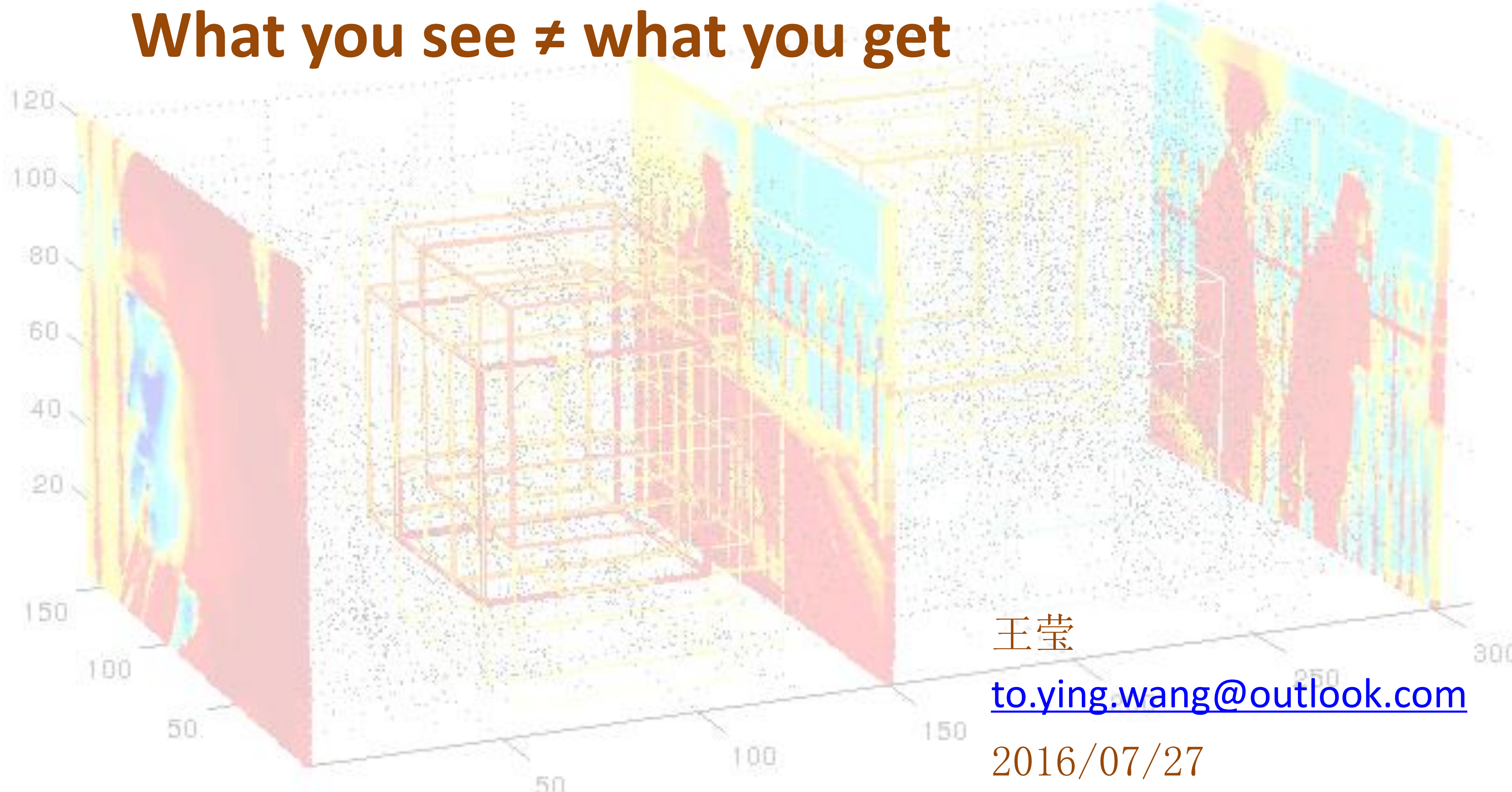
Robotic Vision

What you see ≠ what you get

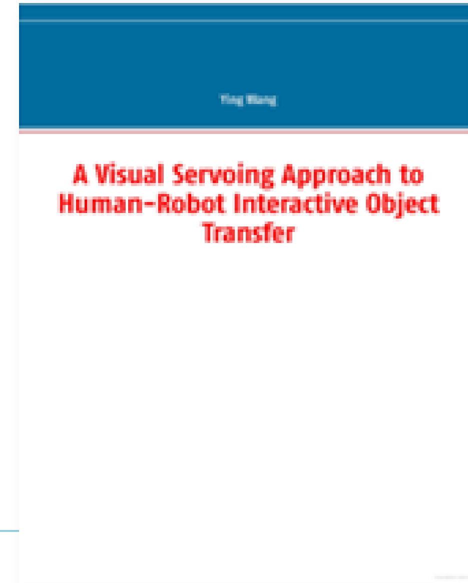


王莹
to.ying.wang@outlook.com
2016/07/27

0. About the Presenter

Dr. Ying Wang
Computer Vision Algorithm Engineer

Research Interests:
Computer Vision, HRI, Robotics, Automation control


A Visual Servoing Approach to Human-Robot Interactive Object Transfer

RWTH Aachen University, Germany

PhD Candidate in Mechanical Engineering, focusing on Robot Vision

Degree completed in 2016.01 with a mark of "sehr gut"

Topic: A Visual Servoing Approach to Human-robot Interactive Object Transfer

Harbin Institute of Technology, China

Master of Science in Mechatronics Engineering (top 5% of class)

Topic: Study of the DSP based Servo Controller of Die Bonding Machine

Wuhan University of Technology, China

Bachelor of Science in Mechanical Engineering and Automation (top 5%)

Topic: Design of Mechanical Structure of the Oil Tube marking Machine

Contents

Demands for Robotic Vision

Dr. Ying Wang
Jul. 27th 2016

Robots are designed and built to complement human abilities.

