Monocular 3D Positioning and Tracking Systems

Tiago Gaspar

President of the jury

Professor Mário Figueiredo

Members of the jury

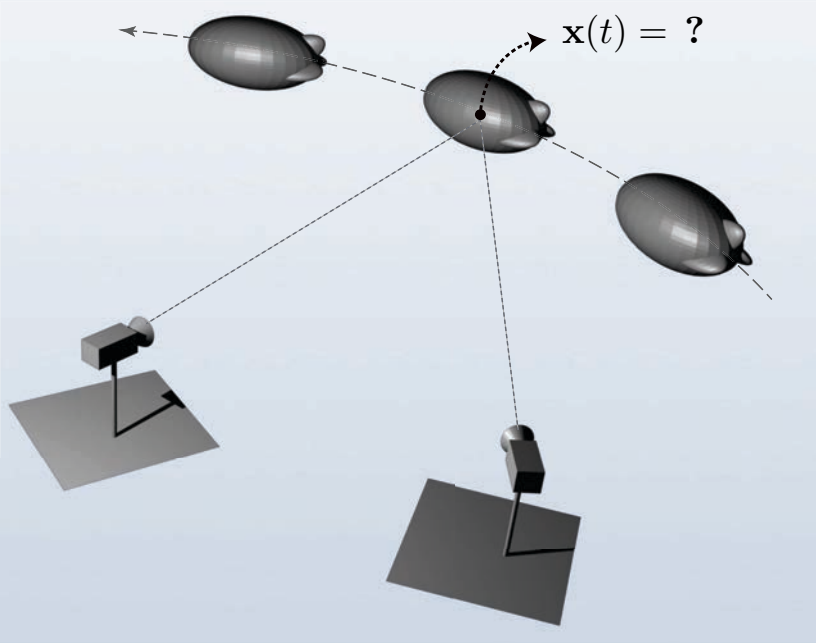
Professor Jorge Dias
Professor José Santos-Victor
Prof. João Rodrigues
Prof. Jacinto Nascimento
Prof. Paulo Oliveira

July 20, 2015

Motivation

Positioning and tracking consists in…

…using measurements provided by one or more sensors, at fixed locations or at moving platforms, to estimate the state of a moving object, which is usually composed of its position, velocity, and sometimes acceleration.

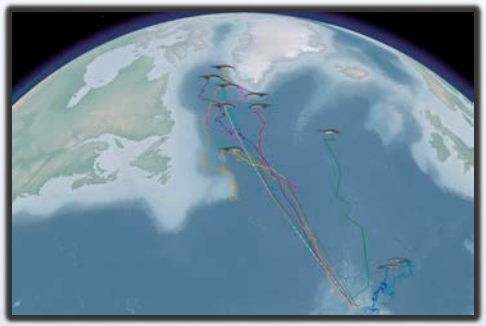


Motivation

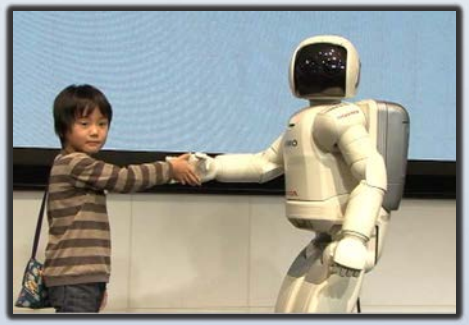
Positioning and tracking systems have several applications

Security and surveillance

Trajectory determination



Air traffic control



Human-computer interaction

Location-based services

Motivation

Outdoors: NAVSTAR-GPS, GLONASS, Galileo

FAIL (multipath & attenuation)

Main approaches:
Infrared
Ultrasound
Radio frequency
Computer Vision

Applications:
Industry
Hospitals
Houses
Shopping malls

Space | Indoors | Caves / tunnels | Underwater

Cons:
Computational power
Extraction of information

vs.

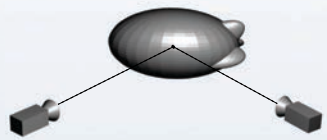
Pros:
Passive
Diversity of information
Cameras everywhere
Cheap

A reliable solution for indoor positioning is yet to be found…

Motivation

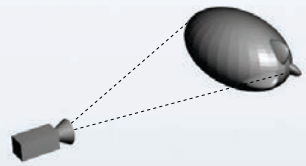
Vision-based 3D positioning systems with static cameras

Stereo systems



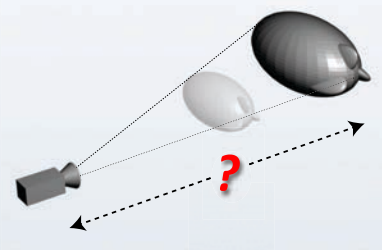
SOLVED
(Triangulation / Epipolar constraint)

Monocular systems with ***known*** target dimensions



SOLVED
(Range and bearing)

Monocular systems *with* ***unknown*** *target dimensions*



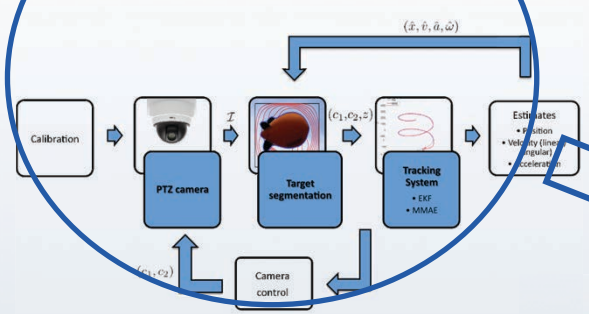
NOT SOLVED
(Bearing only)

With this work we want to answer the question:

What can be done in terms of 3D positioning when a single static camera is used and there is no information about (either the dimensions or the shape of) the target?

Main Contributions

Architecture for monocular 3D positioning and tracking systems

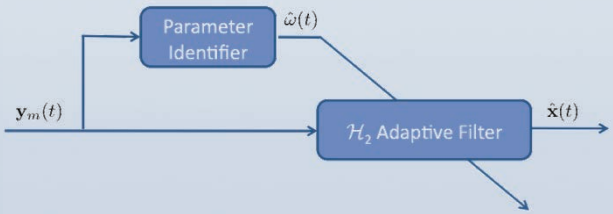


Monocular depth estimation for 3D positioning of targets with unknown dimensions

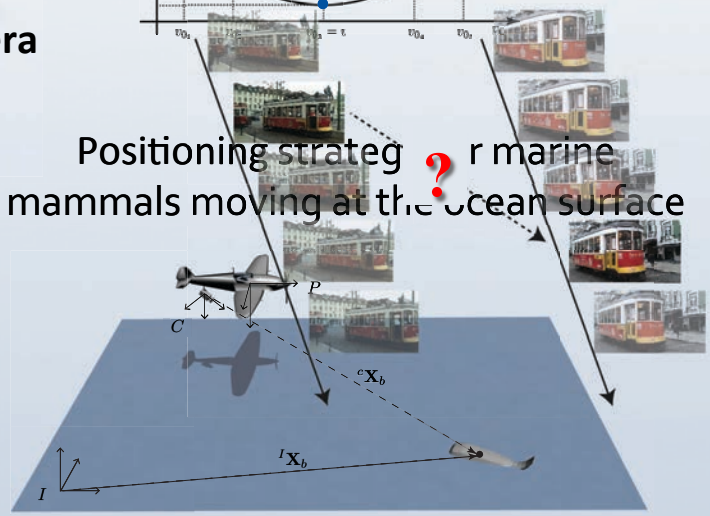
Synchronization of independently moving cameras

Extension to multi-camera configurations

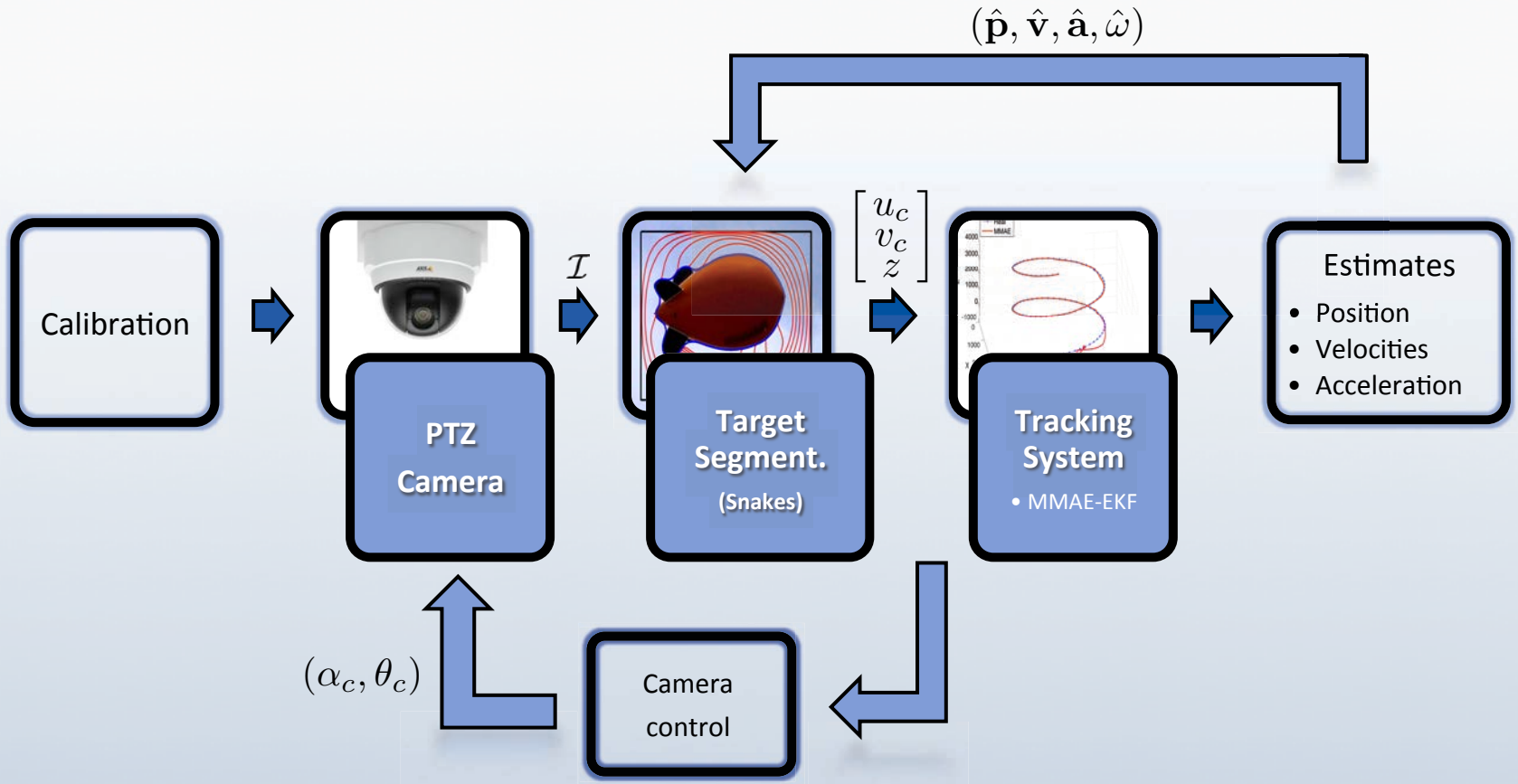
H_2 adaptive filter for 3D positioning of targets with an unstable model



Positioning strateg r marine mammals moving at the ocean surface ?



