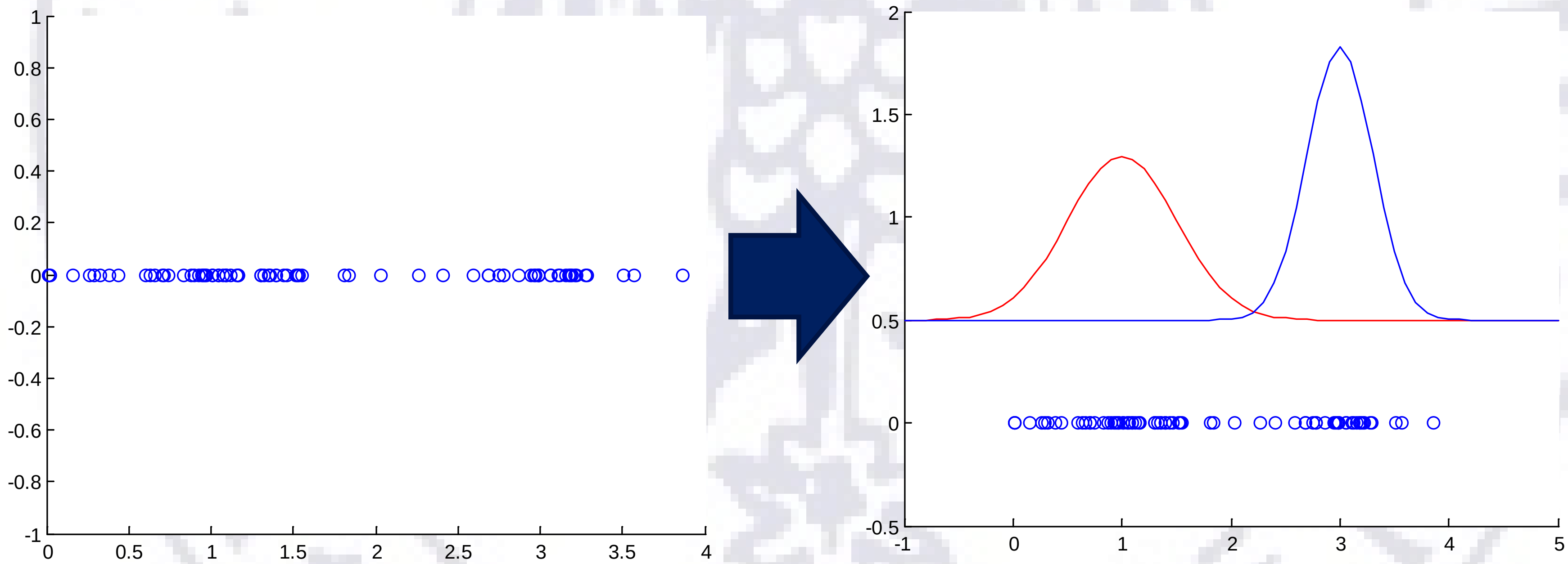
- A popular statistical model used with EM is the mixture of Gaussians:

$$p(x;\Theta) = \sum_{k=1}^{K} c_k G(x;\mu_k,\sigma_k)$$

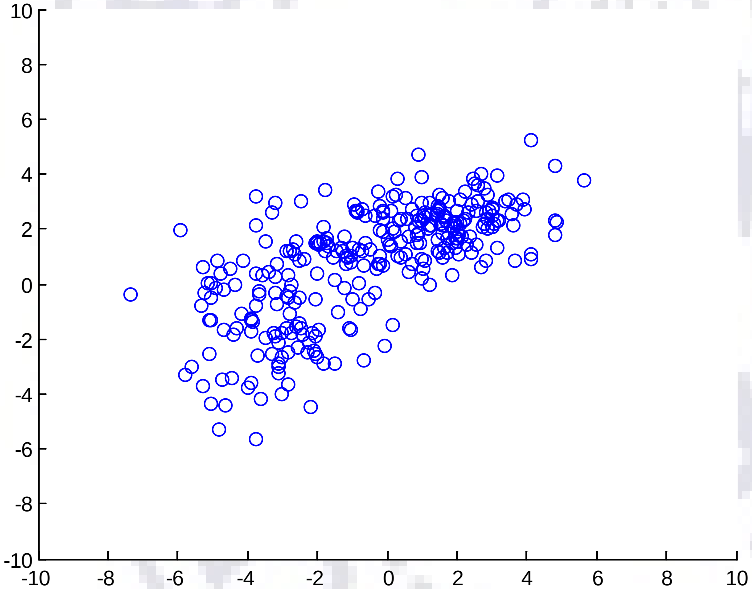
With G Gaussian function and $\Theta = (c_1, \mu_1, \sigma_1, \dots, c_K, \mu_K, \sigma_k)$