Efficiency & Automation

Stacking boxes for shipping

Mobility

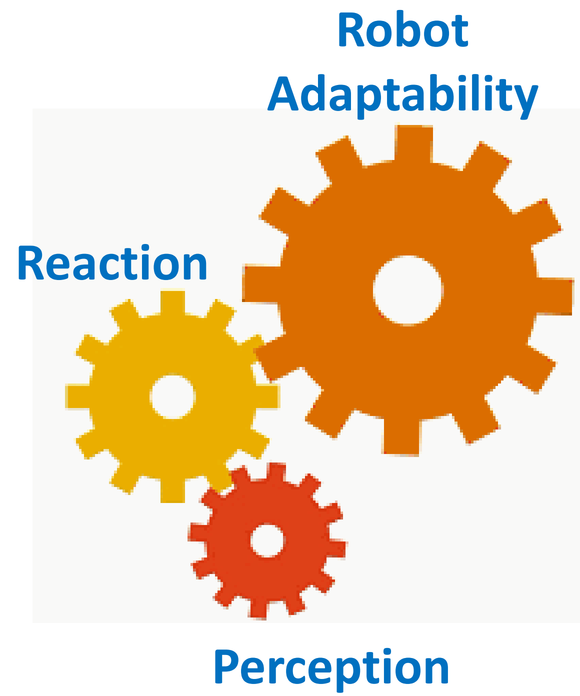
Surveillance

Skills & Accuracy

Medical operation

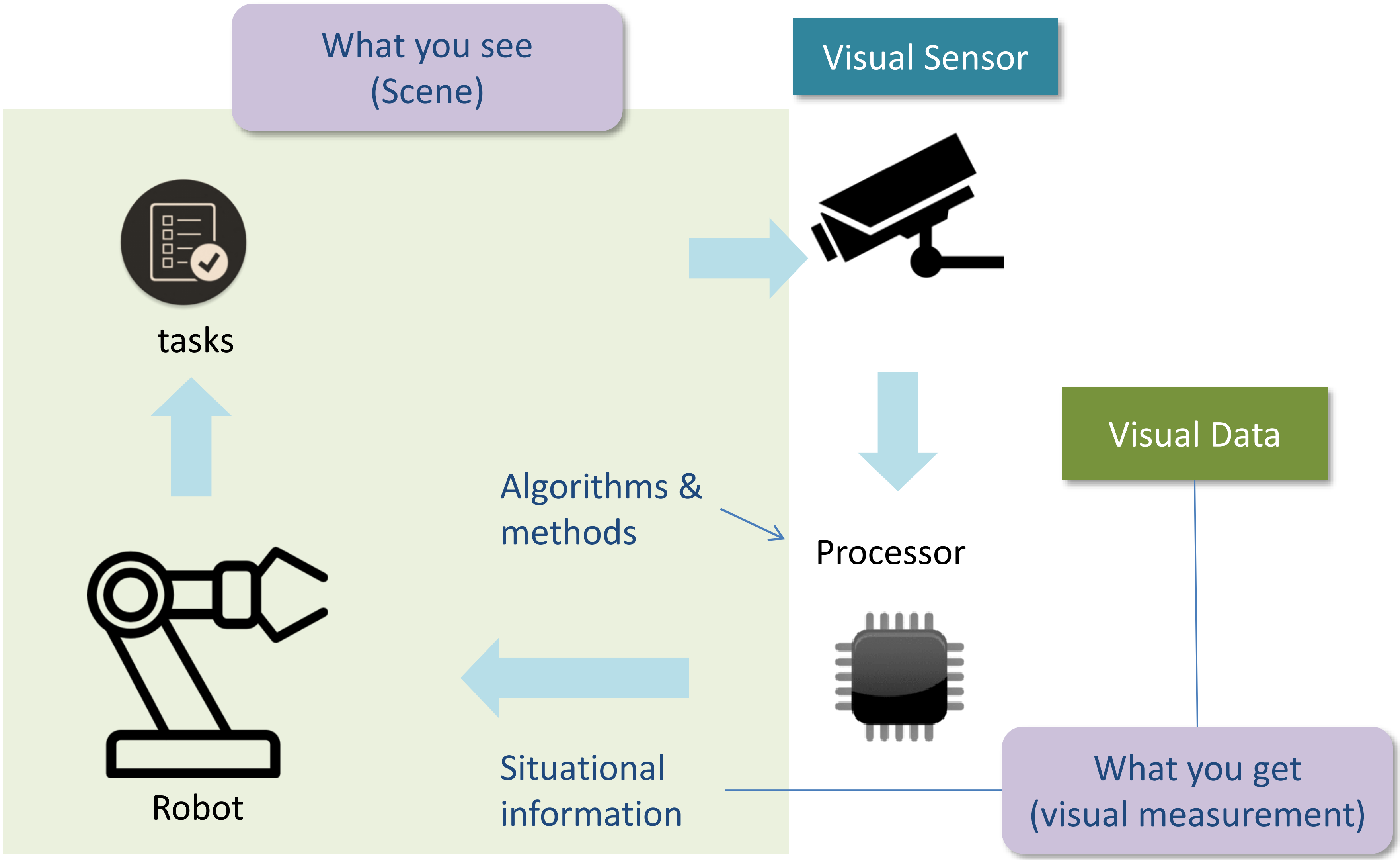
Hazards Tolerance

Spray painting



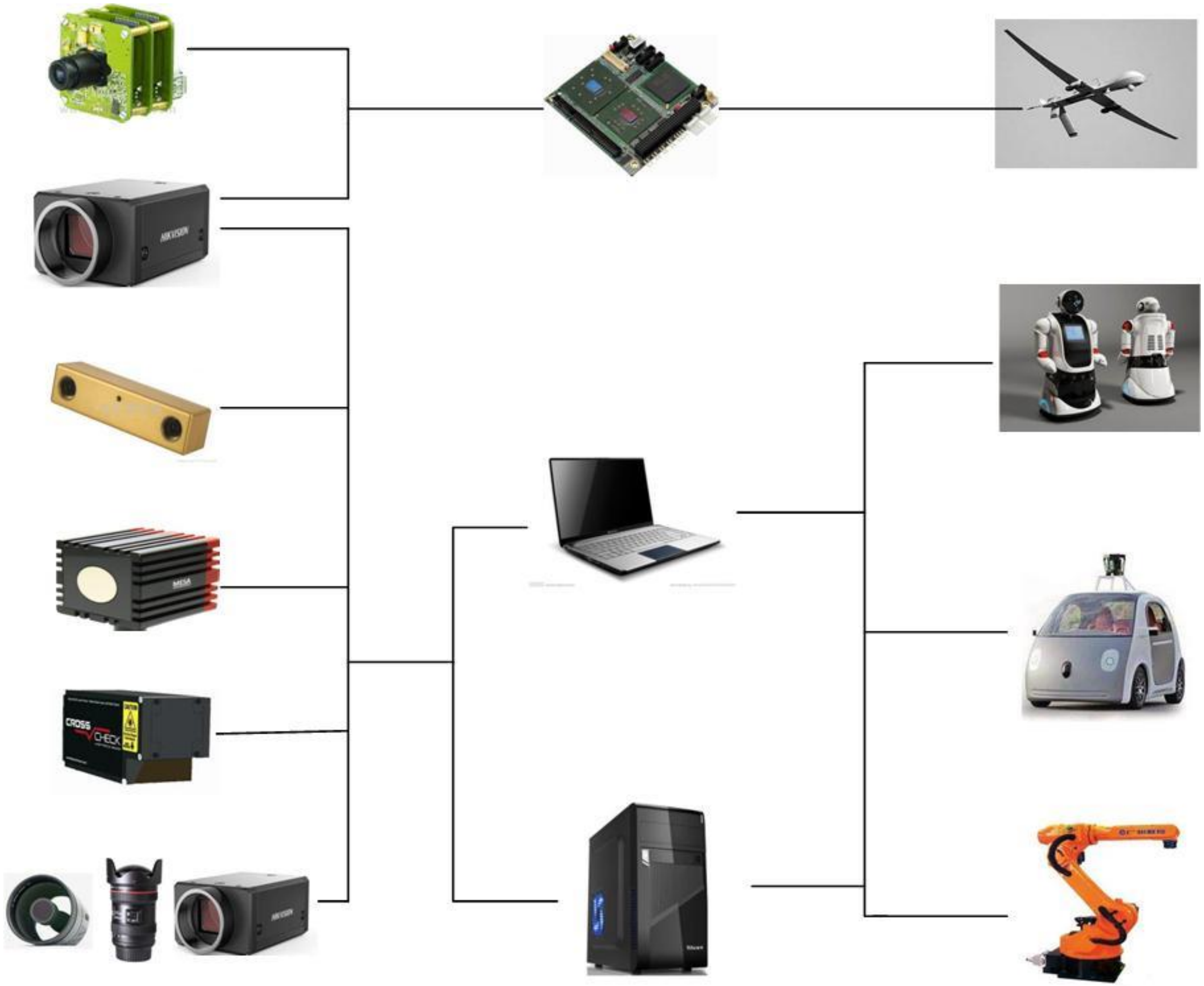
Robotic Vision System

Dr. Ying Wang
Jul. 27th 2016



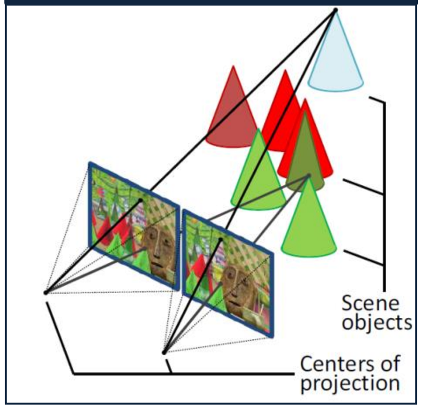
Robotic Vision System

Dr. Ying Wang
Jul. 27^{th} 2016



Dr. Ying Wang
Jul. 27th 2016

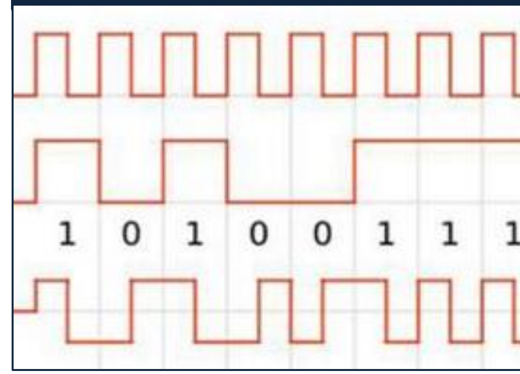
Imaging



Robotic Vision

© Sensopart

Signal Processing



AI

Computer Vision

Machine Vision

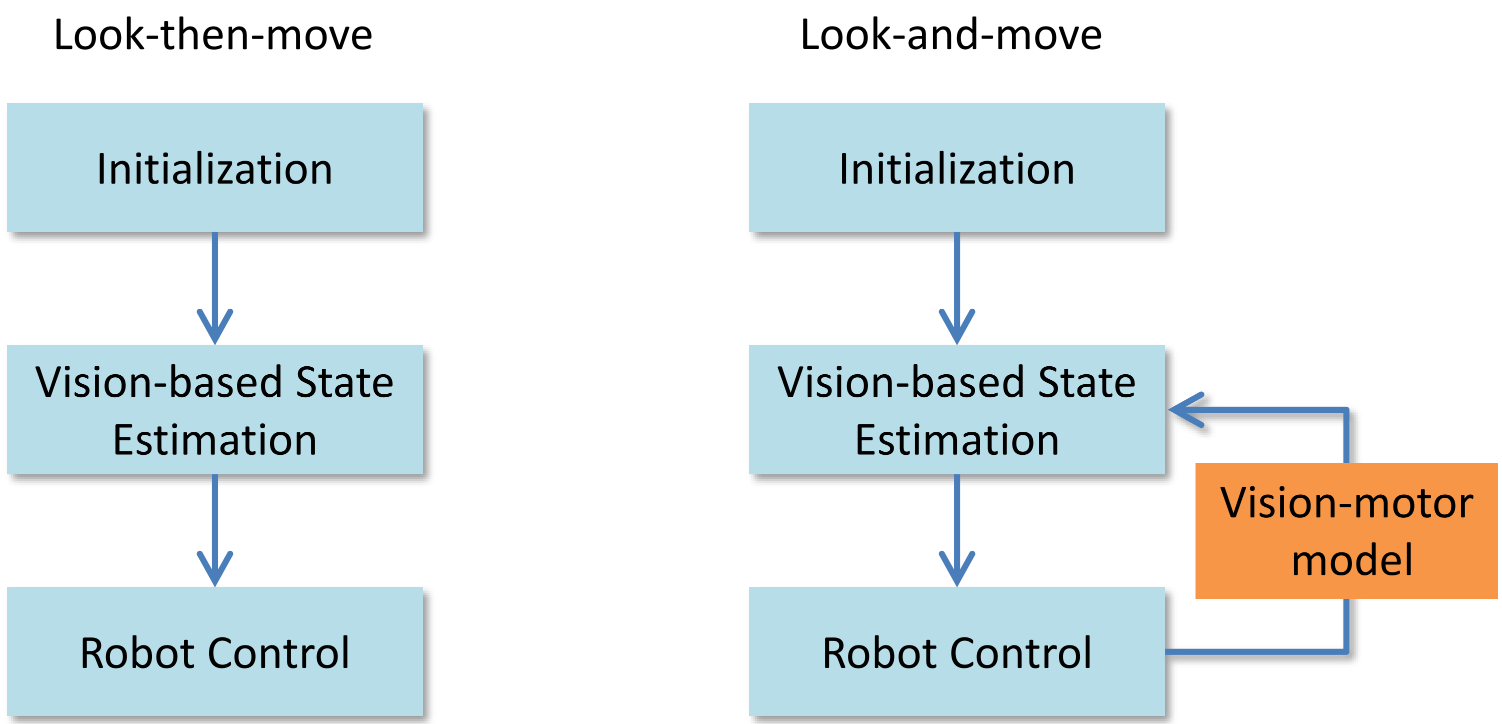
Image Processing

Photogrammetry

© Saab AB

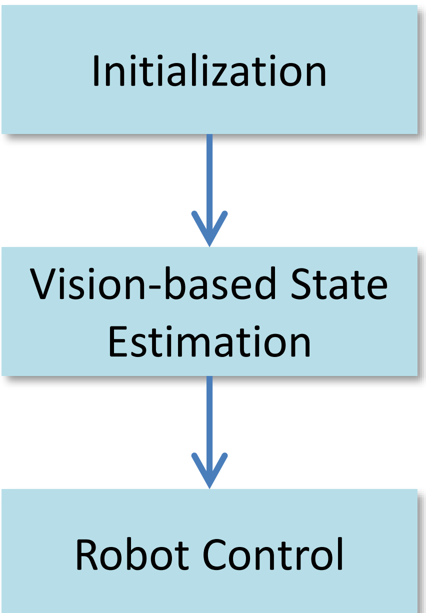
Contents

General Approach to Robotic Vision



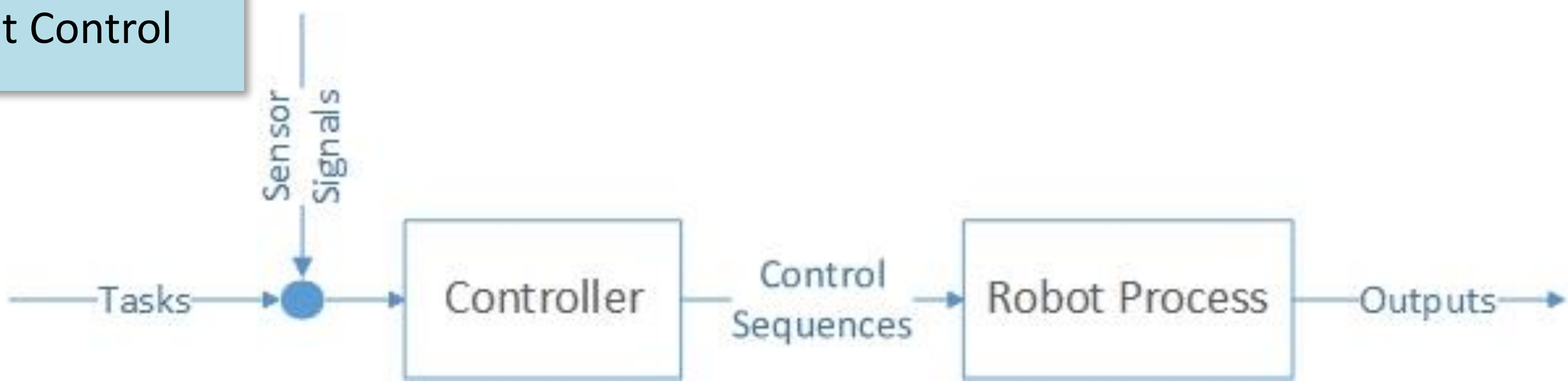
Look-then-move

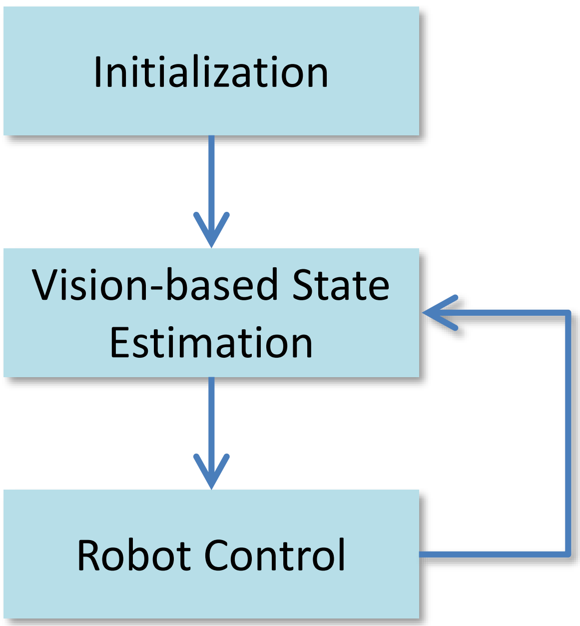
Dr. Ying Wang
Jul. 27th 2016



Look-then-move
Visual sensing and manipulation are combined directly in an open-loop fashion.

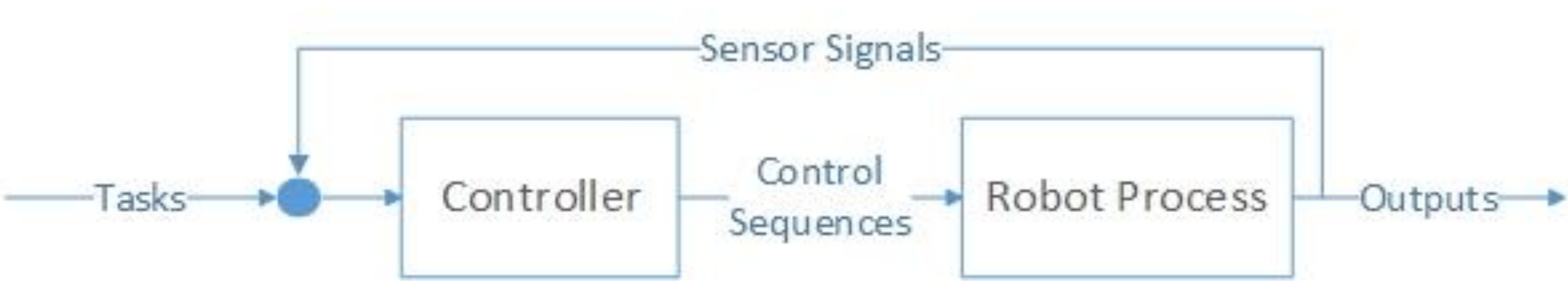
The accuracy of the operation, in such a configuration, depends directly on the accuracy of the hardware, such as the visual sensors, the manipulator and the controller.





Visual Servoing
uses a visual-feedback control loop to increase the overall accuracy of the system - a principal concern in any application.

Visual servoing approaches broaden the application domain of robotic manipulation, as they do not need *a priori* knowledge of the workspace, that is, they are competent of visual control in an unmodeled environment.



Vision-based State Estimation

Dr. Ying Wang

State estimation, in the sensor space, mainly concerns the camera configuration and the image processing algorithms. Various visual information is extracted, such as color, pose, or features of objects, to describe the state of the target object.

Visual Sensors

Processor

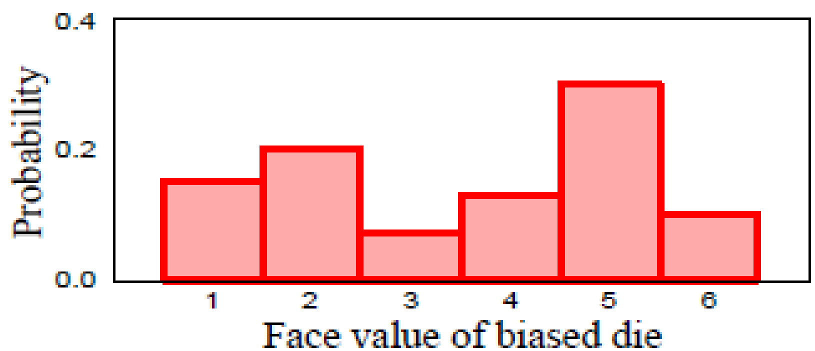
Visual data	2D Image	Depth Image	3D point cloud	video
World state	presence	status	Pose	velocity

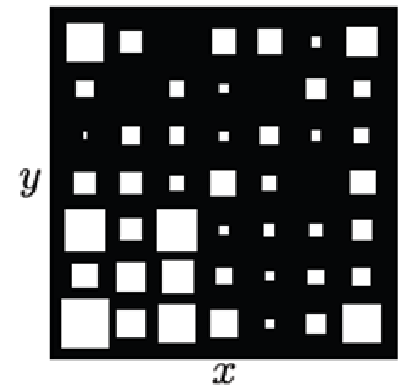
Random variable *x* denotes a quantity that is uncertain. This information is captured by the probability distribution $P_r(x)$ of the random variable. A random variable may be discrete or continuous.