System Architecture



System Architecture

Target Model

3D Planar Constant-Turn Model

$$\dot{\mathbf{a}}(t) = -\omega^2 \mathbf{v}(t) + \mathbf{d}(t)$$

noise

ω	- angular speed
$\mathbf{v}(t)$	- linear velocity
$\mathbf{a}(t)$	- linear acceleration

Assumptions:

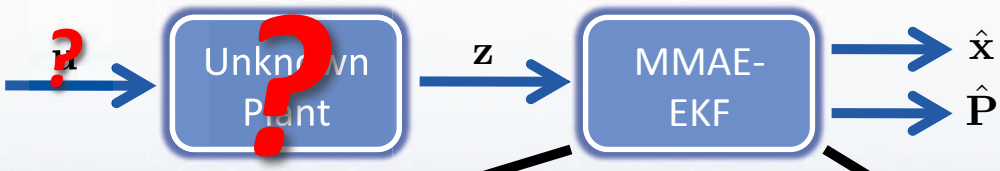
- constant speed
- constant angular velocity
- angular and linear velocity vectors are orthogonal to each other

~~The target moves in a plane orthogonal to the angular velocity vector, the maneuver plane~~

The target can span the 3D space

System Architecture

Tracking System



Target model:

$$\mathbf{z} = \begin{bmatrix} u_c \\ v_c \\ z \end{bmatrix}$$

Center of the target image (mean of the contour)

Distance of the target to the origin of the world reference frame (target dimensions & visual looming strategy)

$$\mathbf{x}_k = \mathbf{F}_T(\omega)\mathbf{x}_{k-1} + \mathbf{w}_{k-1}$$

Parameter dependent system (intricate dependence)

Observation model:

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k)$$

Pinhole model **(nonlinear)**

MMAE-EKF

- bank of EKFs, each one associated with a different angular speed
- the individual estimates are combined using a weighted sum (with the *a posteriori* hypothesis probabilities of each model as weights)

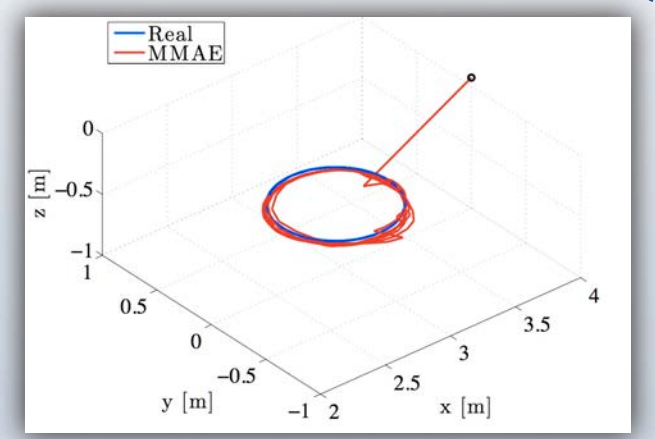
System Architecture

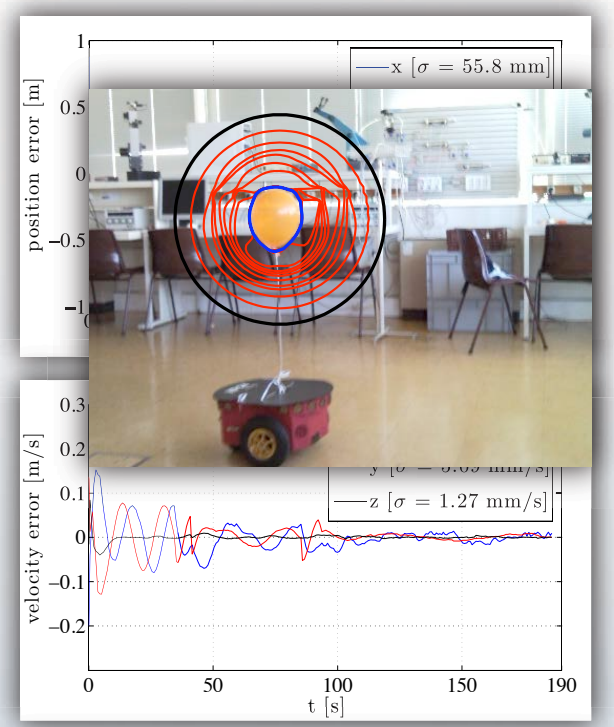
Experimental Results

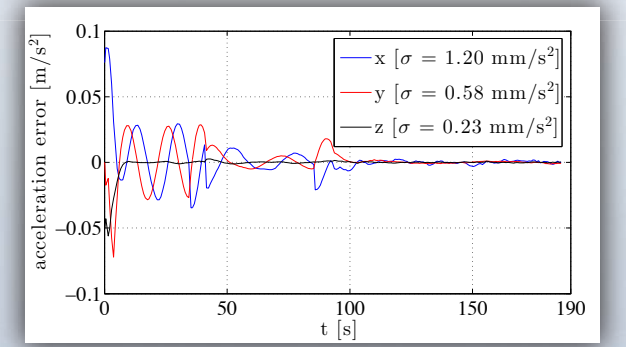
System description

- *Creative WebCam Live! Motion*
- Easily identifiable target
- Manual ground truth calibration
- Sampling interval: 0.5 sec

Estimation error



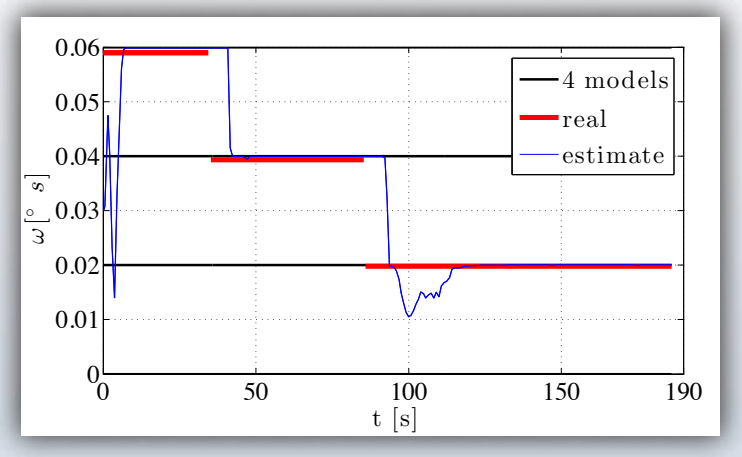




System Architecture

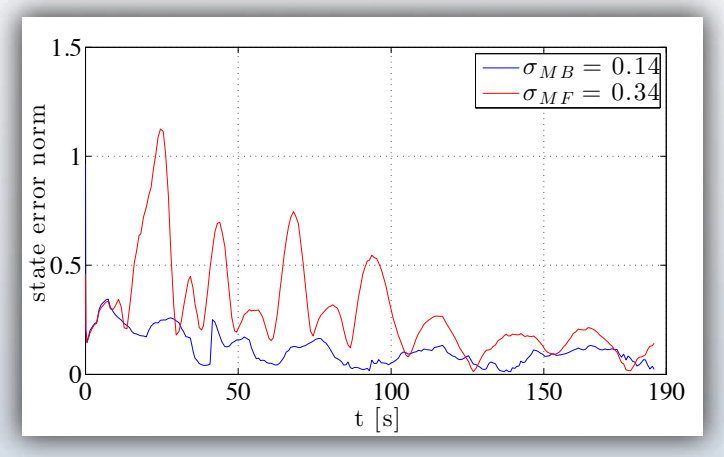
Experimental Results

Model identification



Model-Based vs Model-Free

(double integrator)



The system that we proposed achieves accuracies on the order of a few centimeters, but assumes that the dimensions of the target are known.

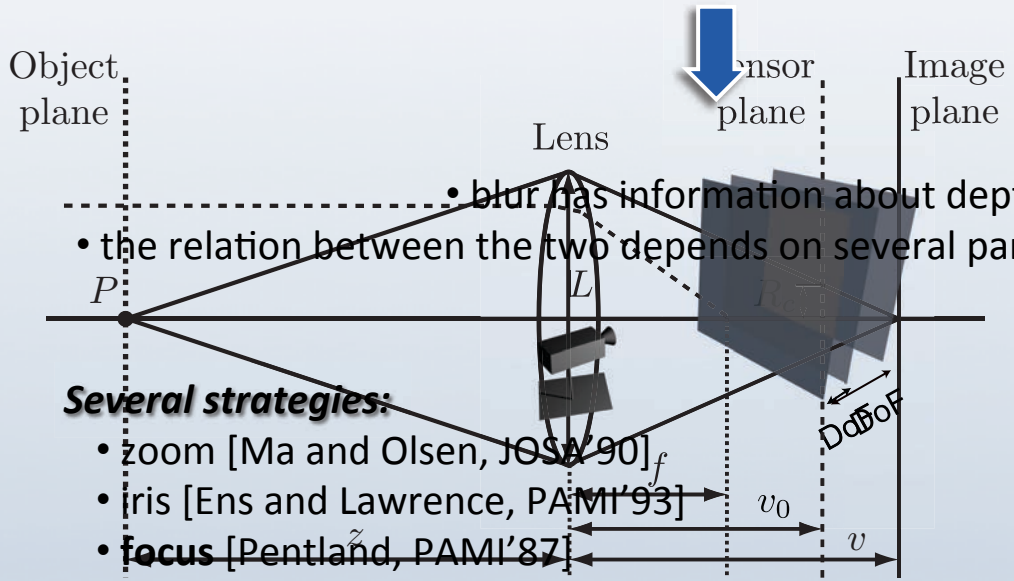
This is very restrictive!