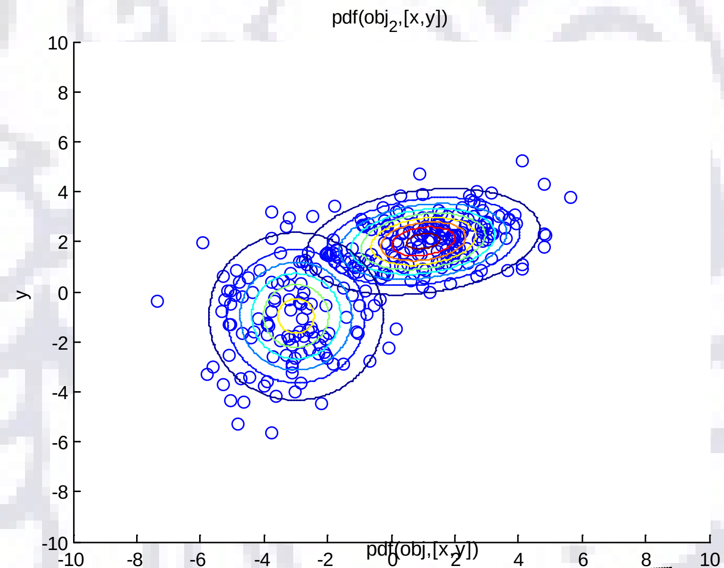
c_k are weights such that $c_k \geq 0, \quad \sum_{k=1}^{K} c_k = 1$

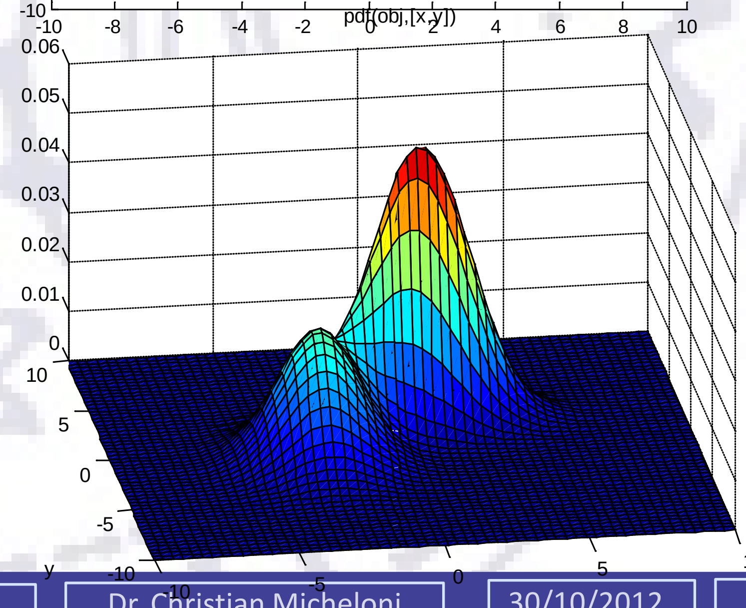
- Given a set of values $X = \{x_1, \ldots, x_N\}$ EM thus searches for the parameters that better describe their distribution in terms of a Mixture-of-Gaussians model



- Also works with bivariate or multivariate Gaussian distributions:







- the likelihood is expressed as:

$$\Lambda(X,\Theta) = \prod_{n=1}^{N} p(x_n,\Theta)$$

In the case of mixture of Gaussians this becomes:

$$\Lambda(X,\Theta) = \prod_{n=1}^{N}\sum_{k=1}^{K} c_k G(x;\mu_k,\sigma_k)$$

and the log-likelihood is:

$$\lambda(X,\Theta) = \sum_{n=1}^{N}\log\sum_{k=1}^{K} c_k G(x;\mu_k,\sigma_k)$$

EM algorithm/5

- Goal: find parameters Θ that maximize the log-likelihood
- Solution: use the Lagrange multipliers method and set to zero the partial derivatives of $\lambda(X,\Theta)$ with respect to each parameter μ_k, σ_k, c_k
- The solving equations, in the simplified case of mixture of isotropic bivariate Gaussian functions, are:

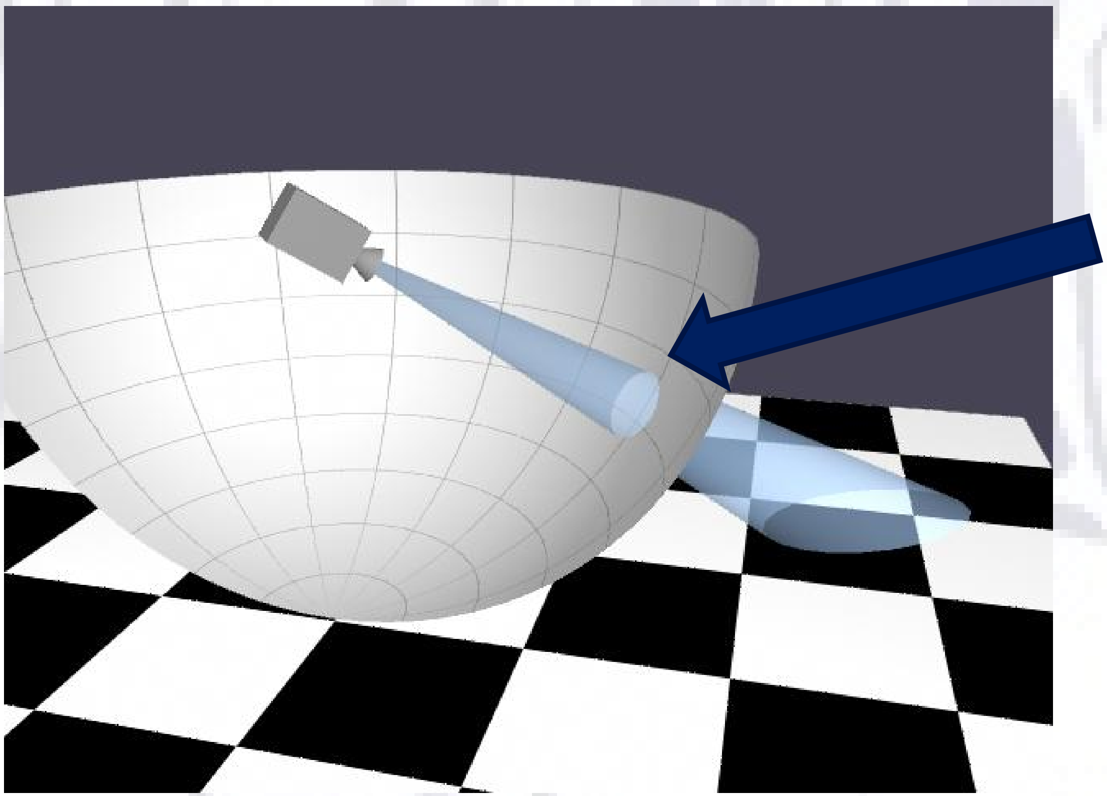
- $\mu_k = \frac{\sum_{n=1}^{N} p(k|n) x_n}{\sum_{n=1}^{N} p(k|n)}$
- $\sigma_k^2 = \frac{\sum_{n=1}^{N} p(k|n) ||x_n - \mu_k||^2}{2 \sum_{n=1}^{N} p(k|n)}$
- $c_k = \frac{1}{N} \sum_{n=1}^{N} p(k|n)$

$$p(k|n) = \frac{c_k G(x_n; \mu_k, \sigma_k)}{\sum_{z=1}^{K} c_z G(x_n; \mu_z, \sigma_z)}$$

mutually dependent,
iterate until convergence

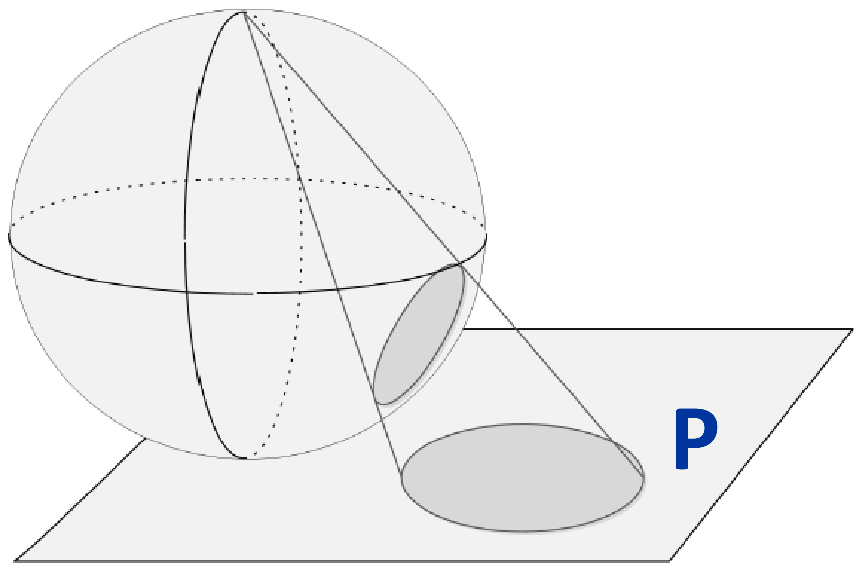
Dropping the distance

- The observation function γ_k does not depend to the distance of point x from the camera. Thus, we can drop a dimension and project all the observed points on the surface of a sphere centered on the camera.