1 D

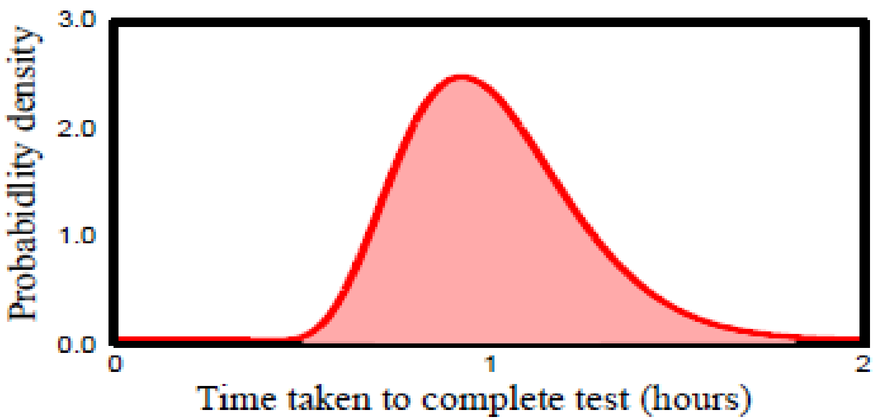
2 D

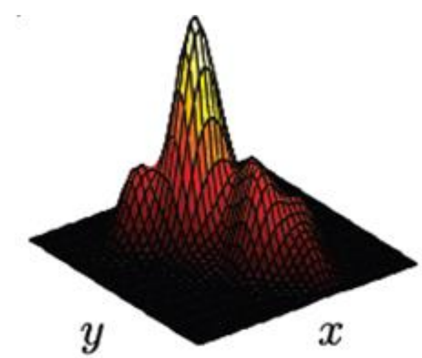
discrete





continuous



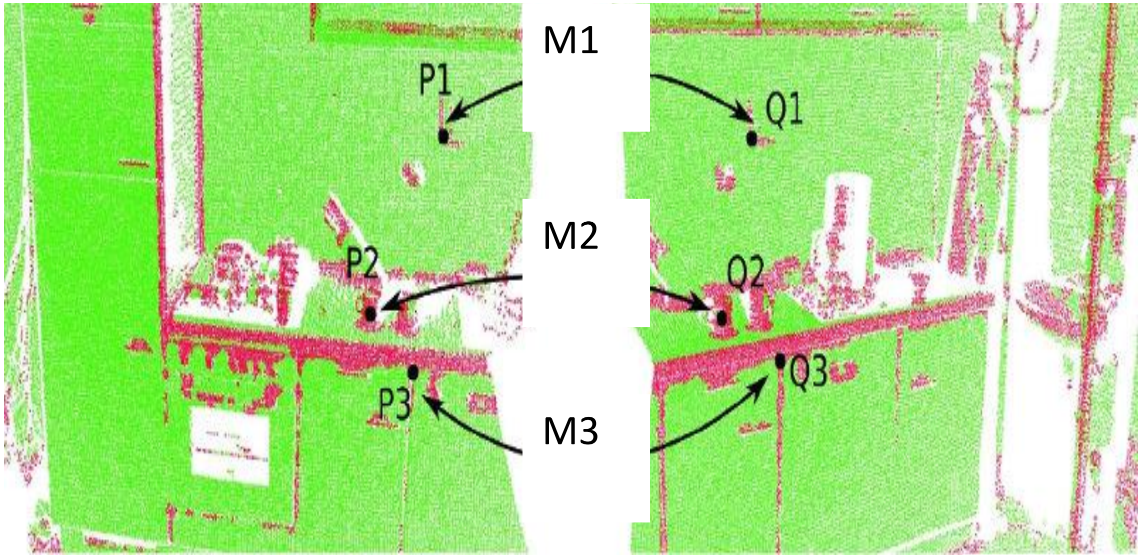


Dr. Ying Wang
Jul. 27th 2016

Noise

©pudn

Many-to-one mapping



©PCL

Dr. Ying Wang
Jul. 27th 2016

taking visual data x and use them to infer the state of the world θ

fitting probability models to data - learning

Maximum Likelihood

the maximum likelihood (ML) method finds the set of parameters $\hat{\theta}$ under which the data $\{x_i\}_{i=1}^{I}$ are most likely.

$$\hat{\theta} = \max_{\theta}[P_r(x_1 \dots x_I|\theta)] = \max_{\theta}\left[\prod_{i=1}^{I} P_r(x_i|\theta)\right]$$

Maximum a posteriori

maximum a posteriori estimation maximizes the posterior probability $[P_r(x_1 \dots x_I|\theta)]$of the parameters

$$\hat{\theta} = \max_{\theta}[P_r(\theta \mid x_1 \dots x_I)] = \max_{\theta}\left[\frac{P_r(x_1 \dots x_I|\theta)P_r(\theta)}{P_r(x_1 \dots x_I)}\right] = = \max_{\theta}\left[\frac{\prod_{i=1}^{I} P_r(x_i|\theta)P_r(\theta)}{P_r(x_1 \dots x_I)}\right]$$

$$\hat{\theta} = \max_{\theta}\left[\prod_{i=1}^{I} P_r(x_i|\theta)P_r(\theta)\right]$$

Dr. Ying Wang

Jul. 27th 2016

Bayesian approach

$$P_r(\theta \mid x_1 \dots x_I) = \frac{\prod_{i=1}^{I} P_r(x_i|\theta) P_r(\theta)}{P_r(x_1 \dots x_I)}$$

Evaluating the predictive distribution is more difficult for the Bayesian case since we have not estimated a single model but have instead found a probability distribution over possible models. Hence, we calculate

$$P_r(x^*|\, x_1 \dots x_I) = \int P_r(x^*|\, \theta) \mathrm{P}_r(\theta|\, x_1 \dots x_I) d\theta$$

General Form

The predictive density calculations for the Bayesian, MAP and ML cases can be unified as

$$P_r(x^*|\, x_1 \dots x_I) = \int P_r(x^*|\, \theta) \delta\left[\theta - \hat{\theta}\right] d\theta = P_r(\mathrm{x}^*|\hat{\theta})$$

State Estimation

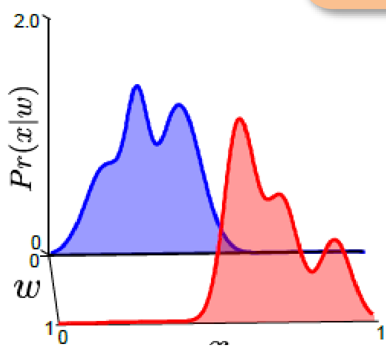
Machine Learning Solution

Dr. Ying Wang
Jul. 27th 2016

model to mathematically relate the visual data x and the world state w. The model specifies a family of possible relationships between x and w and the particular relationship is determined by the model parameters θ.

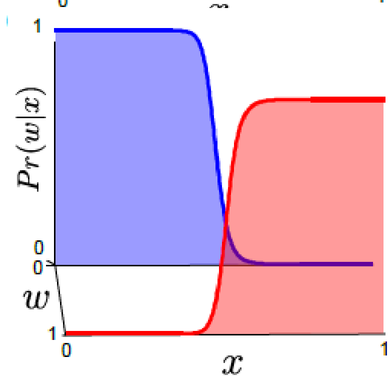
Visual data ***x*** — Model (***θ***) — World state ***w***

Discriminative model
$P_r(w|x)$
– model the contingency of the world state on the data



Regression model
to estimate a continuous quantity from continuous data. e.g. predicting the joint angles from an image of the human body.

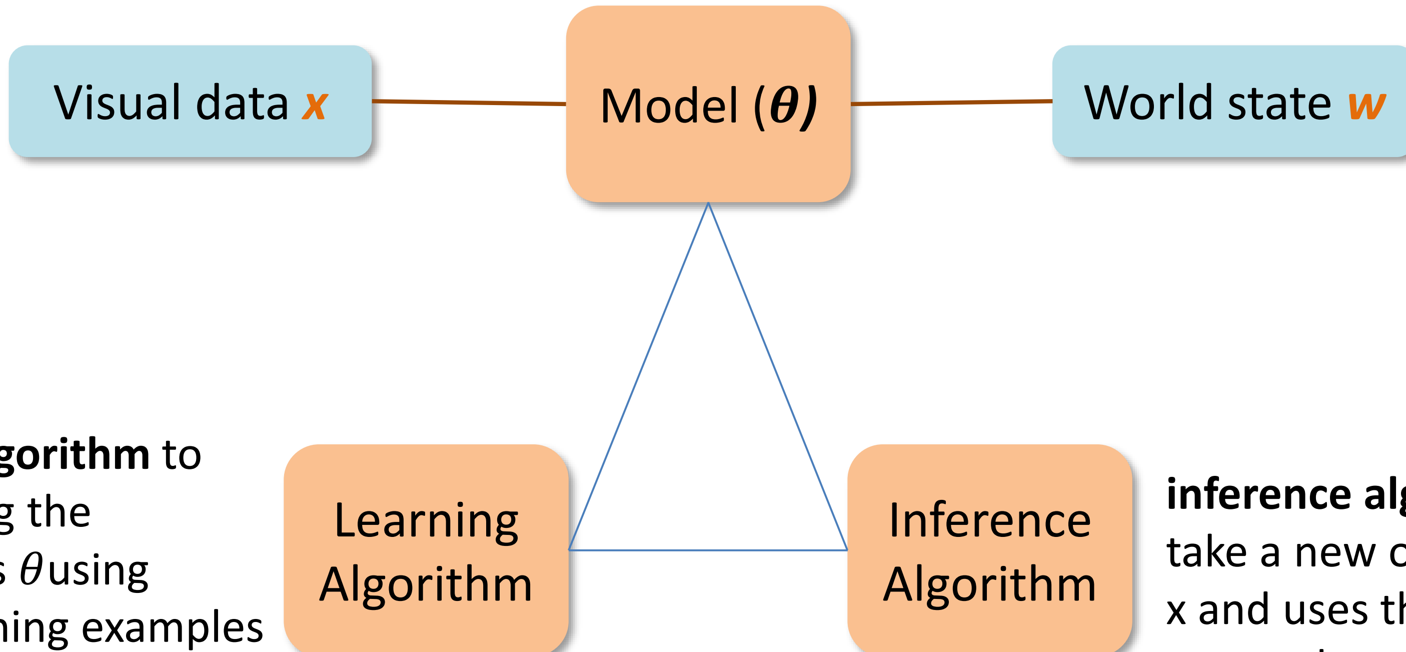
Generative model
$P_r(x|w)$
– model the contingency of the data on the world state



Classification model
to predict a discrete quantity from continuous data. e.g. assigning a label to a region of the image to indicate whether or not a face is present.

Dr. Ying Wang
Jul. 27th 2016

model to mathematically relate the visual data x and the world state w. The model specifies a family of possible relationships between x and w and the particular relationship is determined by the model parameters θ.

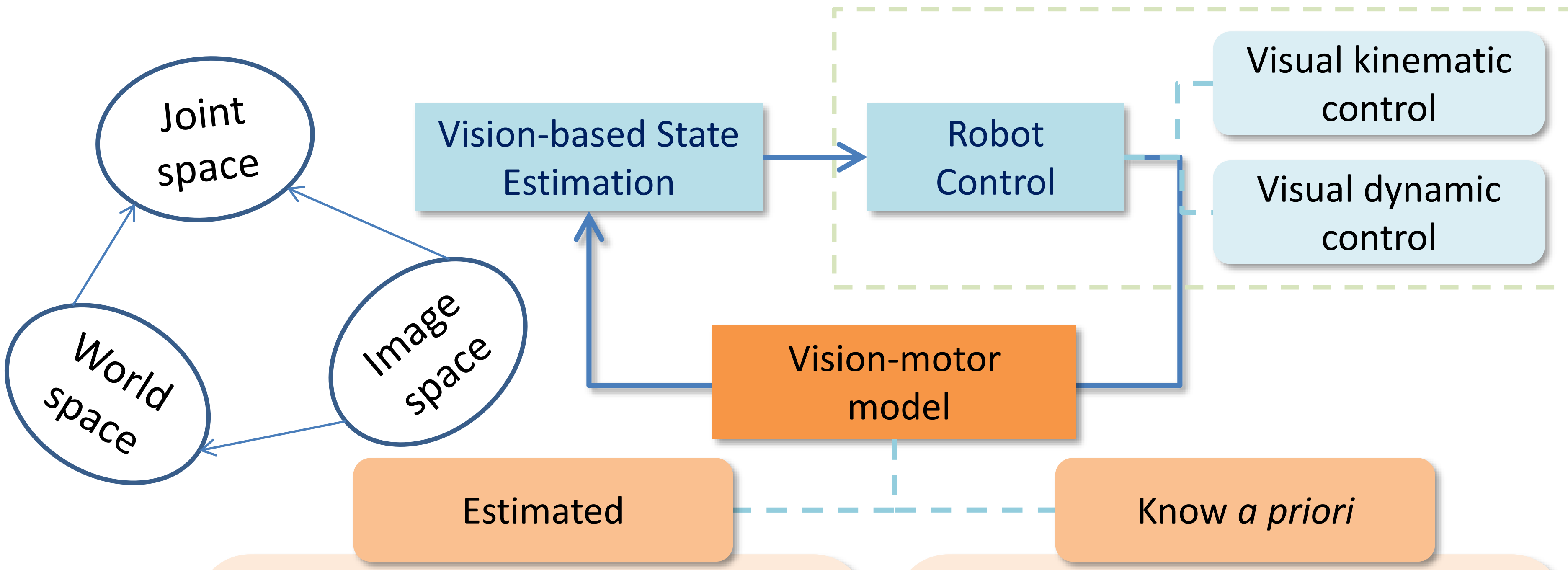


learning algorithm to allow fitting the parameters θ using paired training examples $\{x_i, w_i\}$ where we know both the measurements and the underlying state.

inference algorithm to take a new observation x and uses the model to return the posterior $P_r(w|x)$ over the world state w.