What can be done when no information about the dimensions of the target is available?

Monocular Depth Estimation

Depth of field (of a camera with a given focus value):

- distance between the farthest and nearest planes, with respect to the camera, whose points appear with a satisfactory definition in the image, according to a given criterion
- lenses are unable to simultaneously focus planes at different depths



- blur has information about depth
- the relation between the two depends on several parameters of the camera

Several strategies:

- zoom [Ma and Olsen, JOSA'90]
- iris [Ens and Lawrence, PAMI'93]
- **focus** [Pentland, PAMI'87]

(Thin lens model)

Gaussian lens formula

$$z = \frac{fv}{v - f}$$

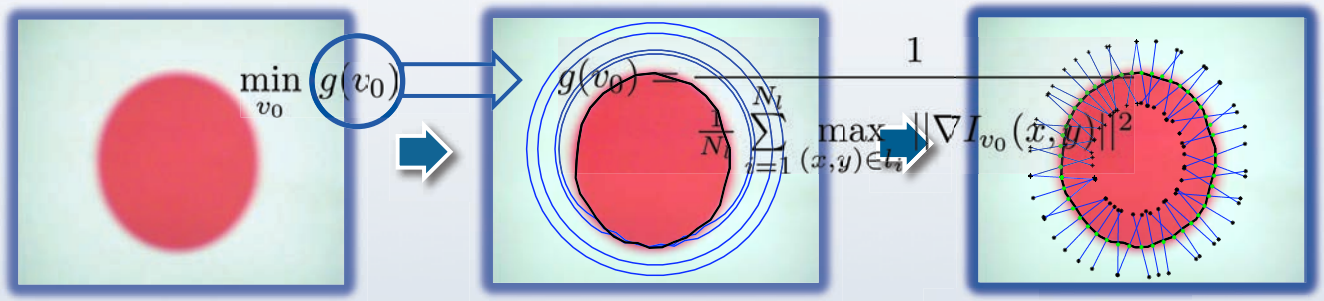
v ?

Monocular Depth Estimation

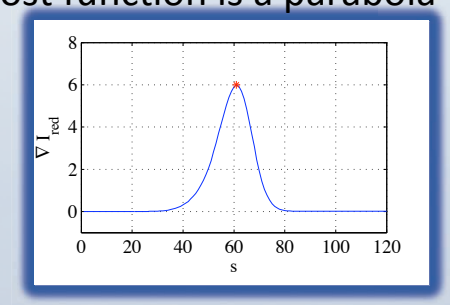
Depth from focus

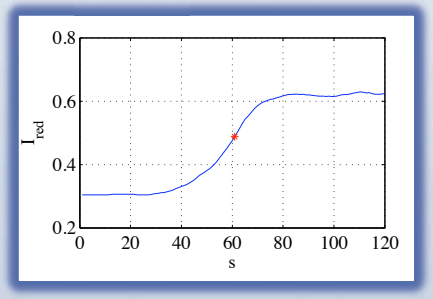
How to find the camera focus value that sharply focuses the target?

Define a quantity that depends on the amount of blur that "corrupts" the boundary of the target



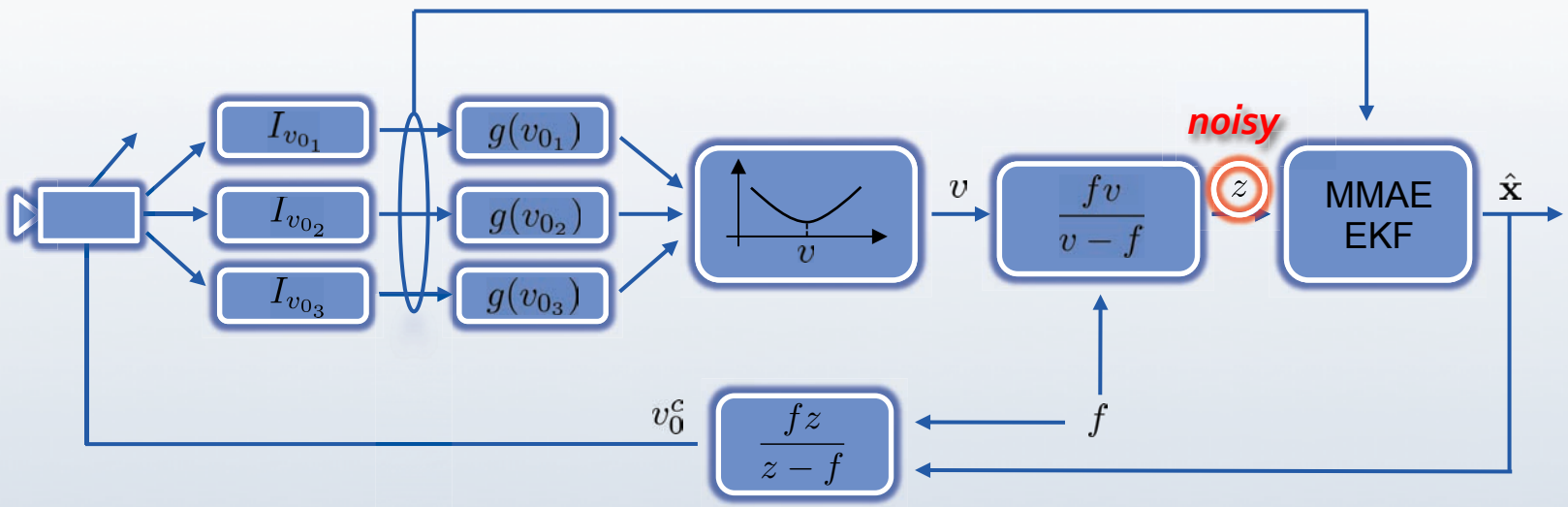
If the PSF is a constant intensity disk, then the exact model for the cost function is a parabola





Monocular Depth Estimation

System architecture



Depth filtering strategies

- Second-order complementary filter
- LPV observer

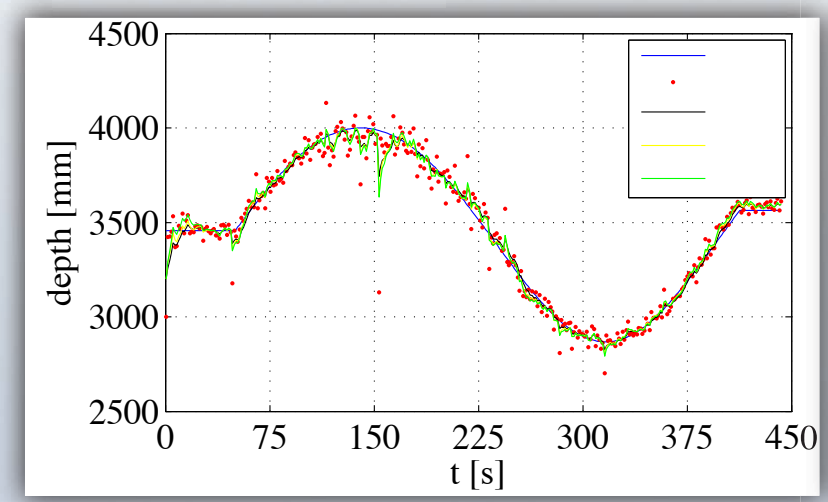
Monocular Depth Estimation

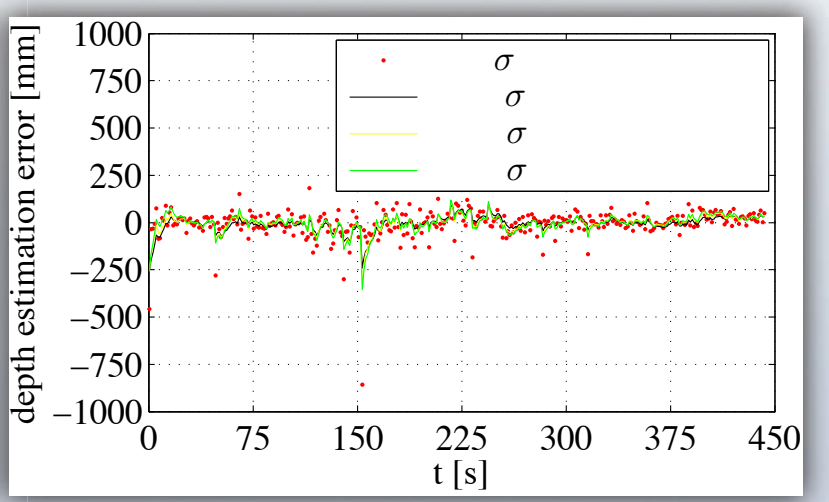
Experimental Results

System description

- AXIS 215 PTZ
  - maximum focal length: 45.6 mm (aperture: F2.7)
- Easily identifiable target
- Manual ground truth calibration
- Sampling interval: 1.2 sec

System performance





Monocular Depth Estimation

Experimental Results

3D Positioning System

The use of depth estimation strategies for targets with unknown dimensions widens the domain of applicability of the 3D positioning system.

However, there is still an important disadvantage: the MMAE-EKF does not have convergence guarantees…

What can be done to obtain a 3D positioning system with convergence guarantees?