Each point can now be described with two angular coordinates. The observed region collapses on a circle on the surface of the sphere.

Projecting the circles on a plane



The circles can be projected on a plane P using a stereographic projection, a circle-preserving projection widely used in cartography.

$$\begin{cases} \phi_k = \arctan\left(\frac{y-Y_k}{x-X_k}\right) \\ \theta_k = \arctan\left(\frac{\sqrt{(x-X_k)^2+(y-Y_k)^2}}{z-Z_k}\right) \end{cases}$$

angles

$$\begin{cases} u_k = 2\tan(\theta_k/2)\cos\phi_k \\ v_k = 2\tan(\theta_k/2)\sin\phi_k \end{cases}$$

coordinates in P

Working on the new plane P

- The equivalent of the observation function on the plane P is thus a circle:

$$\Gamma_k(u,v;\Theta) = \begin{cases} 1 & \text{if } (u-\mu^{\Theta}_{u,k})^2 + (v-\mu^{\Theta}_{v,k})^2 \le (\sigma^{\Theta}_k)^2 \\ 0 & otherwhise \end{cases}$$

- We now approximate this circle with a bivariate, isotropic Gaussian funciton. This approximation helps keeping the optimization problem tractable and intuitively gives more importance to central parts of the image

$$\Gamma_k(\mathrm{x};\Theta) \approx G_k(\mathrm{x};\Theta) = \frac{1}{2\pi{\sigma^{\Theta}_k}^2} e^{-\frac{\|\mathbf{x}-\mu^{\Theta}_k\|^2}{2{\sigma^{\Theta}_k}^2}}$$

- If M(x) is the stereographic projection of point x on the plane P, the score function can now be rewritten as

$$f(\mathbf{x};\Theta,C)=\left(\sum_{k=1}^{K}c_k G_k(M_k(\mathbf{x});\Theta)\right)^{\omega(\mathbf{x})}$$

- This leads to a new global score function to be maximized, whose logarithm is:

$$\lambda(\Theta,C)=\sum_{\mathbf{x}\in\mathcal{H}}\omega(\mathbf{x})\log\sum_{k=1}^{K}c_k G_k(M_k(\mathbf{x});\Theta)$$

The final solving equations (to be iterated until convergence) are:

$$\mu_k^{\Theta} = \frac{\sum_{\mathbf{x}\in\mathcal{H}} \omega(\mathbf{x})\ p(k|\mathbf{x})\ M_k(\mathbf{x})}{\sum_{\mathbf{x}\in\mathcal{H}} \omega(\mathbf{x})\ p(k|\mathbf{x})}$$

$$\sigma_k^{\Theta\,2} = \frac{\sum_{\mathbf{x}\in\mathcal{H}} \omega(\mathbf{x})\ p(k|\mathbf{x})\ \|M_k(\mathbf{x}) - \mu_k^{\Theta}\|^2}{2\sum_{\mathbf{x}\in\mathcal{H}} \omega(\mathbf{x})\ p(k|\mathbf{x})}$$

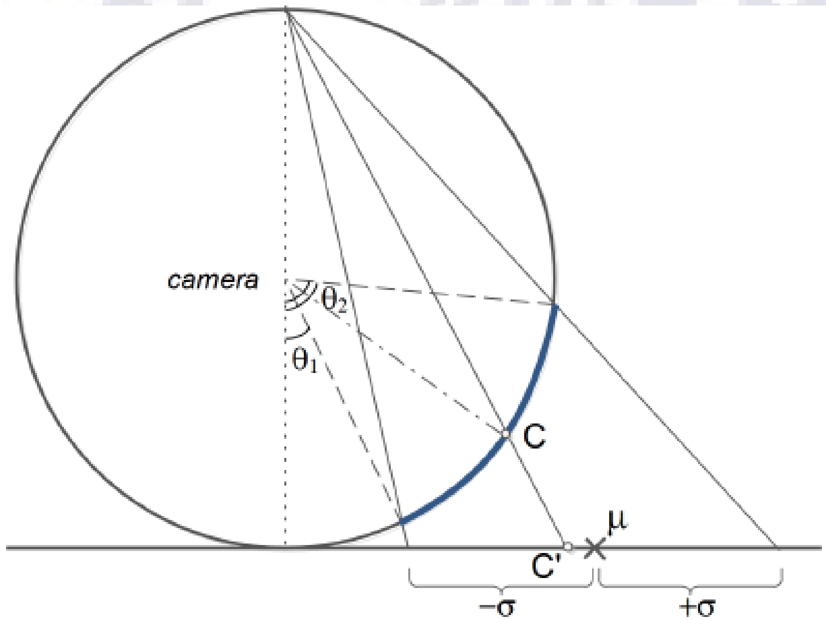
$$c_k = \frac{\sum_{\mathbf{x}\in\mathcal{H}} \omega(\mathbf{x})\ p(k|\mathbf{x})}{\sum_{\mathbf{x}\in\mathcal{H}} \omega(\mathbf{x})}$$