Vision-Motor Model

Dr. Ying Wang
Jul. 27th 2016



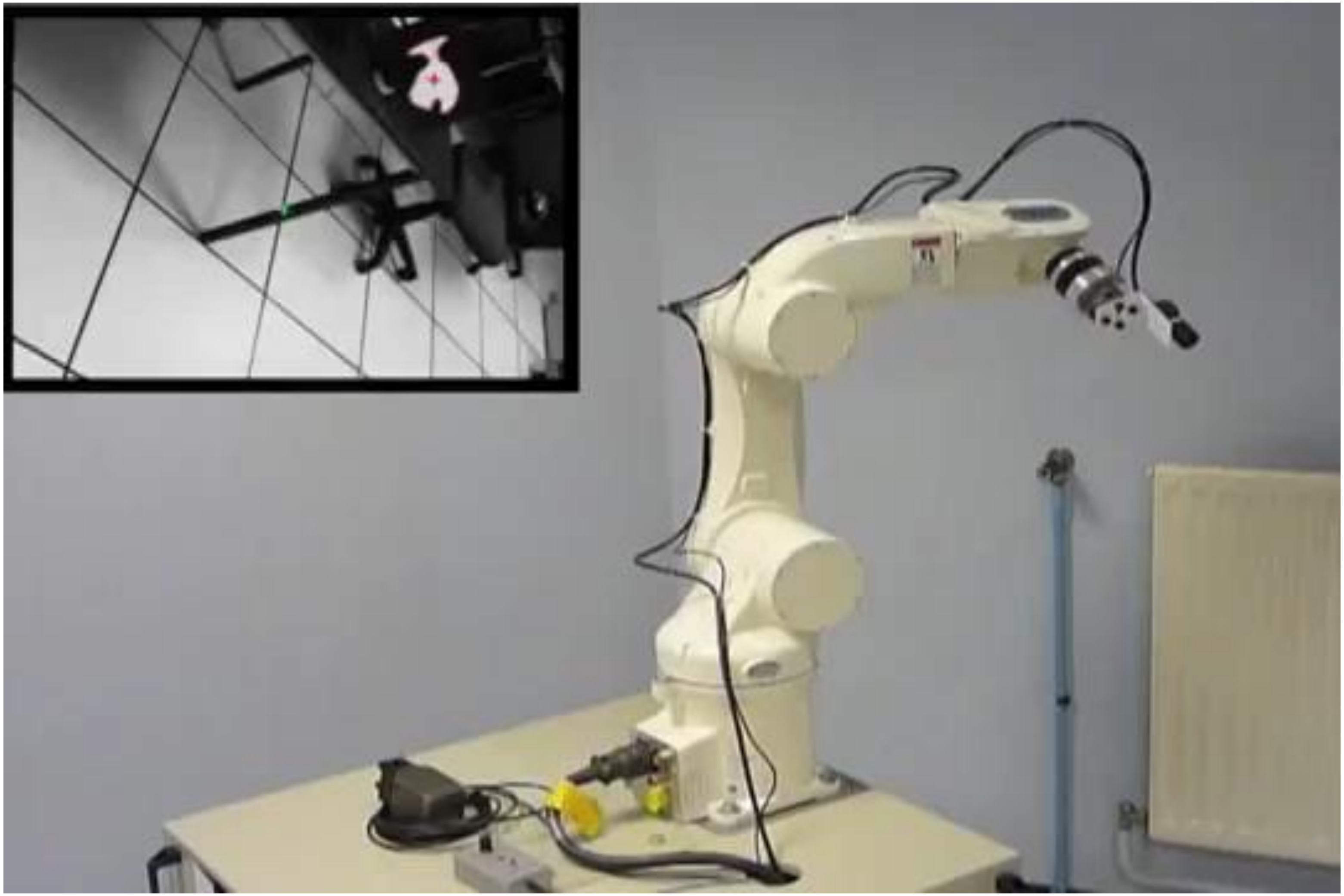
the model estimation can be conducted on- or off-line. The estimated image Jacobian relates the joint velocity directly to image space velocities, which can be estimated from previous measurements.

the forward or inverse kinematics of the robot (robot Jacobian) is available to deduce the differential changes between the joint and Cartesian space. e.g. PBVS, IBVS or 2 1/2D visual servoing

Contents

Solution Overview

Example: Object Following



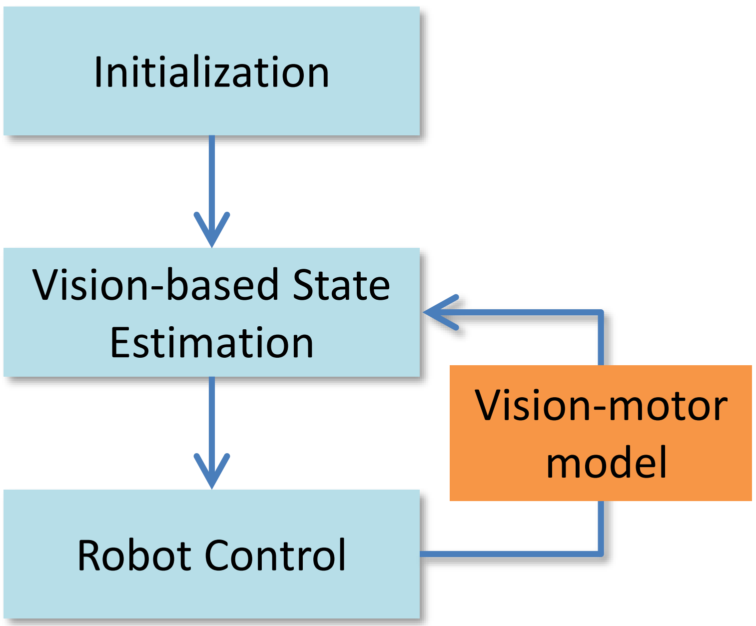
Solution Overview

Example: Object Following

Dr. Ying Wang
Jul. 27th 2016

Task : controlling the robot to follow the target

Analysis Model



Design Model

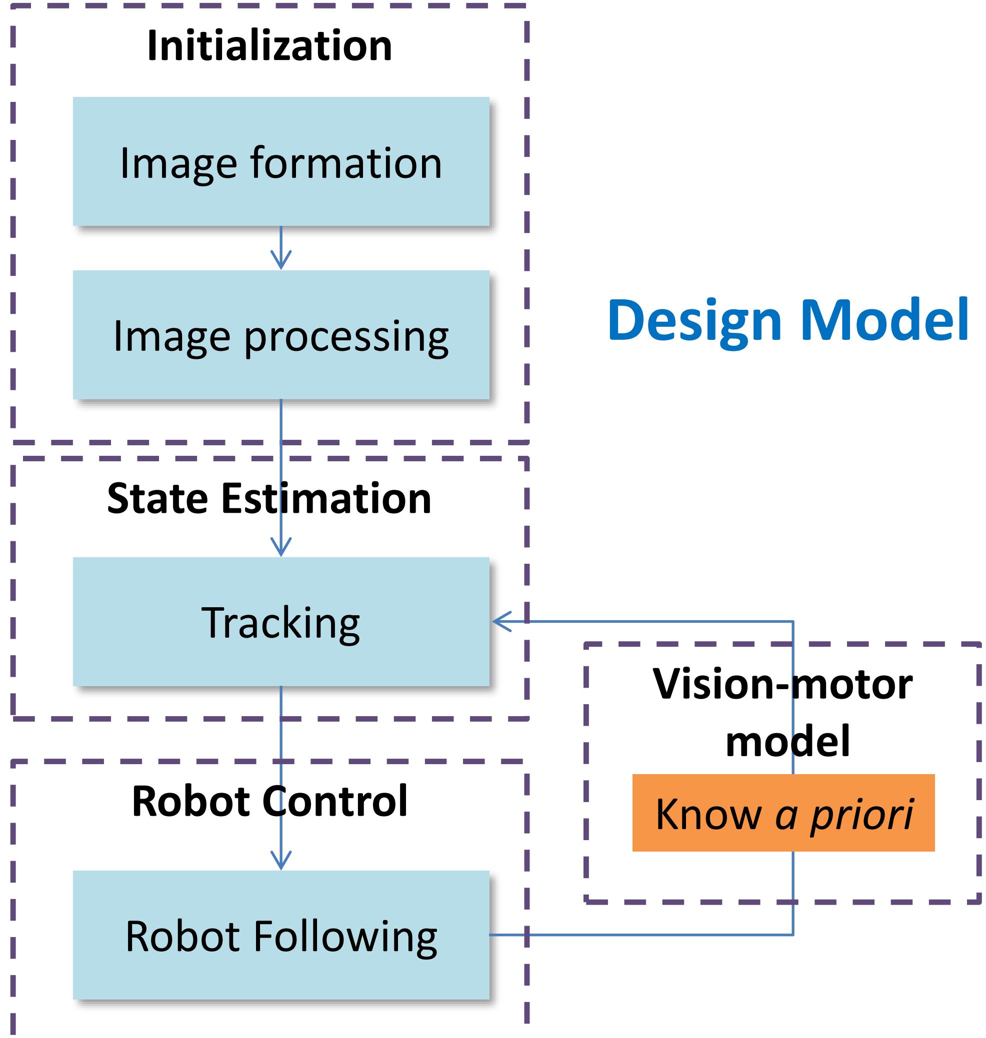


Image Formation

Dr. Ying Wang
Jul. 27th 2016

Perspective Projection

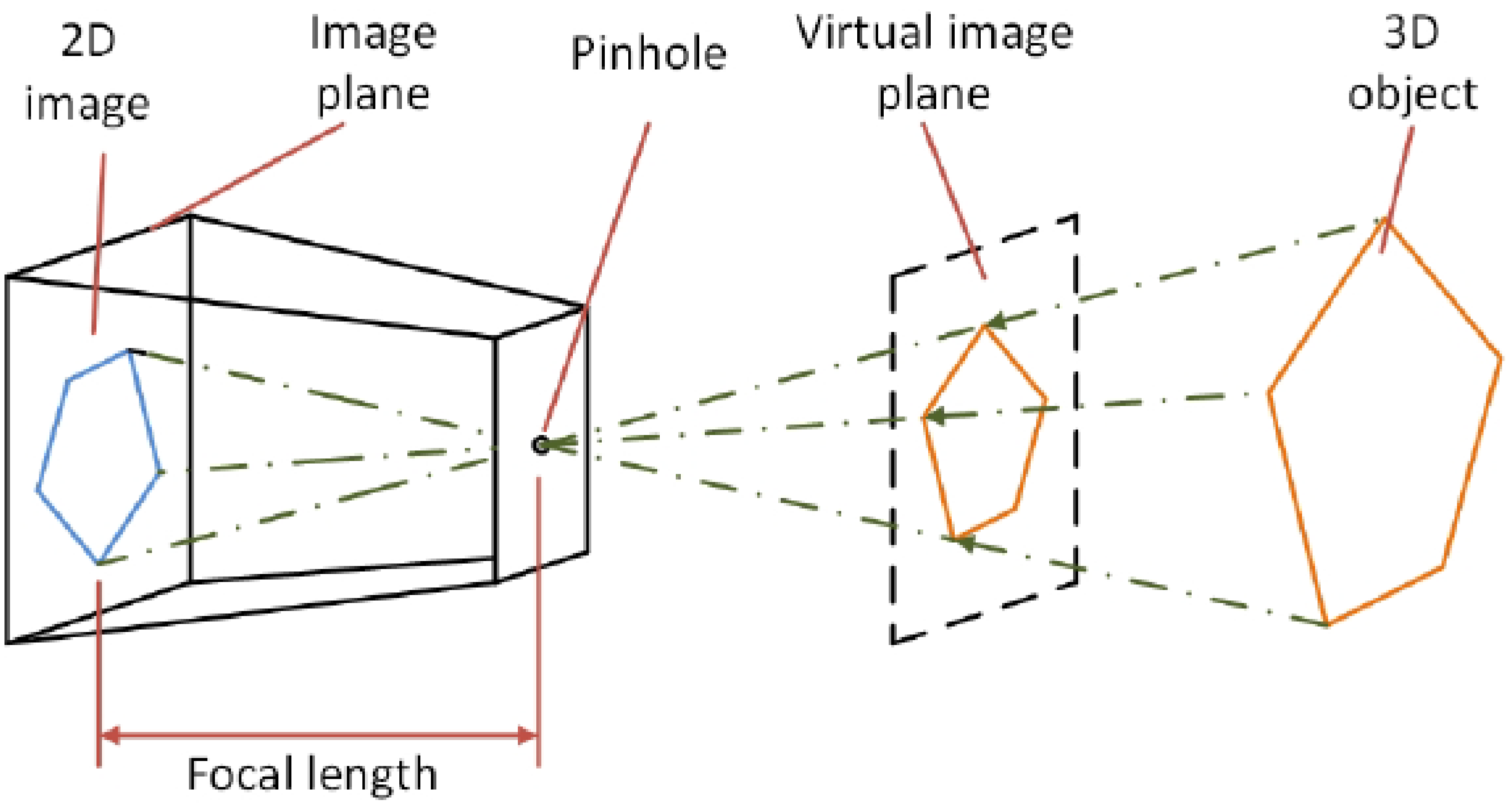
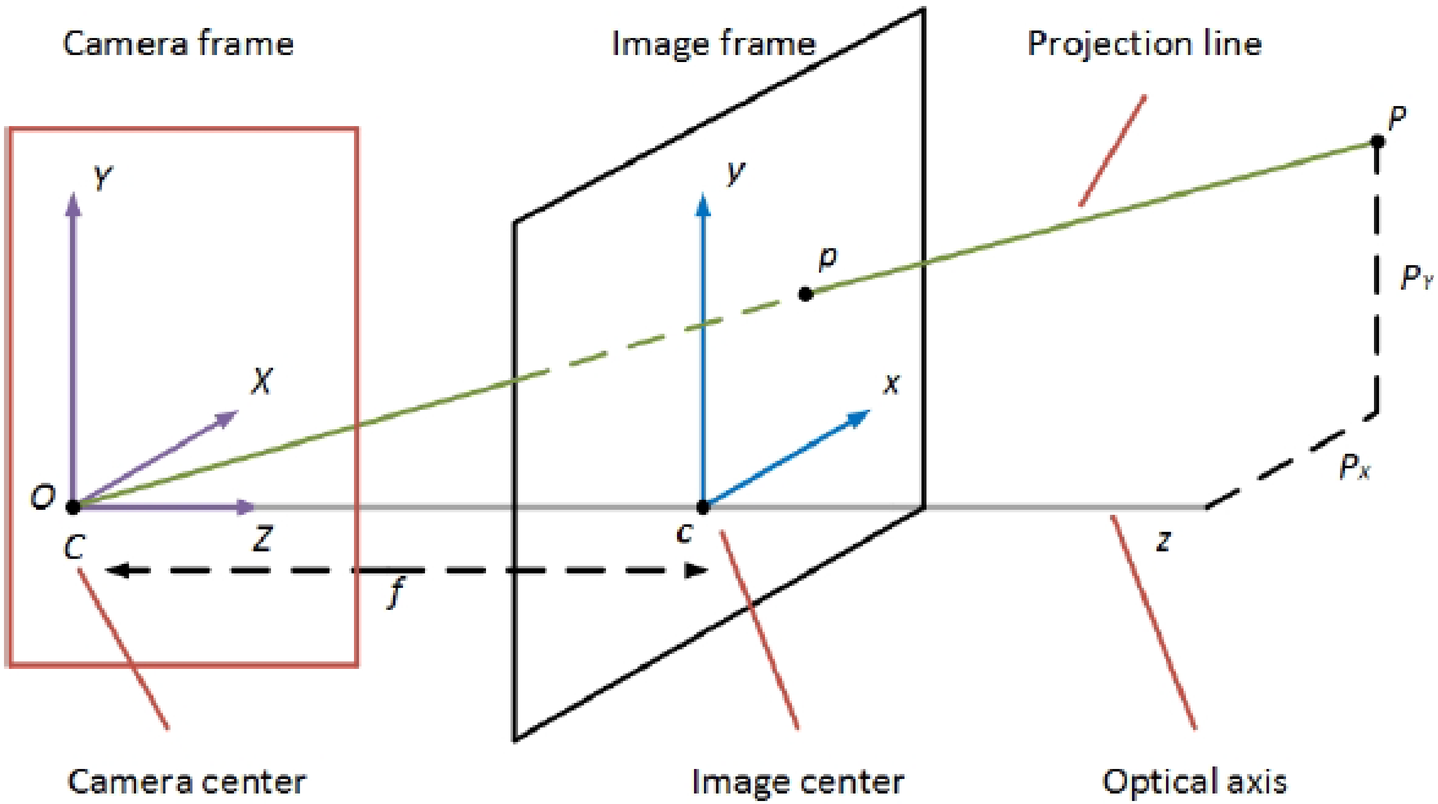