H_2 Adaptive Filter

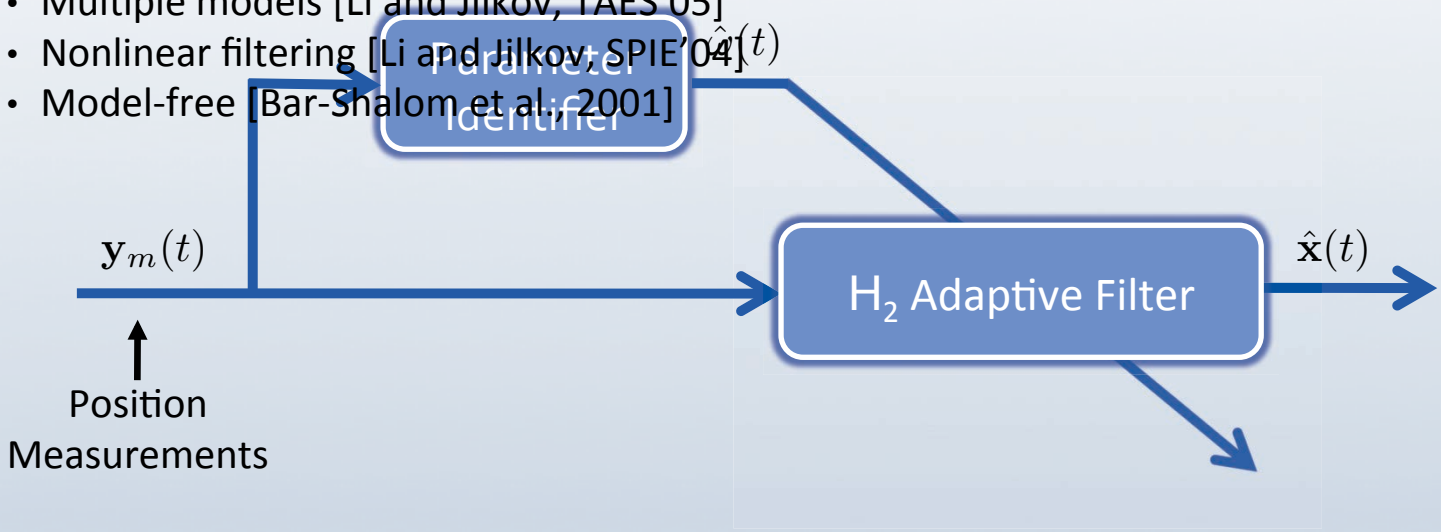
Overview

Target 3D model

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\omega)\mathbf{x}(t) + \mathbf{B}\mathbf{d}(t)$$

Typical approaches

- Multiple models [Li and Jilkov, TAES'05]
- Nonlinear filtering [Li and Jilkov, SPIE'04]
- Model-free [Bar-Shalom et al., 2001]



H_2 Adaptive Filter

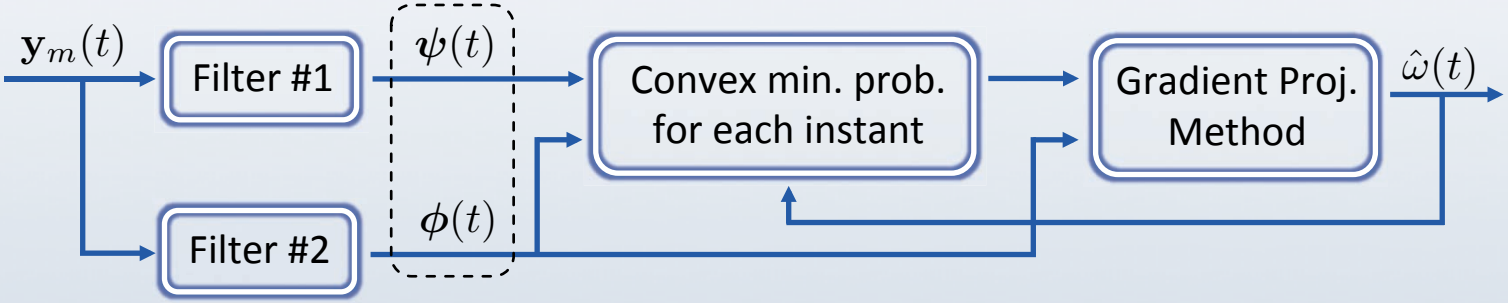
Parameter Identifier

3D Planar Constant-Turn Model:

$$\alpha = -\omega^2$$

$$\underbrace{\dot{\mathbf{a}}(t)}_{?} = \underbrace{\alpha}_{?}\underbrace{\mathbf{v}(t)}_{?} + \mathbf{d}(t)$$

Linear relation in the unknown parameter



Gradient projection method

- $\alpha \leq 0$
- derive convergence guarantees

H_2 Adaptive Filter

Design and Convergence

Filter structure

P. I. Filter gain

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{F}(\hat{\omega}(t))\hat{\mathbf{x}}(t) + \mathbf{L}(\mathbf{y}_m(t) - \mathbf{C}\hat{\mathbf{x}}(t))$$

within a given interval

Filter design

Find the gain that minimizes the maximum of the H_2 norm of the systems obtained from the disturbances to the position error, when the extremes of the interval are considered

Convergence guarantees

For bounded velocities and noise, and for some persistency of excitation conditions, the errors of both the parameter identifier and the H_2 filter

Converge to 0 exponentially fast	Are bounded after the transient
Absence of noise	Presence of noise

H_2 Adaptive Filter

Simulation Results

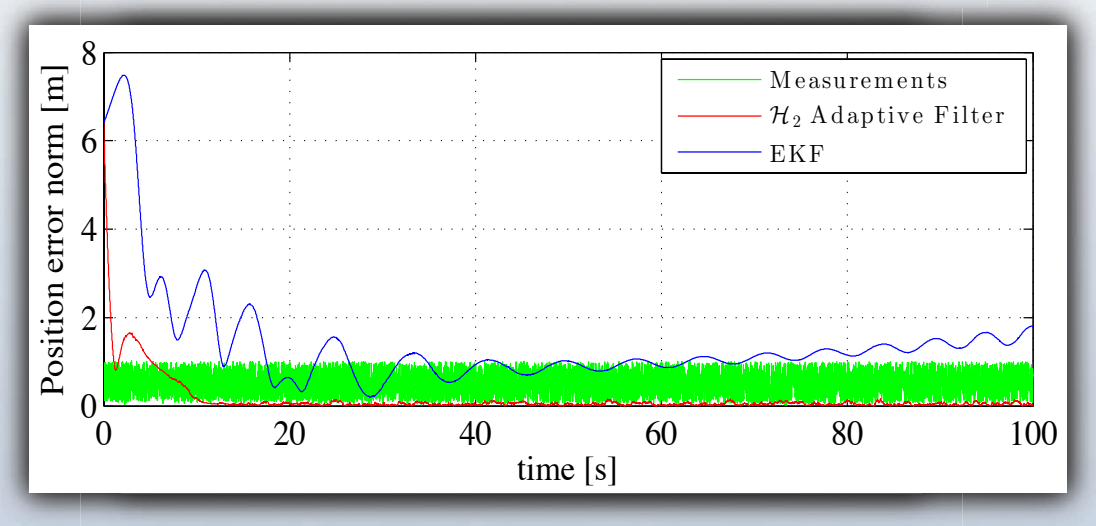
System specifications

- continuous-time simulations (1 PTZ camera)
- observation noise (uniformly distributed)
  - target center: [-10, 10] pixel
  - target depth: [-1, 1] m

System performance

- comparison with an EKF (augment the state with the angular speed)

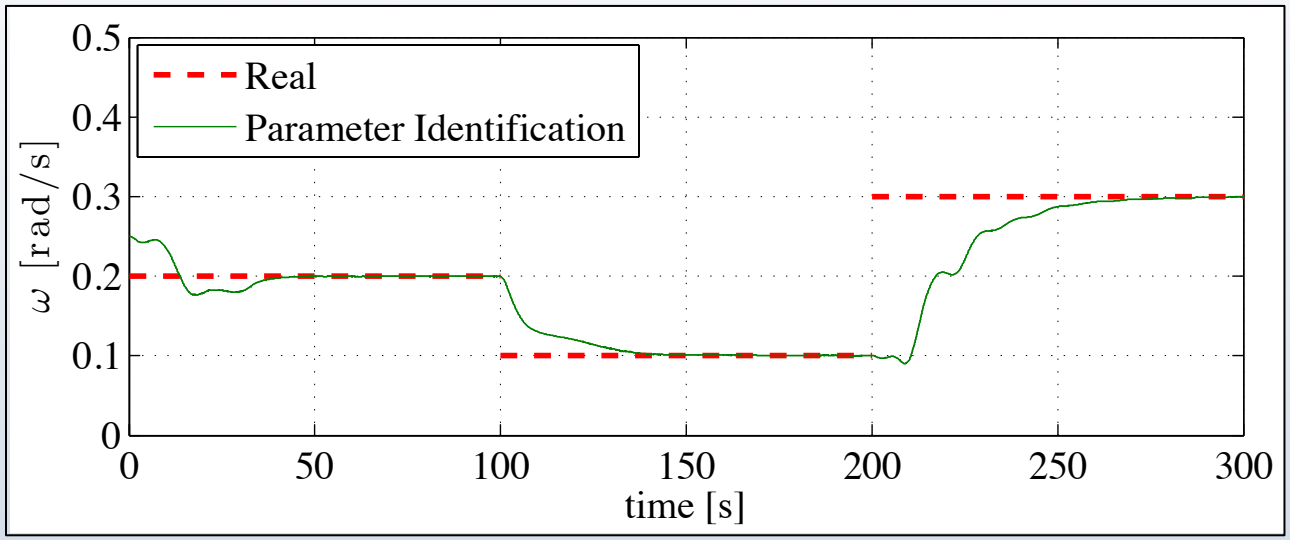
Straight line trajectory experiment ($\omega = 0$ *rad/s*)



H_2 Adaptive Filter

Simulation Results

Trajectory with 3 different angular speeds



The proposed positioning architecture is particularly appropriate for indoor applications, in which the target is usually within a small distance from the camera.

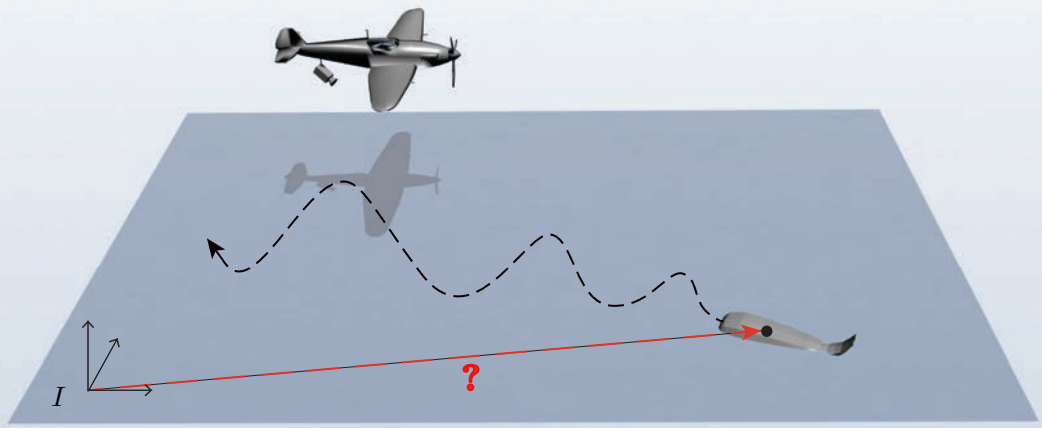
How to deal with outdoor applications, in which the target is usually further away from the camera and it is difficult to put a (e.g. GPS) receiver on it?