Maximization equations

$$\boxed{p(k|\mathbf{x}) = \frac{c_k\omega(\mathbf{x})G(M_k(\mathbf{x});\Theta)}{\sum_{z=1}^{K} c_z\omega(\mathbf{x})G(M_z(\mathbf{x});\Theta)}}$$

Expectation equation

- The results on plane P (in terms of centers and variances of the Gaussian functions) can be projected back using an inverse stereographic projection, in order to obtain the pan, tilt and zoom angles of each camera.
- Note: we cannot simply back-project the center μ of each Gaussian function, as centers are not preserved by stereographic projections.

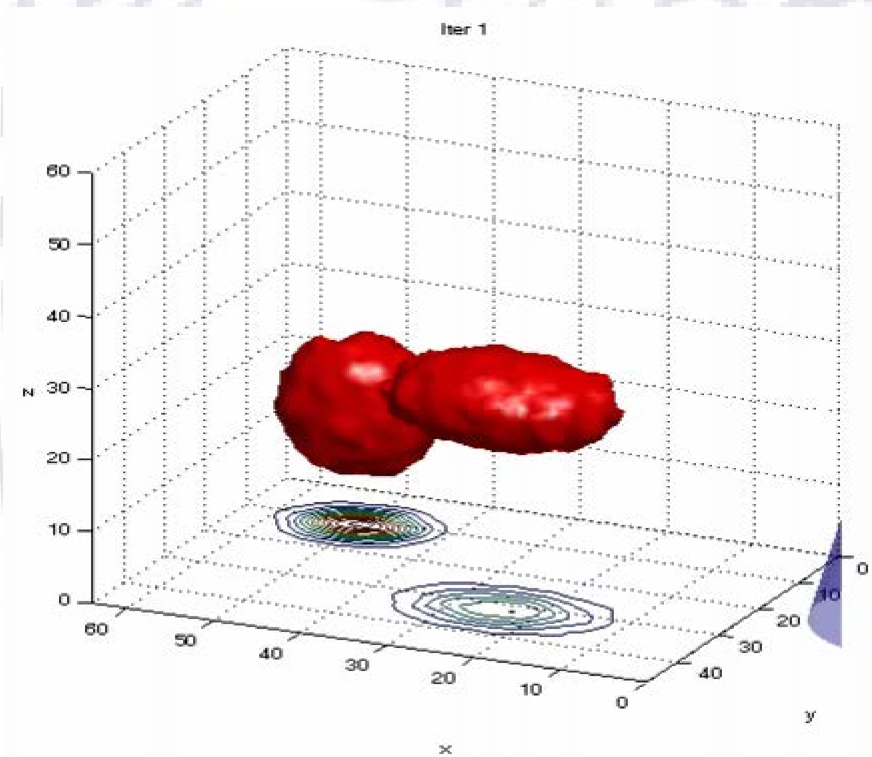


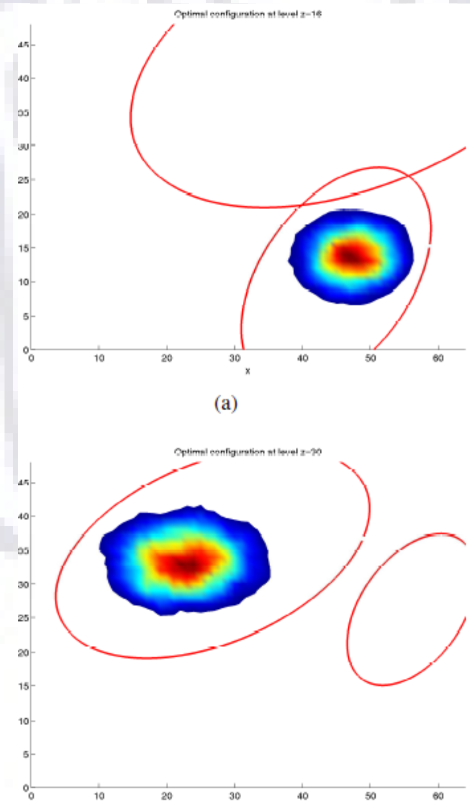
- The problem can be solved by first computing the two θ_1, θ_2 angles