Image Formation

Dr. Ying Wang
Jul. 27th 2016

Perspective Projection



Projection

$$P\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow p\begin{pmatrix} f\frac{X}{Z} \\ f\frac{Y}{Z} \end{pmatrix}$$

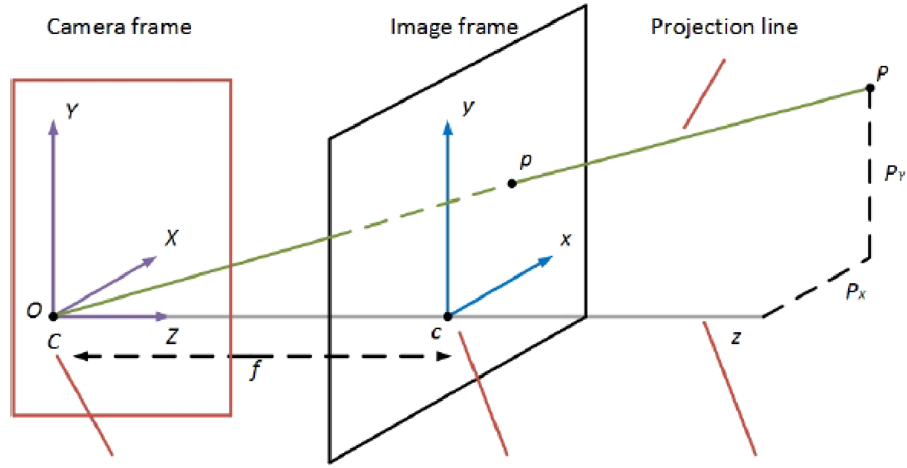
Dr. Ying Wang

Correspondence Points

$$P = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}^T, p = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}^T$$

Camera Intrinsics

$$M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$



Perspective Projection

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R_{CW} \quad T] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

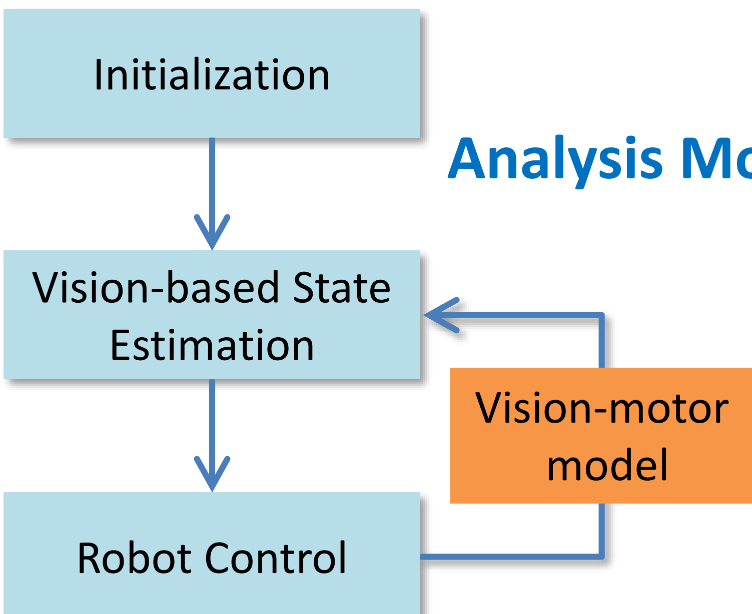
Solution Overview

Dr. Ying Wang
Jul. 27^{th} 2016

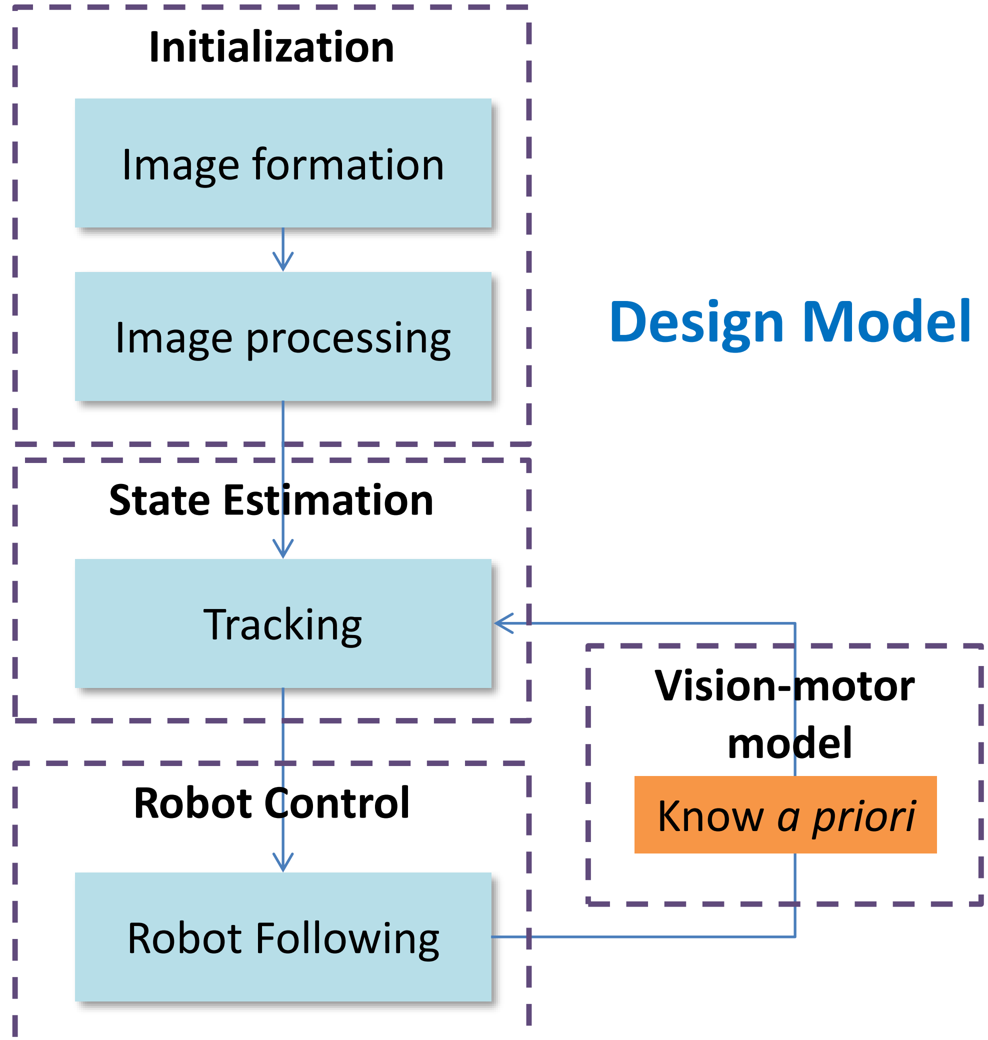
Example: Object Following

Task : controlling the robot to follow the target

What does the scene look like?
Which is the target object there?
Where is it located?
How should the robot be controlled?



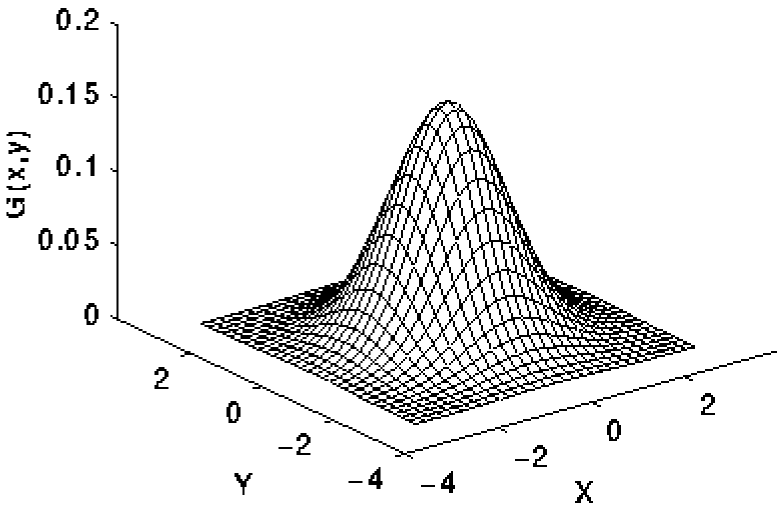
Analysis Model



Design Model

Dr. Ying Wang
Jul. 27th 2016

Gaussian Filter $G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$



$\frac{1}{273}$

1	4	7	4	1
4	16	26	16	4
7	26	41	26	7
4	16	26	16	4
1	4	7	4	1

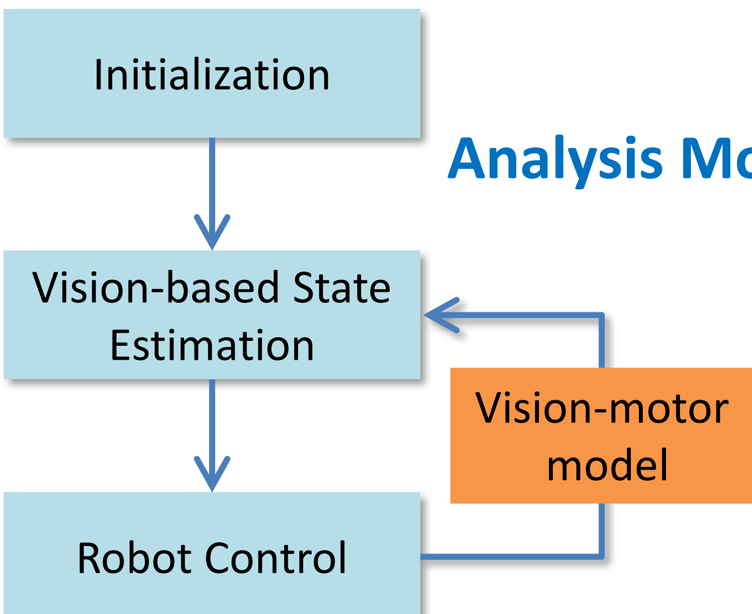
Solution Overview

Example: Object Following

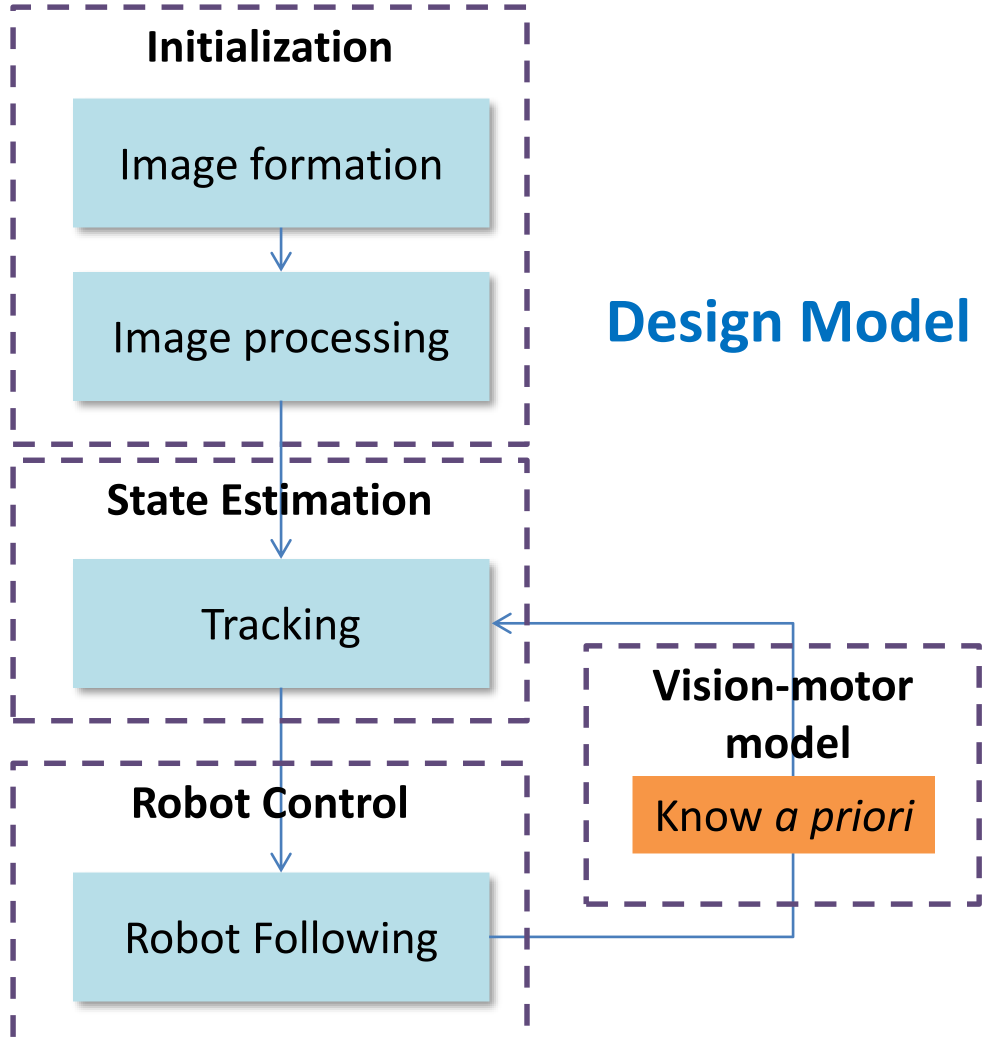
Dr. Ying Wang
Jul. 27th 2016

Task : controlling the robot to follow the target

What does the scene look like?
Which is the target object there?
Where is it located?
How should the robot be controlled?