GPS/AHRS Aided Application

Problem Statement

Goal: estimate the state (position and velocity) of a marine mammal over time

- UAV instrumented with
  - GPS: UAV position
  - AHRS (Attitude and Heading Reference System): UAV orientation
  - PTZ camera: center of the image of the target



The distance from the target to the camera is not measured, but we know that the target moves on a plane (ocean surface)

GPS/AHRS Aided Application

Filtering Approach

Models:

- Target: *2D Horizontal Constant-Turn Model with Known Turn Rate*
- UAV: 3D kinematic model (suits different types of aircrafts)
- Camera: pinhole model

Nonlinear problem

(nonlinear observation equation and unknown angular speed)

Two different filters

Isolated KF

1) Transform measurements into linear functions of the target state
2) Design a **time-invariant KF** for the state of the target
3) MMAE (unknown angular speed)

Joint KF

1) **Augment the state** of the system with the state of the UAV
2) Transform measurements into linear functions of the system state
3) Design a **time-varying KF** for the augmented state
4) MMAE (unknown angular speed)

GPS/AHRS Aided Application

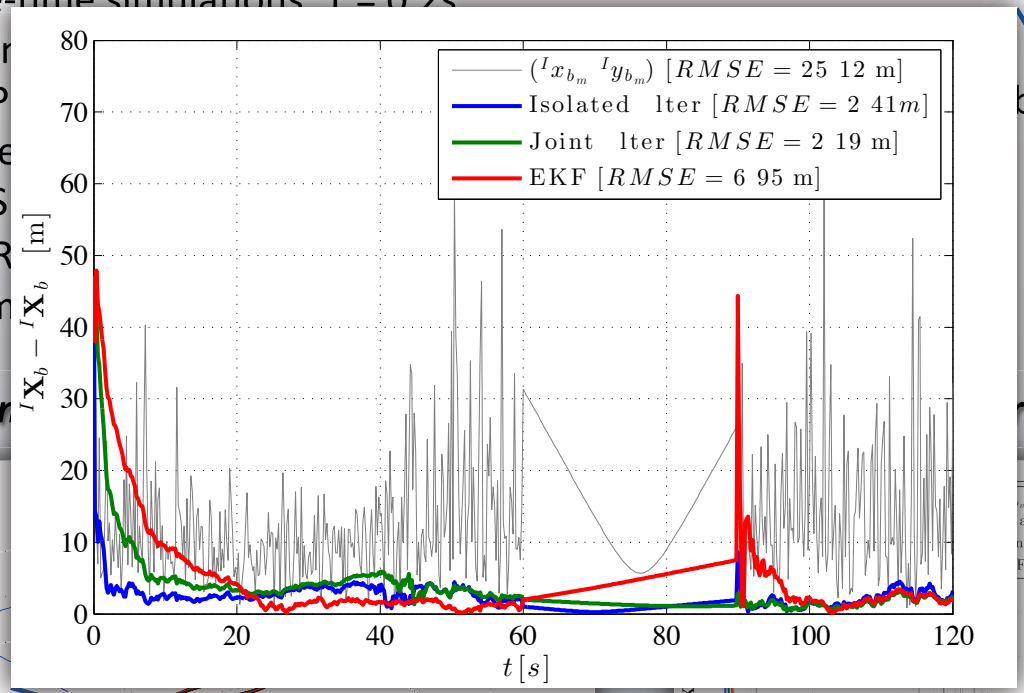
Simulation Results

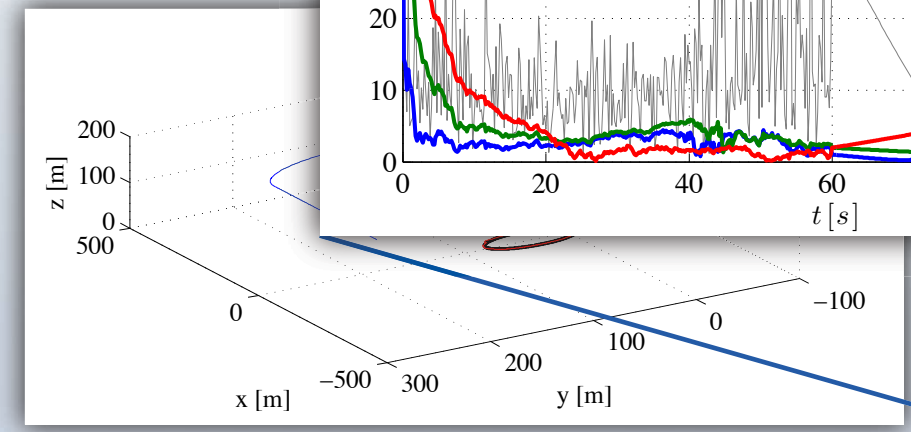
Setup description

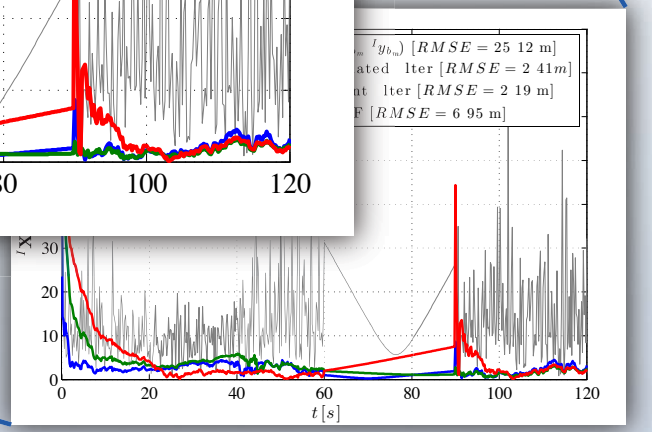
- discrete-time simulations: T = 0.2s
- target ar
- Baram P
- Measure
- GPS
- AHR
- Cam

EKF
augmented
with the
target
angular

Performance comparison (occlusions)







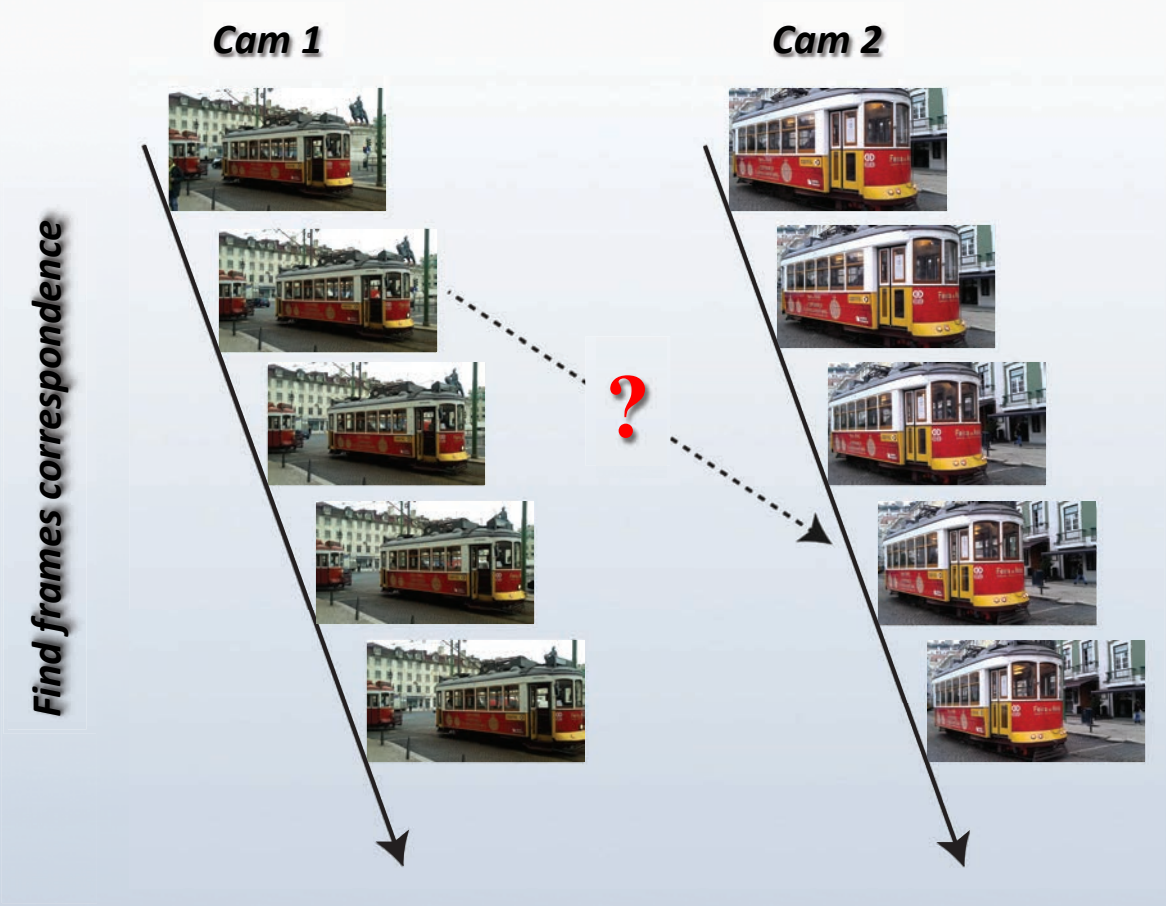
The presented algorithms are mostly appropriate for monocular systems.
A natural next step is their extension to multi-camera configurations.

This requires that all the cameras are synchronized (the target is dynamic).

How to synchronize a set of cameras that track a moving object without using additional hardware?