$$\begin{cases} \theta_1 & = 2\arctan\left(\frac{||\mu_k||+\sigma_k}{2}\right) \\ \theta_2 & = 2\arctan\left(\frac{||\mu_k||-\sigma_k}{2}\right) \end{cases}$$

$$\begin{cases} \phi_k & = \arctan\left(\mu_{v,k}/\mu_{u,k}\right) \\ \theta_k & = (\theta_1+\theta_2)/2 \\ \zeta_k & = (\theta_1-\theta_2)/2 \end{cases}$$

- A simple example where the relevance map is a mixture of two Gaussian trivariate functions. The red surface is the isosurface of relevance w=0.1



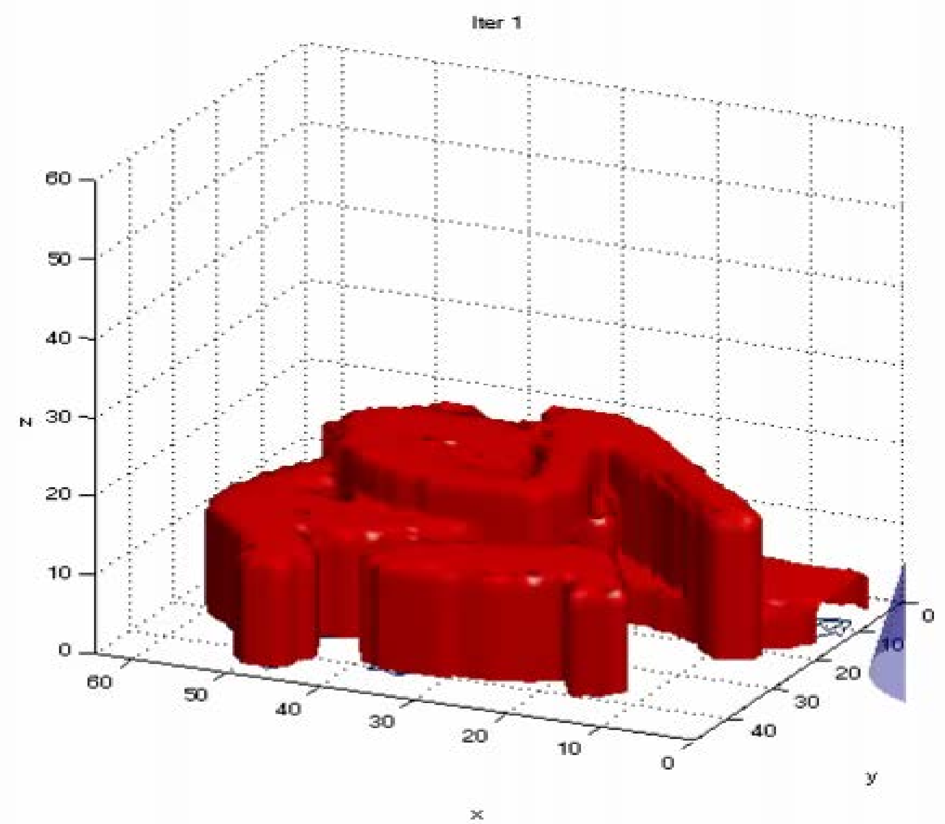


(a)