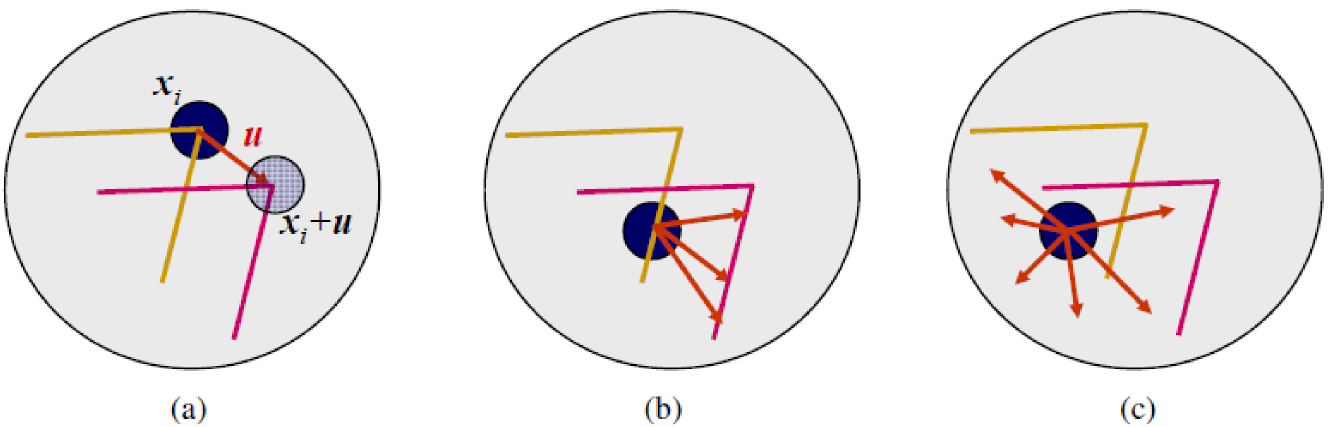
Analysis Model



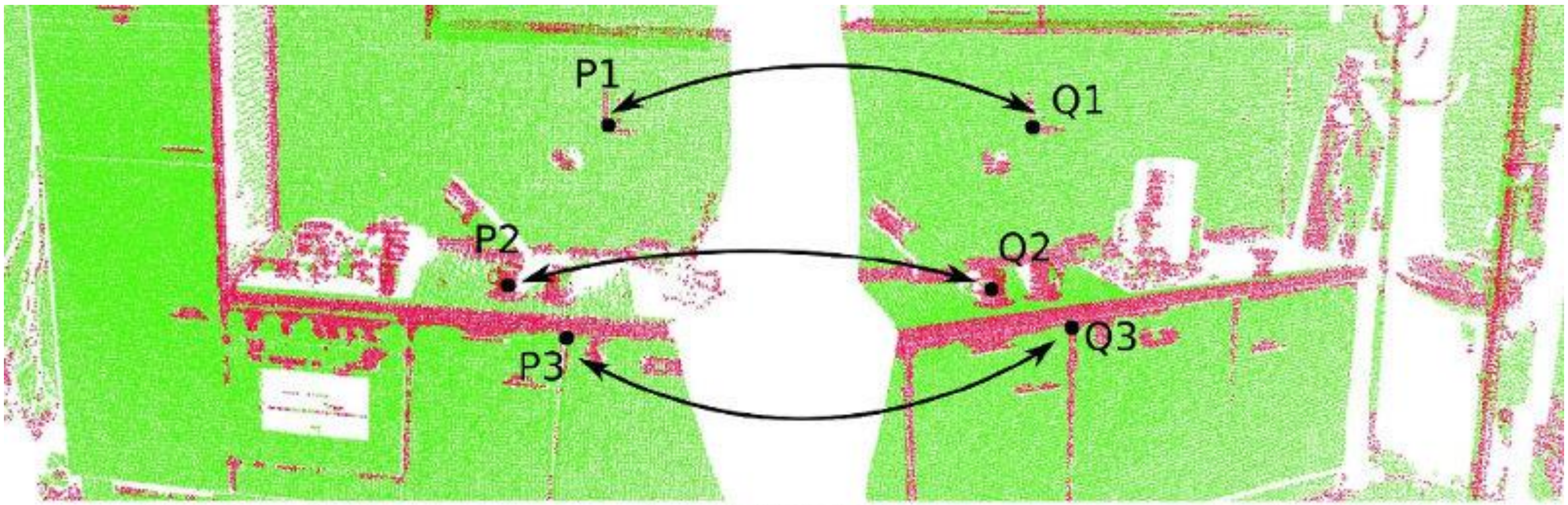
Design Model

Tracking
Keypoints

Dr. Ying Wang
Jul. 27th 2016



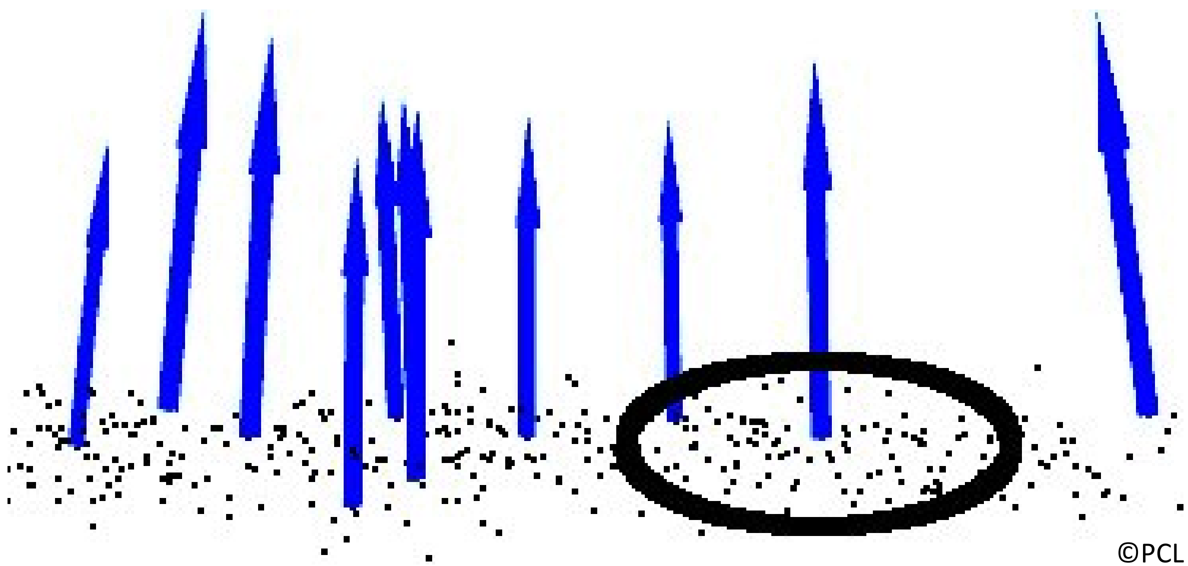
T = t1 T = t2



By including the surrounding neighbors, the underlying sampled surface geometry can be inferred and captured in the feature formulation, which contributes to solving the ambiguity comparison problem. Ideally, the resultant features would be very similar (with respect to some metric) for points residing on the same or similar surfaces, and different for points found on different surfaces, as shown in the figure below.

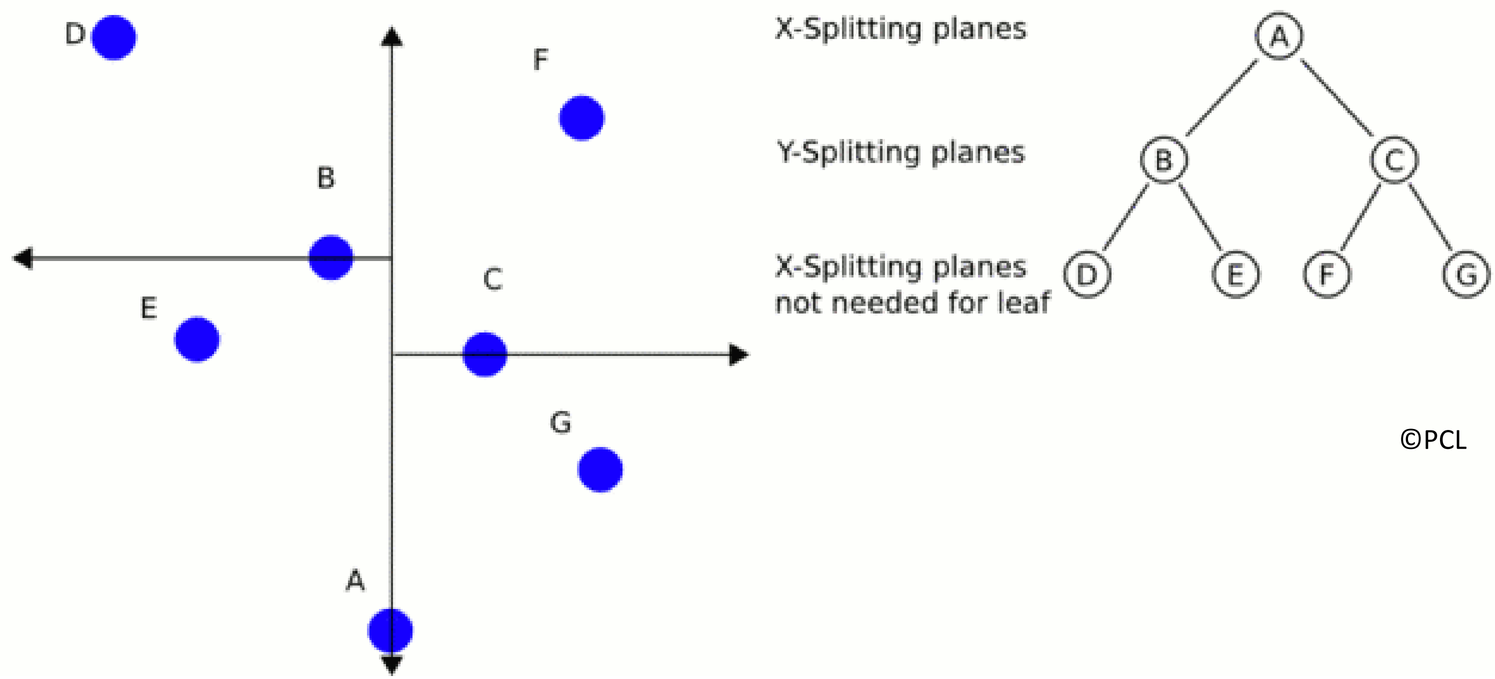
Dr. Ying Wang
Jul. 27th 2016

3D features are representations at a certain 3D point or position in space, which describe geometrical patterns based on the information available around the point. The data space selected around the query point is usually referred as the **k-neighborhood**.