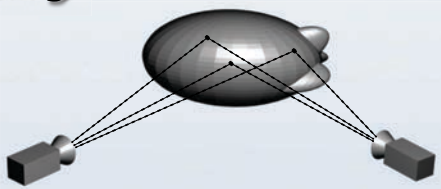
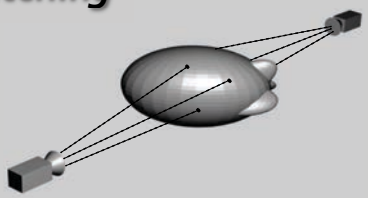
Video Synchronization

Problem Description



Video Synchronization

Previous Work

Features \ Cameras	Static	Independently Moving
Matching 	• Caspi and Irani, PAMI'02 • Tresadern and Reid, BMVC'03 SOLVED	• Tuytelaars and Van Gool, CVPR'04 • Meyer et al., BMVC'08 SOLVED
Non-Matching 	• Wolf and Zomet, VMDS'02 • Caspi and Irani, PAMI'02 SOLVED	***Most general and complex problem*** NOT SOLVED

Video Synchronization

Algorithm Outline

Main idea: use the relative motion between two objects as clue for the synchronization

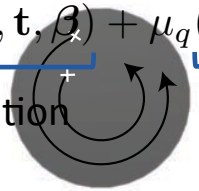
1. **Track features** in two rigid objects with independent motions
2. **Get their 3D motion** (apart from a scaling factor) w.r.t. each camera
3. **For each time delay $\delta \in [-\Delta, \Delta]$, find the constant rigid body transformation $(\mathbf{R}, \mathbf{t})$** between the cameras in the initial time instant

Nonlinear problem

The features cannot be used directly

Cam 1 ***Cam 2***

$$E_m(\delta) = \min_{(\mathbf{q},\mathbf{t},\boldsymbol{\beta})} \mu_R \underbrace{E_R(\delta, \mathbf{q})}_{\text{Rotation}} + \mu_T \underbrace{E_T(\delta, \mathbf{q}, \mathbf{t}, \boldsymbol{\beta})}_{\text{Translation}} + \mu_q \underbrace{(\mathbf{q}^T\mathbf{q} - 1)^2}_{\text{Unit quaternion}}$$



- $\mathbf{q}$: unit quaternion associated with $\mathbf{R}$
- $\boldsymbol{\beta}$: vector with non-negative scaling factors

4. **Choose the delay** that verifies $\hat{\delta} = \arg\min_{\delta} E_m(\delta)$

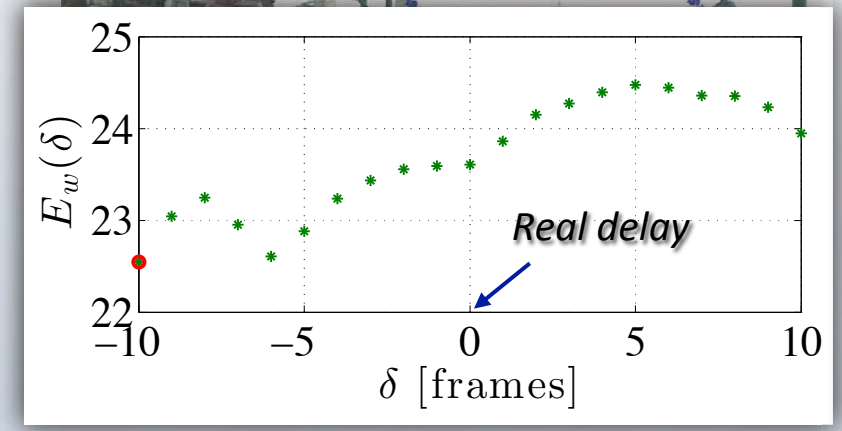
Video Synchronization

Experimental Results

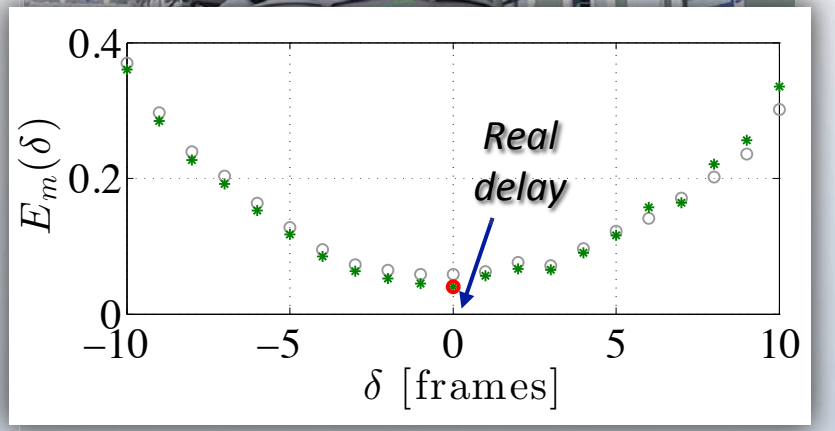
System description

- Regular cameras (29 fps)
- Small video sequences: 96 frames
- Delay in the interval [-10, 10] frames
- Ground truth: mark some frames with a photo-flash

Wolf and Zomet, in VMDS'02



Our method



Summary

Architecture for monocular 3D positioning and tracking systems based on a suboptimal nonlinear multiple-model adaptive estimator

Depth estimation methods for **monocular target tracking and positioning**

- no information about the geometry of the target is required

Model-based **H_2 adaptive filter** for target tracking and positioning

- **convergence guarantees** (hold even when the EKF diverges)

Positioning and tracking system for **outdoor** applications

- particularly useful when placing a (e.g. GPS) receiver on the target is difficult or not recommended

Synchronization algorithm for **independently moving cameras**

- first step towards the extension of the presented methods to multi-camera configurations

Summary

What can be done in terms of 3D positioning when a single static camera is used and there is no information about (either the dimensions or the shape of) the target?

It is possible to obtain full 3D position estimates, with convergence guarantees.

Directions for Future Research

Limitations:

- **Planar target model:** extension to nonplanar models, but does the gain in authenticity compensate for the additional complexity?
- **Slowly rotating target:** infer the shape of the target (e.g. structure from motion) and estimate its attitude
- **Continuous-time H_2 adaptive filter:** search for a discrete-time version that retains the convergence guarantees

Possible directions of research:

- Use 3D pose to help finding the direction of travel, [Fossati and Fua, ECCV'08]
- Multiple targets
- Extension to multi-camera configurations
- Video synchronization without retrieving the 3D motion of the objects

Publications

Journal

T. Gaspar, P. Oliveira, and C. Silvestre. Model-based filters for 3D positioning of marine mammals using UAVs. **TAES** (in press).

T. Gaspar and P. Oliveira. New depth from focus filters in active monocular vision systems for indoor 3D tracking. **TCST** (in press).

T. Gaspar and P. Oliveira. Model-based H_2 adaptive filter for 3D positioning and tracking systems. **Automatica**, 50(1):225–232, 2014.

T. Gaspar and P. Oliveira. Single pan and tilt camera indoor positioning and tracking system. **EJC**, 17(4): 414 – 428, 2011.

Under Review

T. Gaspar, P. Oliveira, and P. Favaro. Synchronization of independently moving cameras via motion recovery. **PAMI**.

Conference

T. Gaspar, P. Oliveira, and P. Favaro. Synchronization of two independently moving cameras without feature correspondences. In **ECCV**, 189–204, 2014.

T. Gaspar and P. Oliveira. LMI-based H_2 adaptive filtering for 3D positioning and tracking systems. In **ACC**, 4183–4188, 2013.

T. Gaspar, P. Oliveira, and C. Silvestre. UAV-based marine mammals positioning and tracking system. In **ICMA**, 1050–1055, 2011.

T. Gaspar and P. Oliveira. Depth estimation in active monocular vision systems for indoor 3D tracking. In **IFAC-WC**, 9470–9475, 2011.

T. Gaspar and P. Oliveira. Monocular depth from focus estimation with complementary filters. In **ICRA**, 4986–4991, 2011.

T. Gaspar and P. Oliveira. New dynamic estimation of depth from focus in active vision systems. In **VISAPP**, 484–491, 2011.

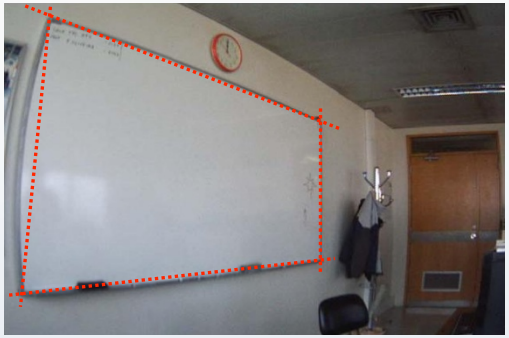
T. Gaspar and P. Oliveira. Single pan and tilt camera indoor positioning and tracking system. In **ECC**, 2792–2797, 2009.

T. Gaspar and P. Oliveira. A single pan and tilt camera architecture for indoor positioning and tracking. In **VISAPP**, vol. 1, 523–530, 2009.

Thank you.