Final results at heights 16 and 30

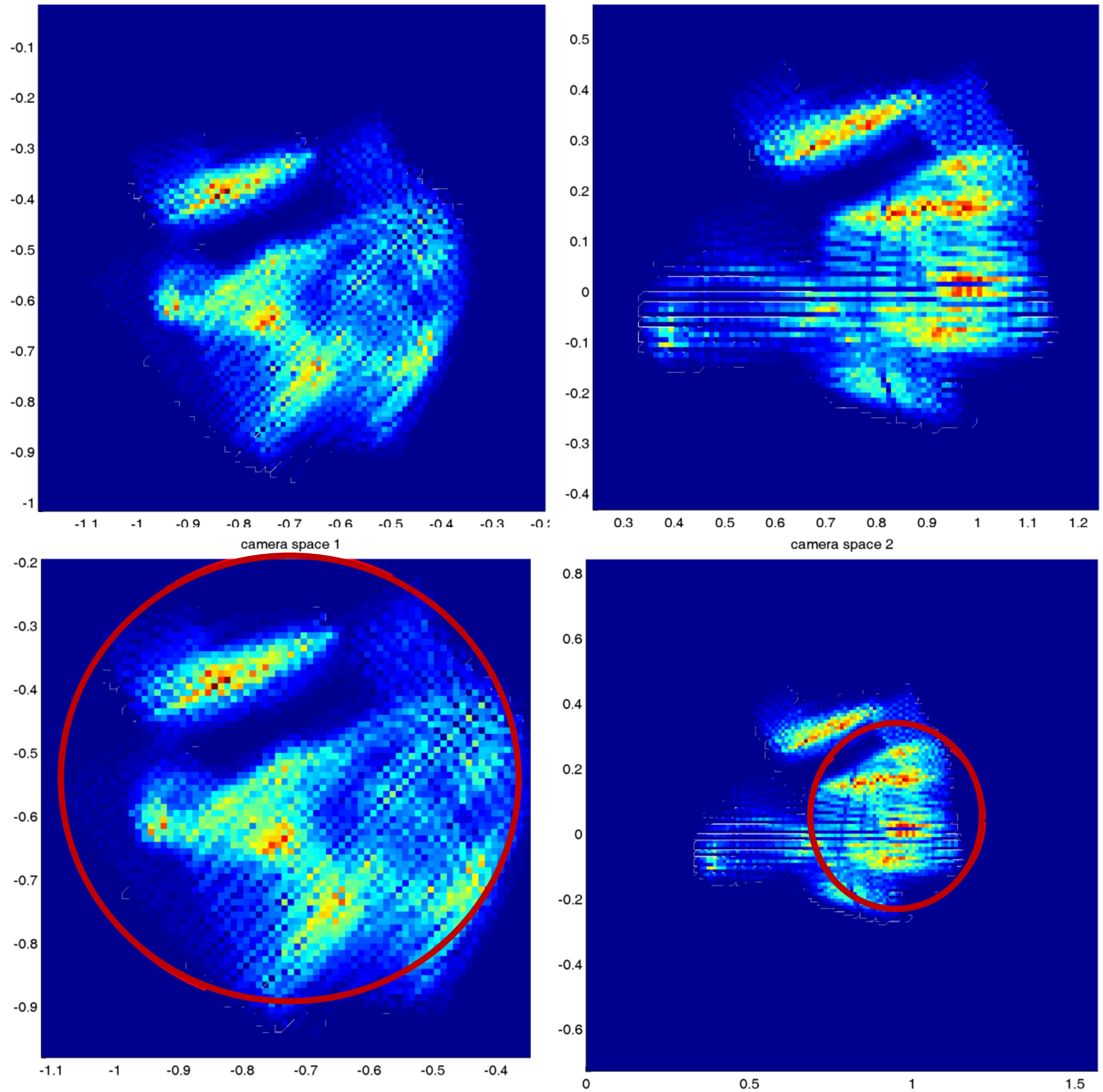
Results

- In this example the relevance map denotes the 3D location of moving objects accumulated over a long period of time