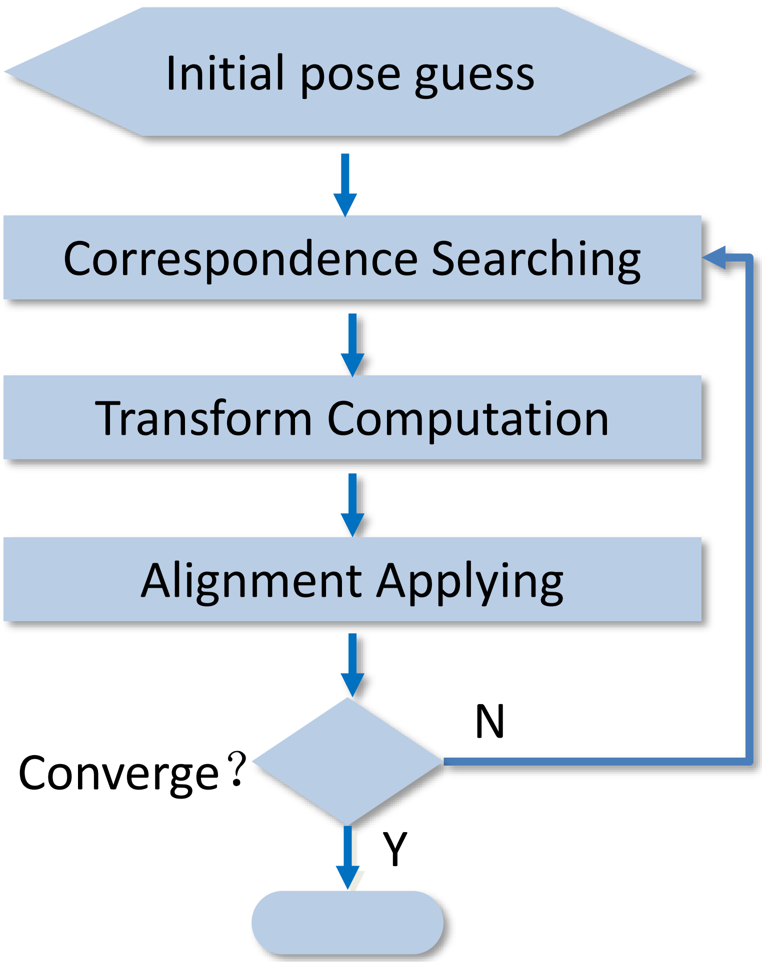
Tracking

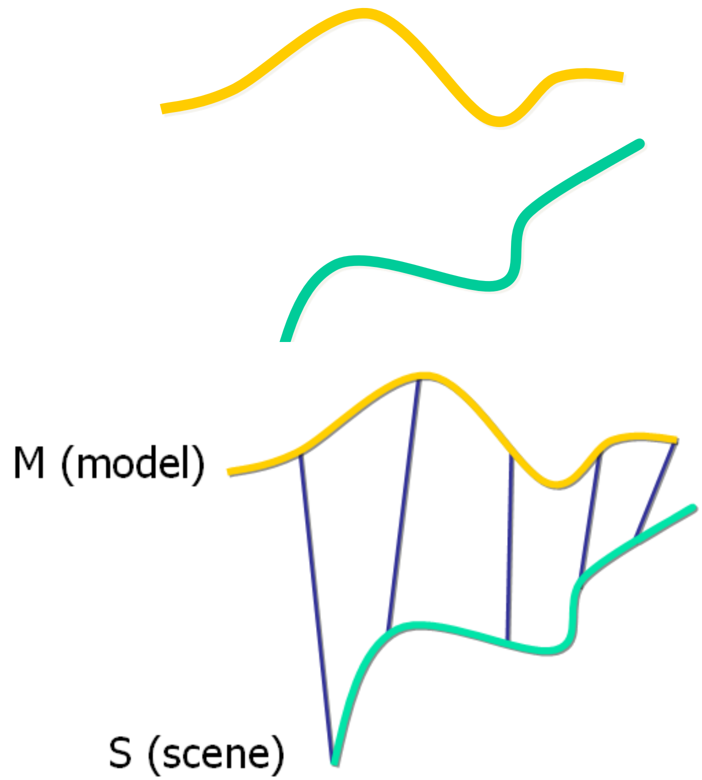
Local Descriptor - K-neighborhood



Tracking

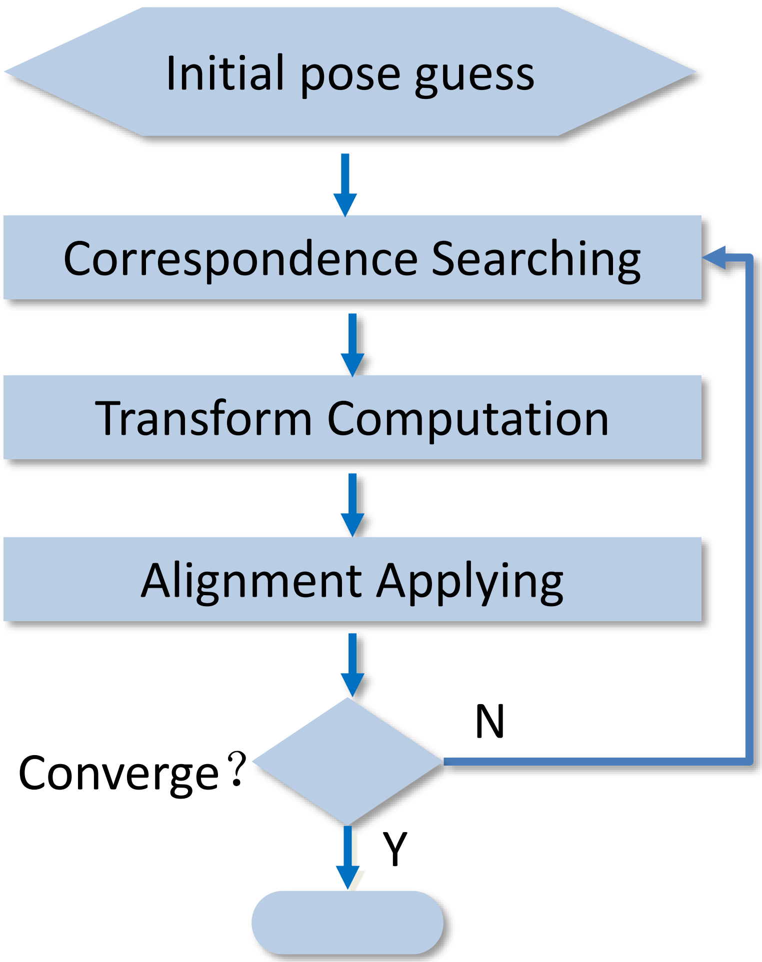
Matching – Iterative Closest Point

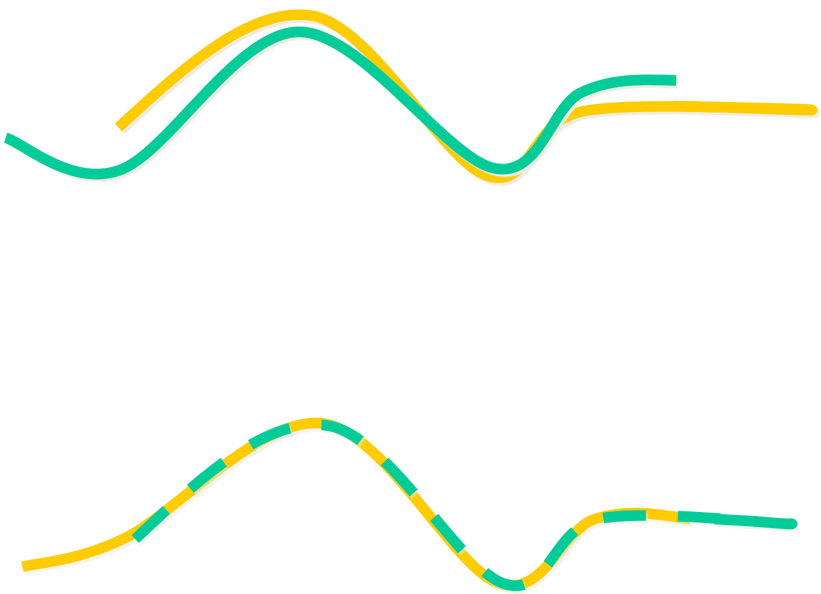




Tracking

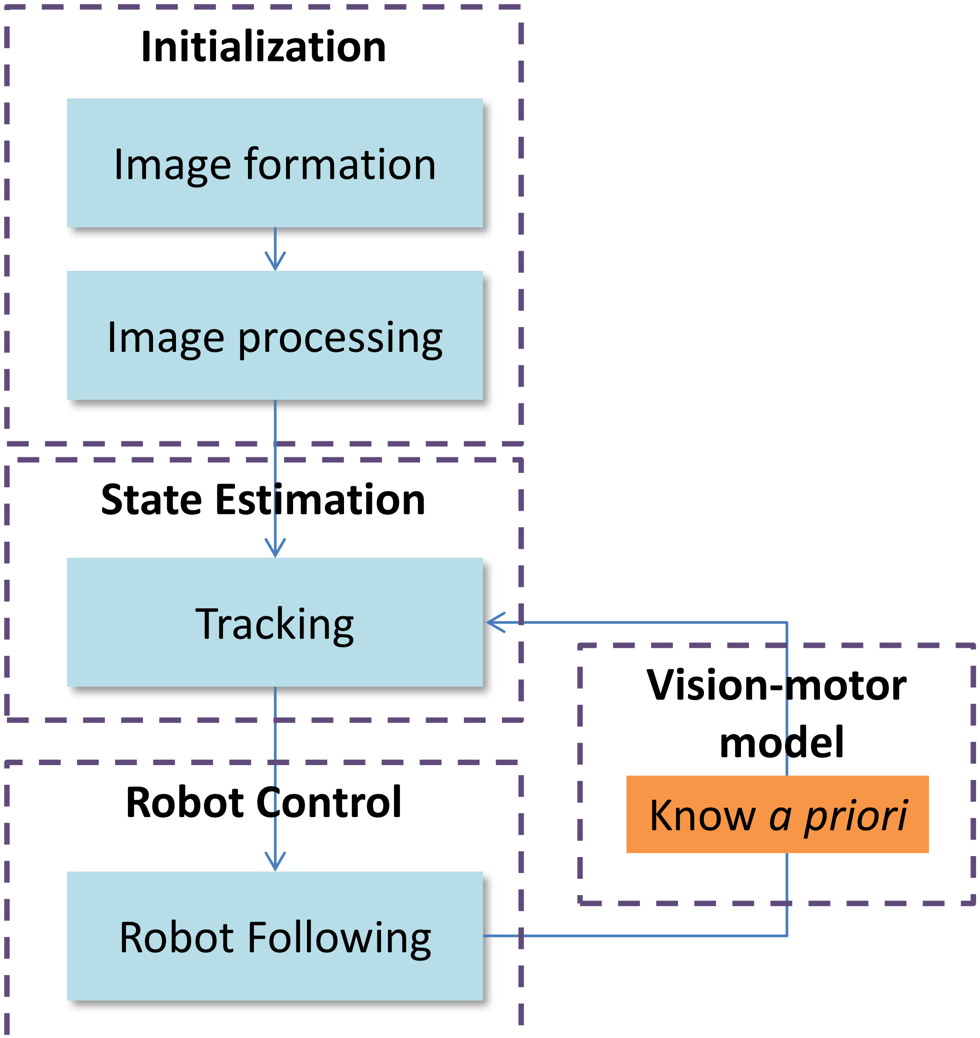
Matching – Iterative Closest Point

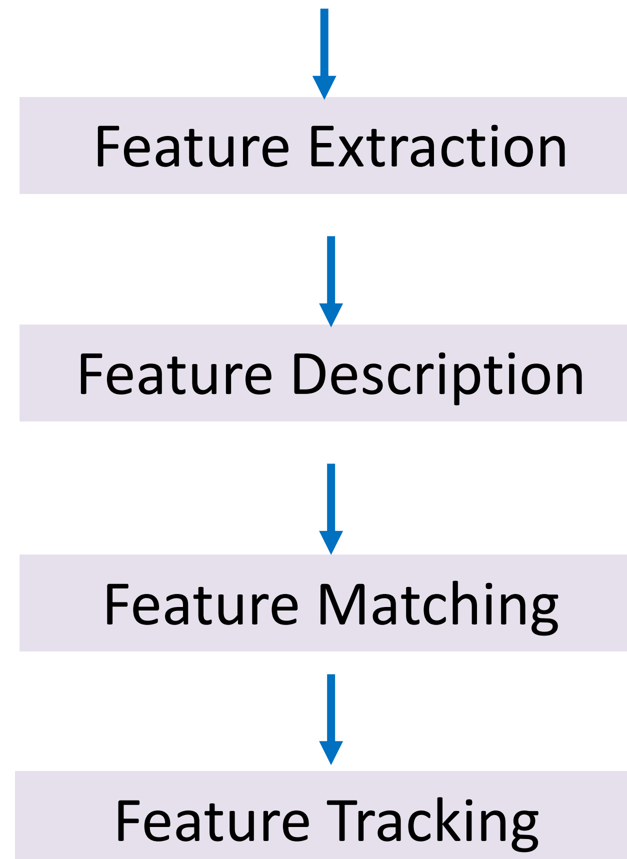




Tracking

Workflow





Solution Model of Machine learning

Dr. Ying Wang
Jul. 27th 2016

Model

World state w is continuous (3D pose) -> Regression model
Taking a generative approach, the likelihoods are described as

$$P_r(x \mid \omega = k)$$

Learning algorithm

the parameters from training data pairs $\{w_i, x_i\}_{i=1}^{I}$ where the pixels have been manually labeled. The prior parameter is learned from the world states $\{w_i\}_{i=1}^{I}$.

Inference algorithm aims to calculate the 3D pose of the object in the video stream.

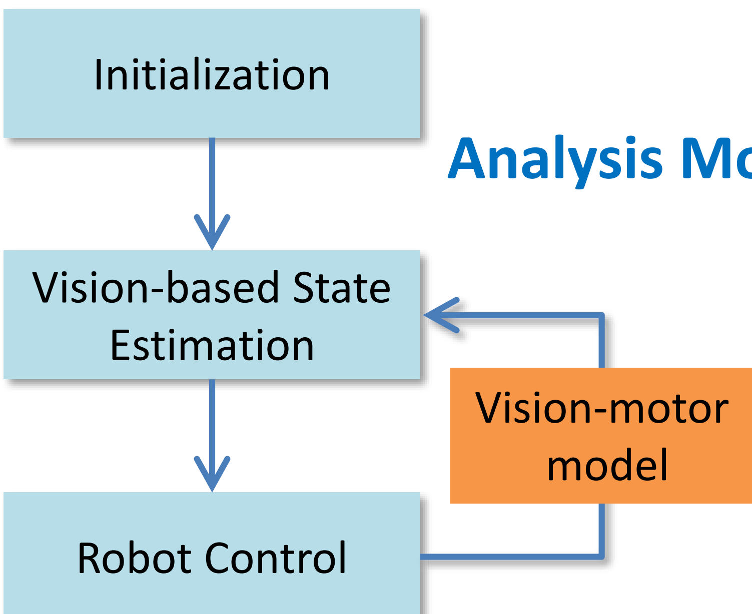
Solution Overview

Dr. Ying Wang
Jul. 27th 2016

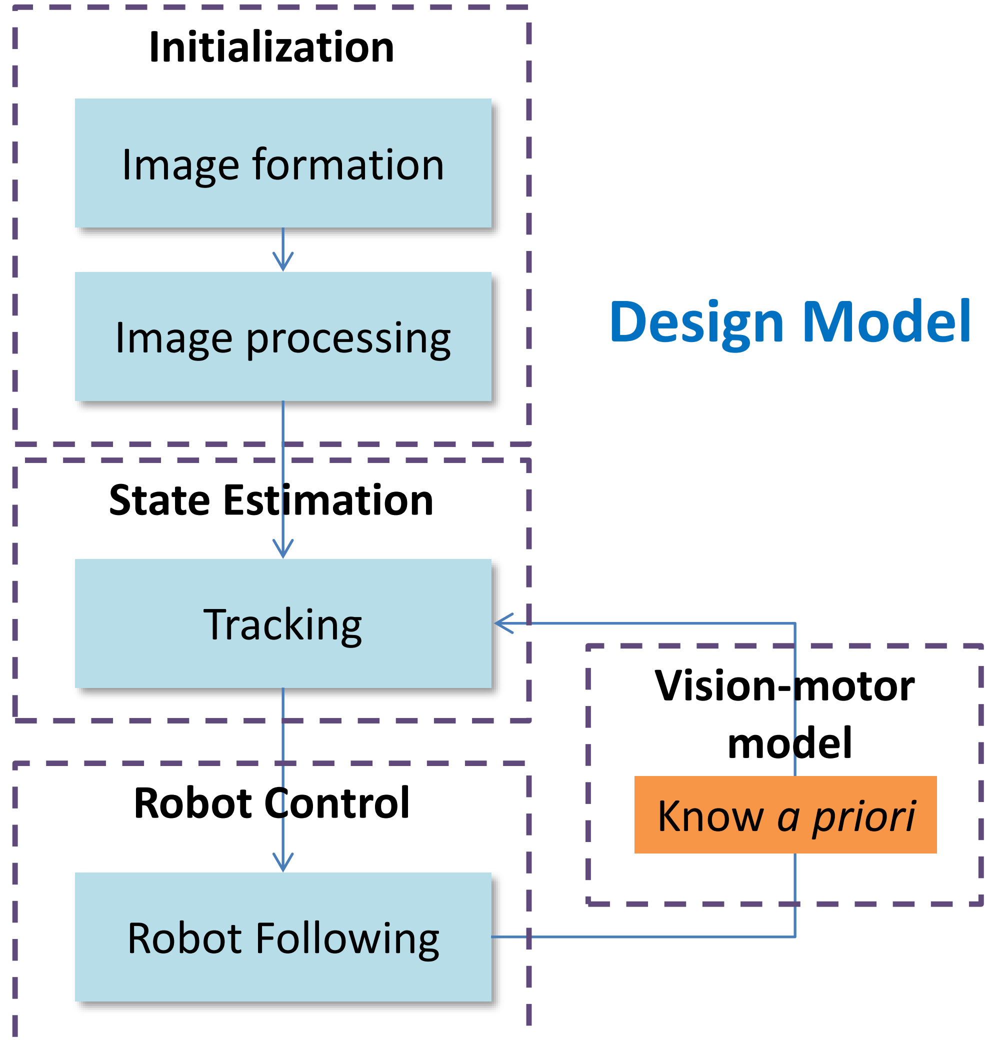
Example: Object Following

Task : controlling the robot to follow the target

What does the scene look like?
Which is the target object there?
Where is it located?
How should the robot be controlled?