System Architecture

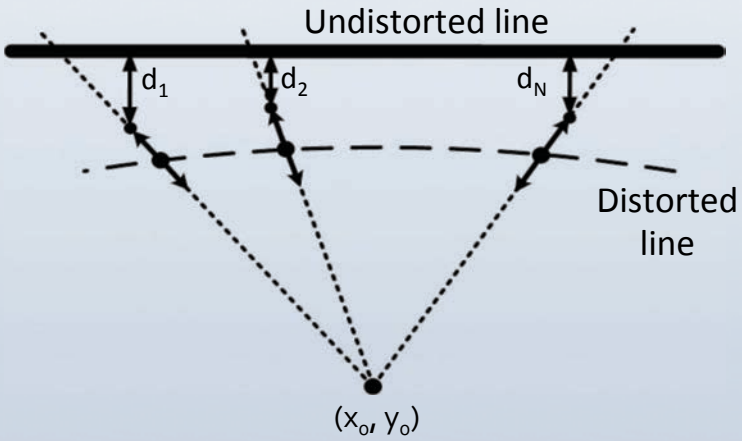
Lens Distortion

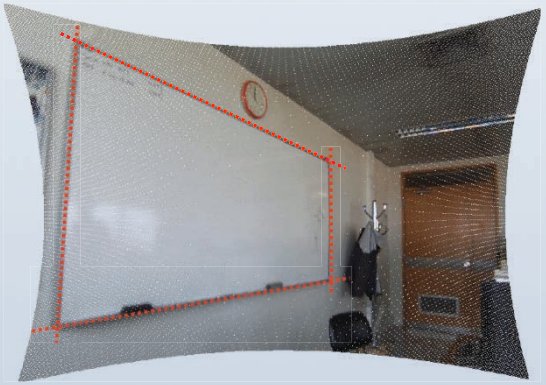


Low cost PTZ camera

- Radial distortion (*barrel*)

Distortion compensation: straight lines must be preserved





System Architecture

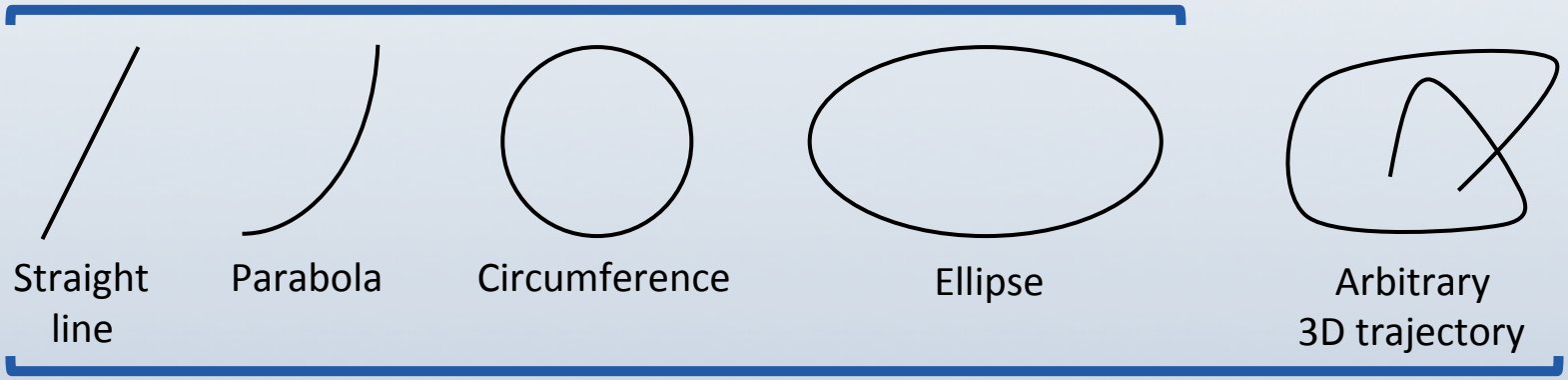
Target Model

Why this model?

- most targets move on a plane (e.g. people, animals)
- also adequate to track targets that span the 3D space
- based on basic kinematic relations (not particular for a given type of target)
- more complex models are usually highly nonlinear (poor performance with agile targets at low data rates)

3D trajectories considered by the model

in the absence of noise: $\mathbf{d}(t) = 0$



in the presence of noise: $\mathbf{d}(t) \neq 0$

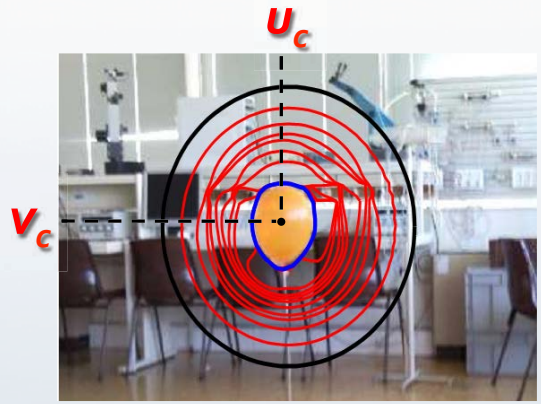
System Architecture

Measurements

The target is segmented using active contours

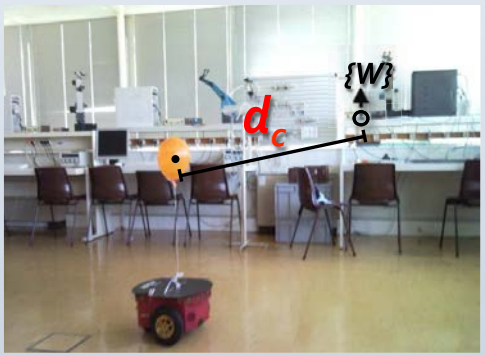
Mean of the contour

Target center coordinates



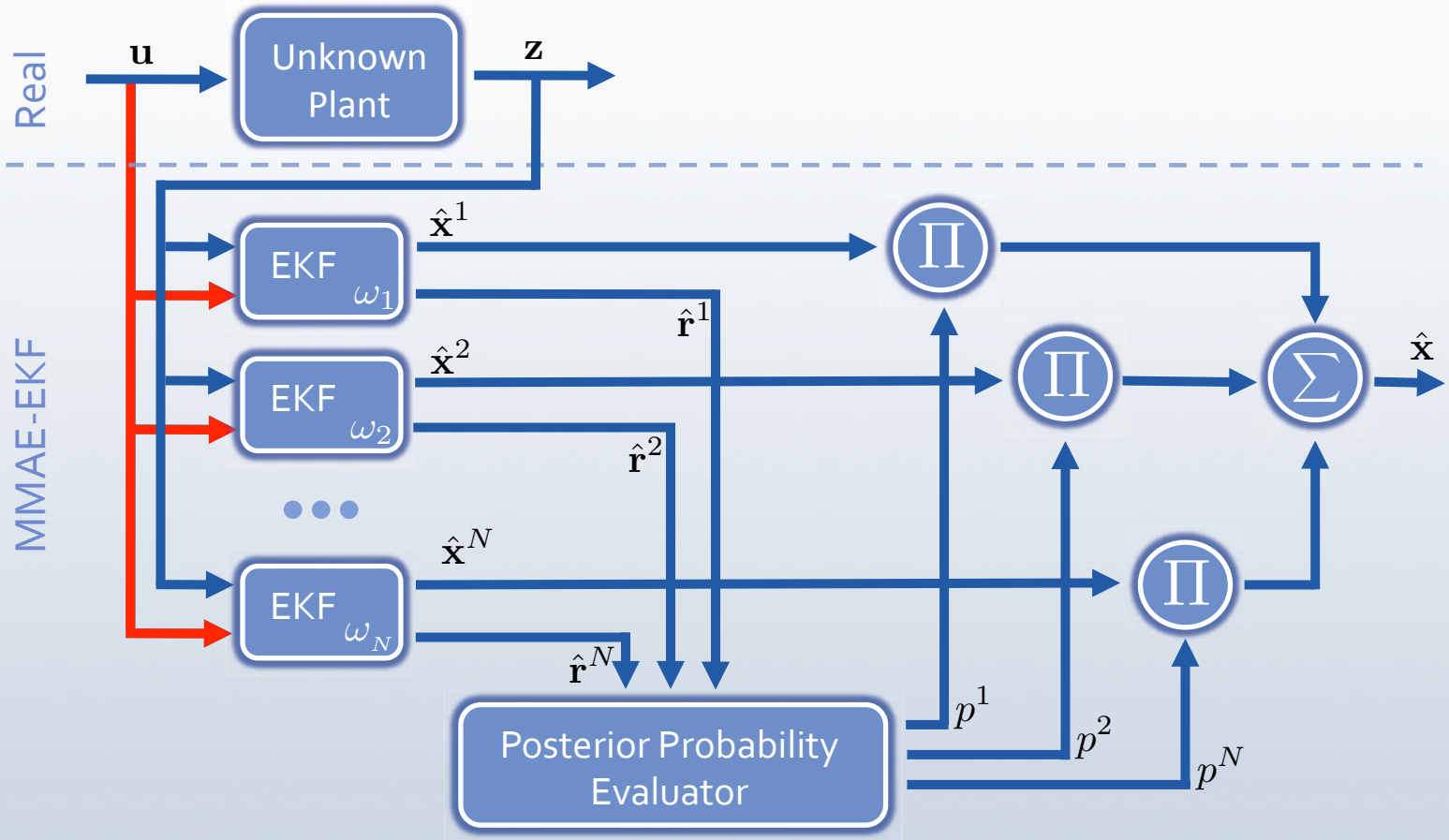
Target dimensions
Visual looming strategy

Distance of the target to the origin of the world reference frame



System Architecture

Multiple-Model Adaptive Estimator (MMAE)



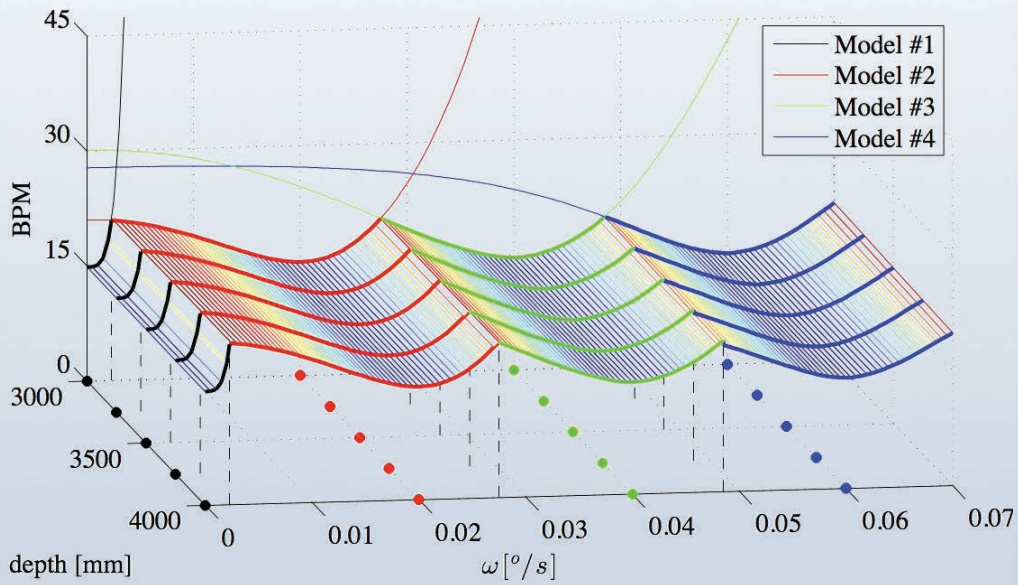
System Architecture

Multiple-Model Adaptive Estimator (MMAE)