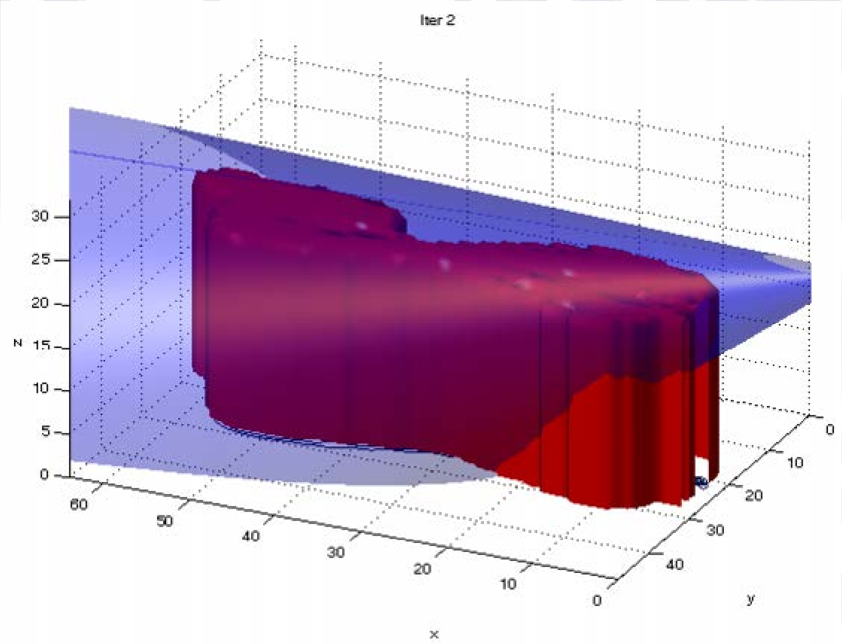


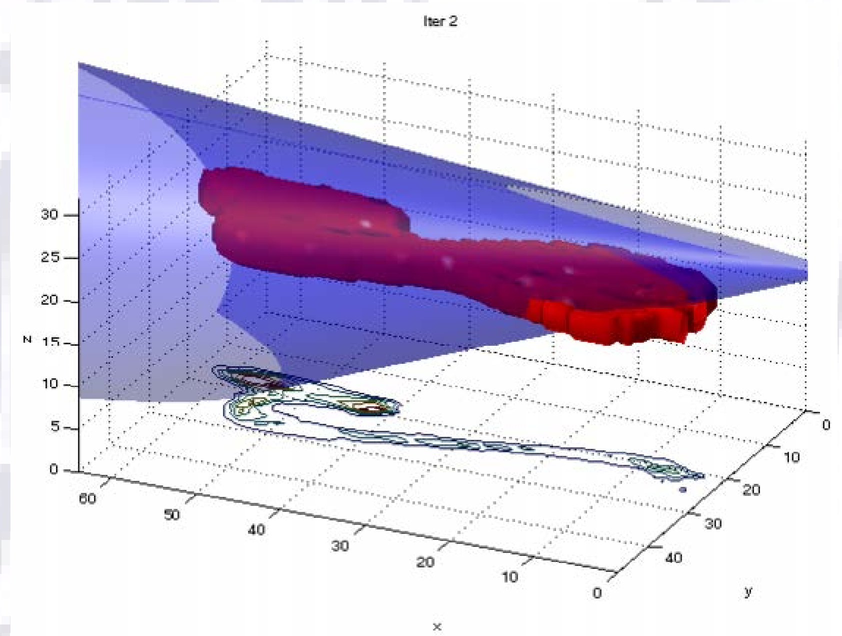
Results

- Relevance Maps computed for the cameras positions
- Coverage achieved by the two cameras



- 3D maps can help focusing only on specific regions. The data below are obtained mapping the 3D volume occupied by a moving person. If we consider only the upper region the camera can be more effectively focused on the head of the person.

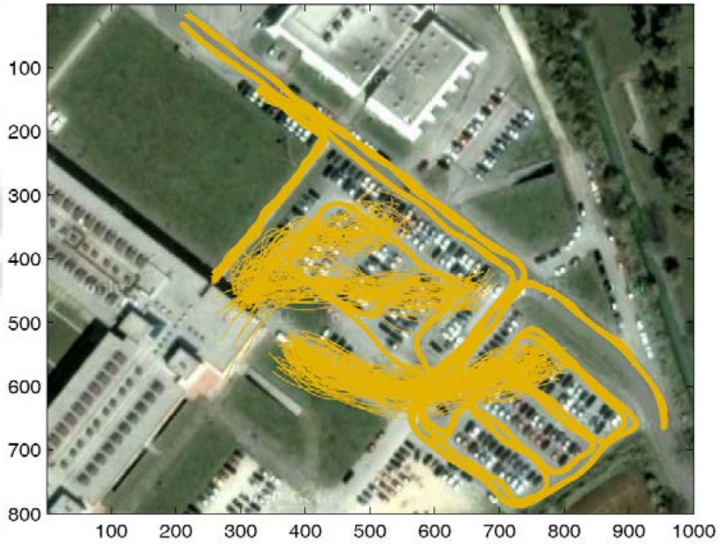




Performance

- In order to give a measurable performance parameter, the concept of *total coverage* can be adopted, defined as the ratio among the relevance of all the considered points covered by at least one camera and the total relevance.
- EM achieve a total coverage 93%

dataset	total coverage	dataset	total coverage
1	0.9693	11	0.9987
2	0.9998	12	0.9545
3	0.9723	13	0.9785
4	0.9923	14	0.9614
5	0.9977	15	0.9735
6	0.9824	16	0.9788
7	0.9775	17	0.9817
8	0.9891	18	0.9937
9	0.9815	19	0.9798
10	0.9304	20	0.9615



PTZ Reconfiguration Conclusions

- Activities can be detected and exploited to computed related 3D relevance maps.
- Relevance maps are projected in a space where the execution of EM algorithm is easier and more efficient.
- The reconfiguration algorithm optimally computes the best cameras parameters to best cover the monitored area considering the activities occurring in it.
- A subspace of the 3D environment can be selected to focus the reconfiguration on desired tasks or objects’ parts.