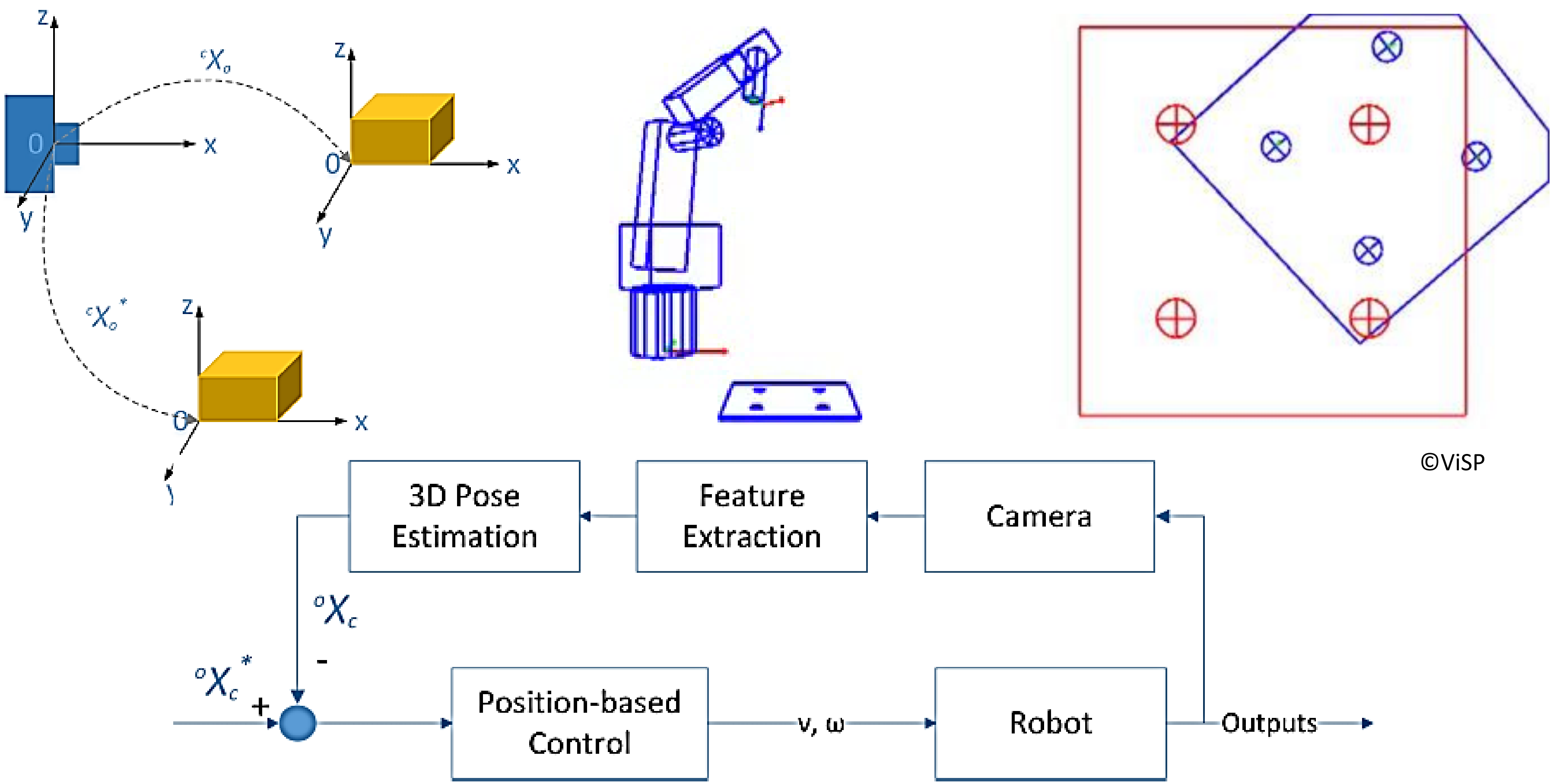
Analysis Model



Solution Model of Machine learning

Visual Servoing

Vision-motor model: PBVS & IBVS



Solution Model of Machine learning

2 ½ D Visual Servoing

Pseudo codes - Initialization

1. set *projModel* ← *perspectiveProjwithDistortion*
2. set *robot* ← *projModel*
3. set *point[4]* //3D points
4. set *dot[4]*
5. compute *cMo*
6. set *P* ← (0, 0, 0)
7. set *cdMo*
8. compute *pd* ← *cdMo, P*
9. compute *Zd* from *P*
10. compute *p* ← *cMo, P*
11. compute *Z* from *P*
12. compute *depth, tu*
13. set *task.addFeature* ← *(p, pd, depth, tu)*

Solution Model of Machine learning

Visual Servoing

Dr. Ying Wang
Jul. 27th 2016

Pseudo codes – Control design

1. set *lamda* ← *(2.5, 0.2, 40)*
2. set *task.setServo* ← *EYEINHAND_L_cVe_eJe*
3. set *task.set_cVe(cVe)* ← *robot.set_cVe(cVe)*
4. set *task.set_eJe(eJe)* ← *robot.set_eJe(eJe)*
5. set *robot.setRobotState* ← *STATE_VELOCITY_CONTROL*

Solution Model of Machine learning

Visual Servoing

Dr. Ying Wang
Jul. 27^{th} 2016

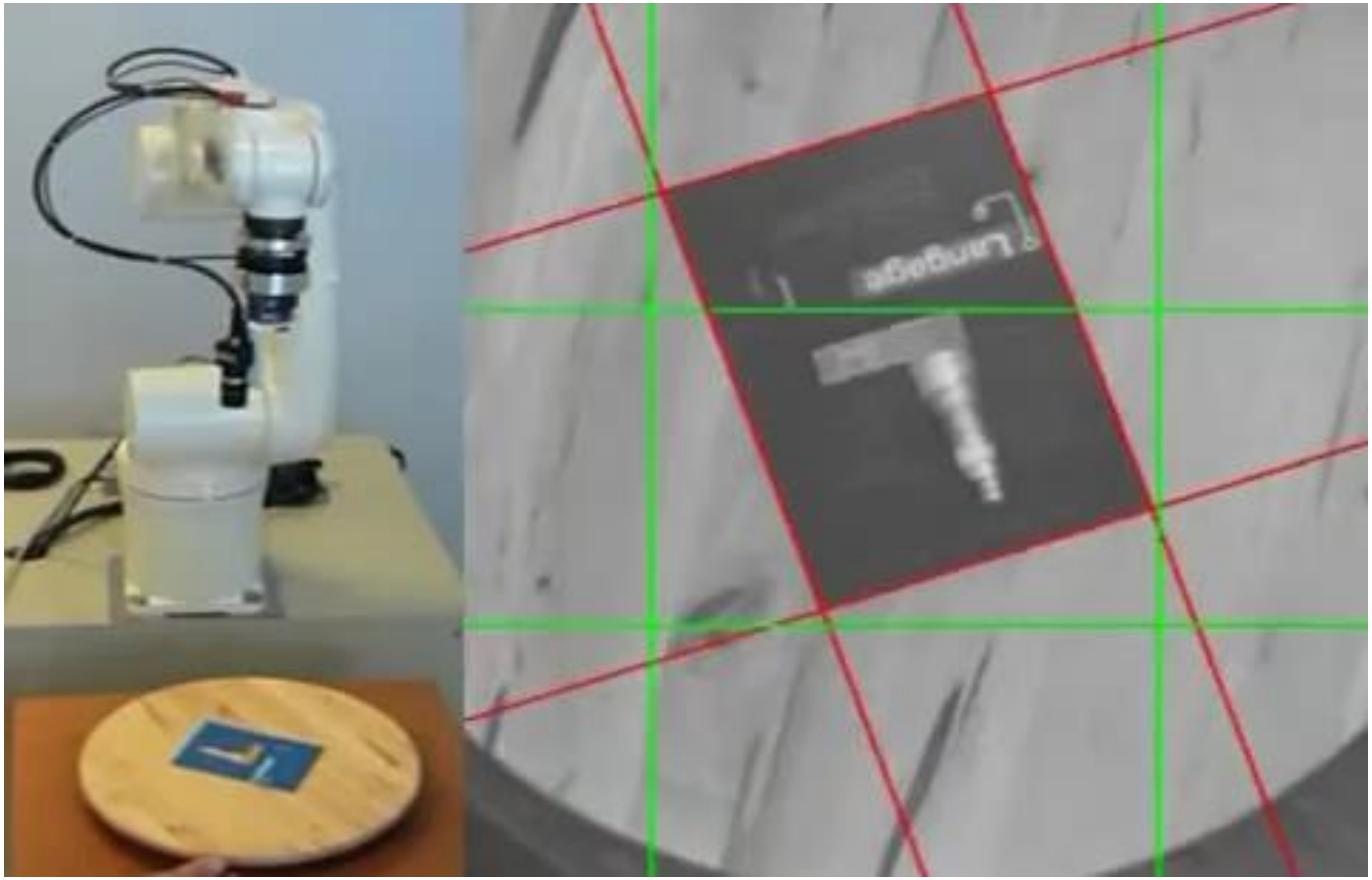
Pseudo codes – Control loop

```
1.  while true
2.          for all feature points
3.                get dot[i].x
4.                get dot[i].y
5.          compute & update cMo
6.          Compute & update p
7.          Compute & update tu
8.          Compute & update depth
9.          update task.set_cVe(cVe) ← robot.set_cVe(cVe)
10.         update task.set_eJe(eJe) ← robot.set_eJe(eJe)
11.         compute v
12.         set robot.setVelocity
```

Solution Model of Machine learning

Dr. Ying Wang
Jul. 27th 2016

Object Following



©ViSP

Robotic Vision by ROS

- Drivers
  - 2D/3D range finders
  - RGB-Depth cameras
  - monocular and stereo cameras

- API
  - Tools (pcl, visp, opencv with ros)
  - Support packages (calibration, recognition, image conversion, visualizer)
  - Messages
  - Topics
  - Services
  - parameters

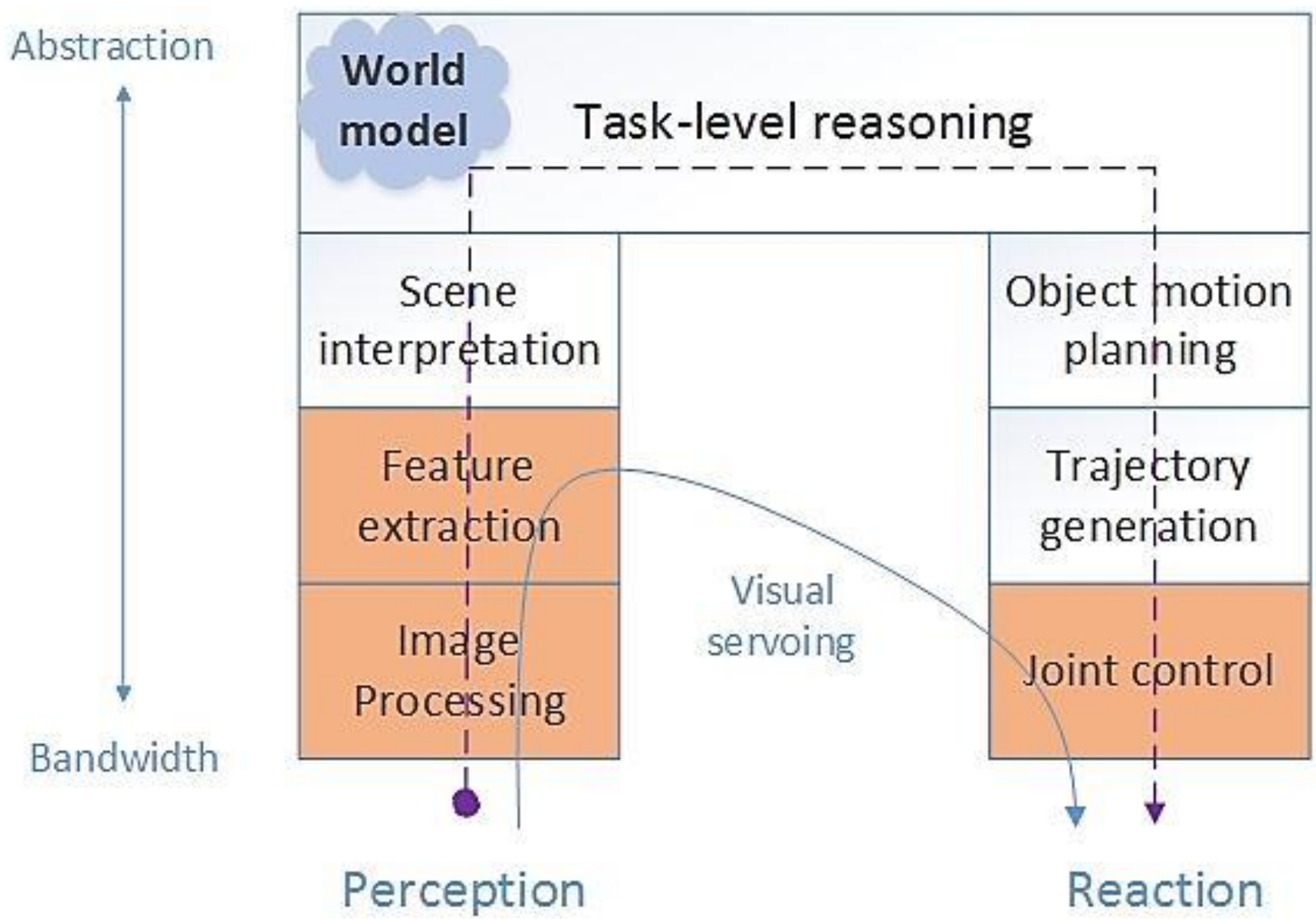
- Tutorials & support

www.**roswiki**.com/

Contents

1. Introduction
2. Modeling a Robotic Vision Problem
3. Solution Proposal
4. Summary

Solution Model of Machine learning



谢谢！

Thank you for your attention!

Vielen Dank für Ihre Aufmerksamkeit!