Design of the bank of EKFs

Division of the parameter set: use Baram Proximity Measure (BPM) for insight (nonlinear system with unstable target model)

- Linearized versions of the filter for several depths
- No convergence guarantees



System Architecture

Experimental Results

System description

- *Creative WebCam Live! Motion*
- Easily identifiable target
- Sampling interval: 0.5 sec

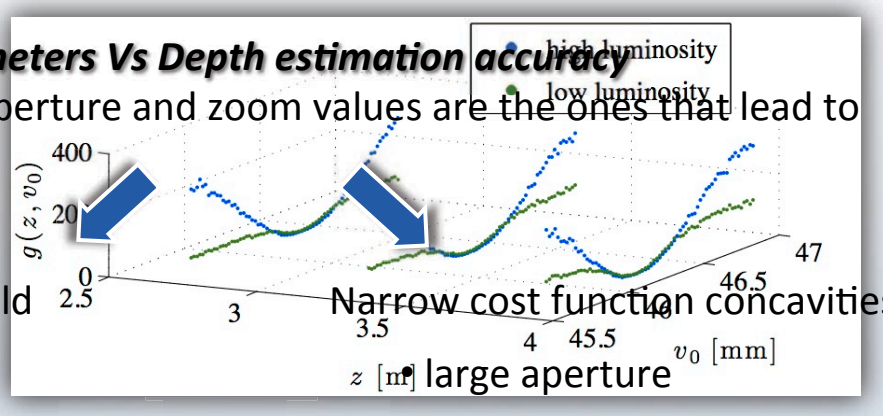
Monocular Depth Estimation

Cost function shape

- If the PSF is a constant intensity disk, then a parabola is the exact model for g(v0)
- The illumination of the scene does not change the position of the cost function minimum

Camera parameters Vs Depth estimation accuracy

- the best aperture and zoom values are the ones that lead to

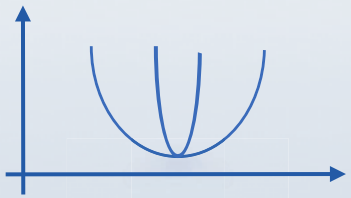


Small depths of field

- large aperture
- large focal length ⇨ ***tradeoff*** ⇦ • small focal length

Narrow cost function concavities

- large aperture



Monocular Depth Estimation

Depth Filtering Strategies

Second-order complementary filter

Depth from focus measurements at low frequencies

&

Depth derivative (visual looming) at high frequencies

- Mimics the natural frequency decomposition of the measurements
- Structure of a stationary KF (under some noise conditions)

LPV observer

Target depth dynamics written as a function of the measurements of the dimensions of the image of the target

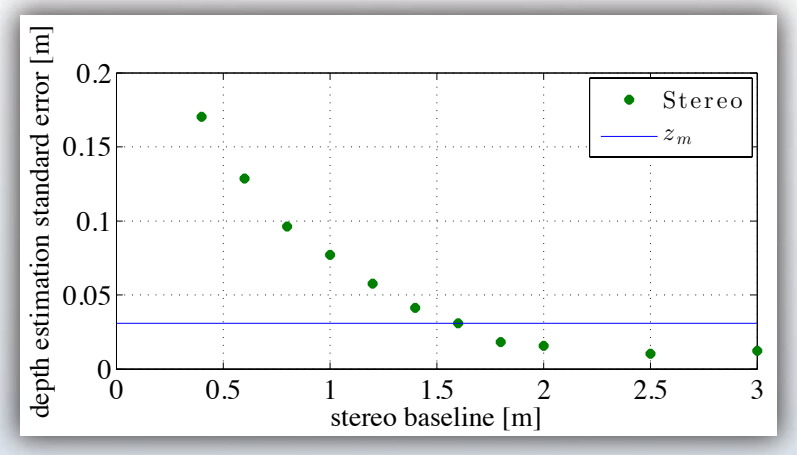
Parameter dependent system

UES (inequality on the gain of the observer)

Monocular Depth Estimation

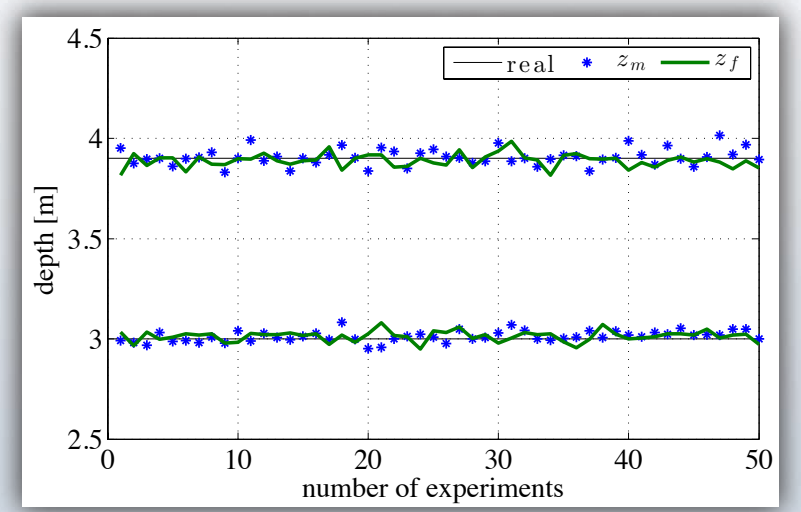
Comparisons

Classical stereo



Monocular (state-of-the-art)

[Favaro and Soatto in PAMI 2005]



Relative RMSE

- *Our method: 1.03%*
- *Favaro: 0.94%, but slower (3x)*

Monocular Depth Estimation

Comparisons

Monocular depth estimation methods (static scenes and real data)

Method	*RMSE (%)*	*Depth range [m]*	*Real-time positioning*
Pentland	2.5	up to ~1	yes
Ens	1.3	0.8 to 0.95	no
Subbarao	2.3 to 20	0.6 to 5	yes
Favaro	~1	3 to 4	no
Krotkov	~1	1 to 3	no
Our method	~1	3 to 4	yes

H_2 Adaptive Filter

Problem Statement

Target dynamics

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\overset{?}{\omega})\overset{?}{\mathbf{x}}(t) + \mathbf{B}\mathbf{d}(t)$$

Consider:

- a ***target maneuvering in the 3D space*** according to the previous model
- ***measurements of the target position***

Goal:

- design a state (position, velocity, and acceleration) estimator such that its ***error***
  1. ***converges exponentially fast to zero*** in the absence of noise
  2. ***is bounded*** in the presence of noise

Assumptions:

- constant, unknown, and bounded angular speed
- bounded noise
- bounded linear velocity

H_2 Adaptive Filter

Other Approaches

*Target dynamics (**unstable** and **unknown**)*

$$\dot{\mathbf{x}}(t) = \underbrace{\mathbf{F}(\omega)}\mathbf{x}(t) + \mathbf{B}\mathbf{d}(t)$$

eigenvalues in the imaginary axis

MMAE approach

Performance degradation when the true plant is not within the set of plants

No convergence guarantees since the dynamics is unstable

Linear filter

Biased unless the true angular speed is used

No convergence guarantees since the true angular speed is unknown

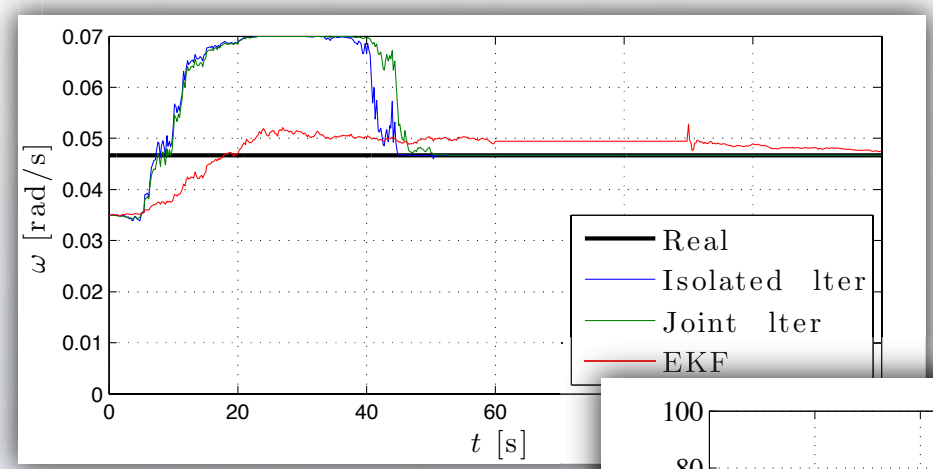
State augmentation

Augment the state with the angular speed

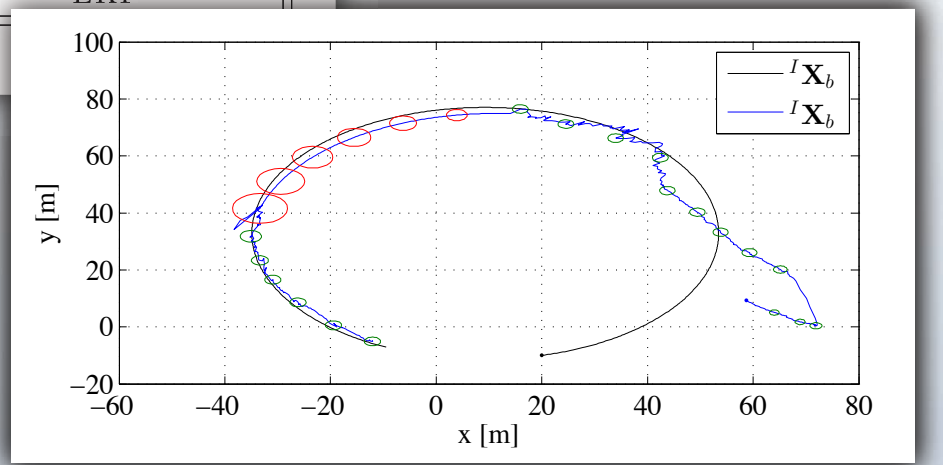
No convergence guarantees since the system that results is highly nonlinear

GPS/AHRS Aided Application

Simulation Results



Angular speed estimates (occlusions)



Target position covariance (occlusions)

Video Synchronization

Independently Moving Cameras