Assistive Robotics

Visual Servoing

Área de Ingeniería de Sistemas y Automática
Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza

Departamento de Informática e Ingeniería de Sistemas Universidad Zaragoza

Gonzalo López Nicolás

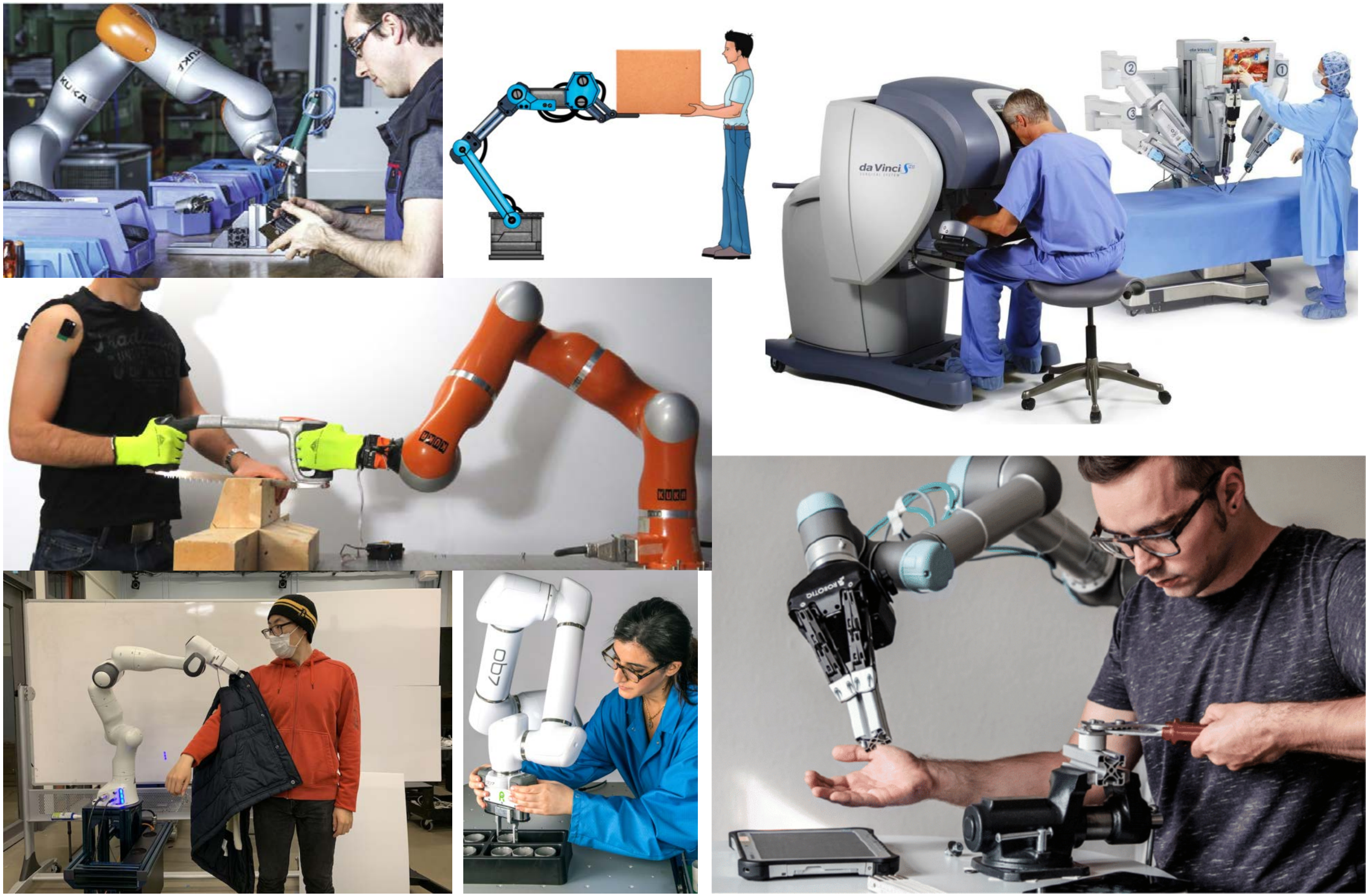
VISUAL SERVOING

- Introduction
- Robotic manipulation
- Vision systems
- Classification of visual control systems
  - Position-Based Visual Servoing
  - Image-Based Visual Servoing
    - Stability analysis
  - Hybrid Visual Servoing
- Conclusion and bibliography

Index

Introduction: Assistive robotics

Introduction: Assistive robotics

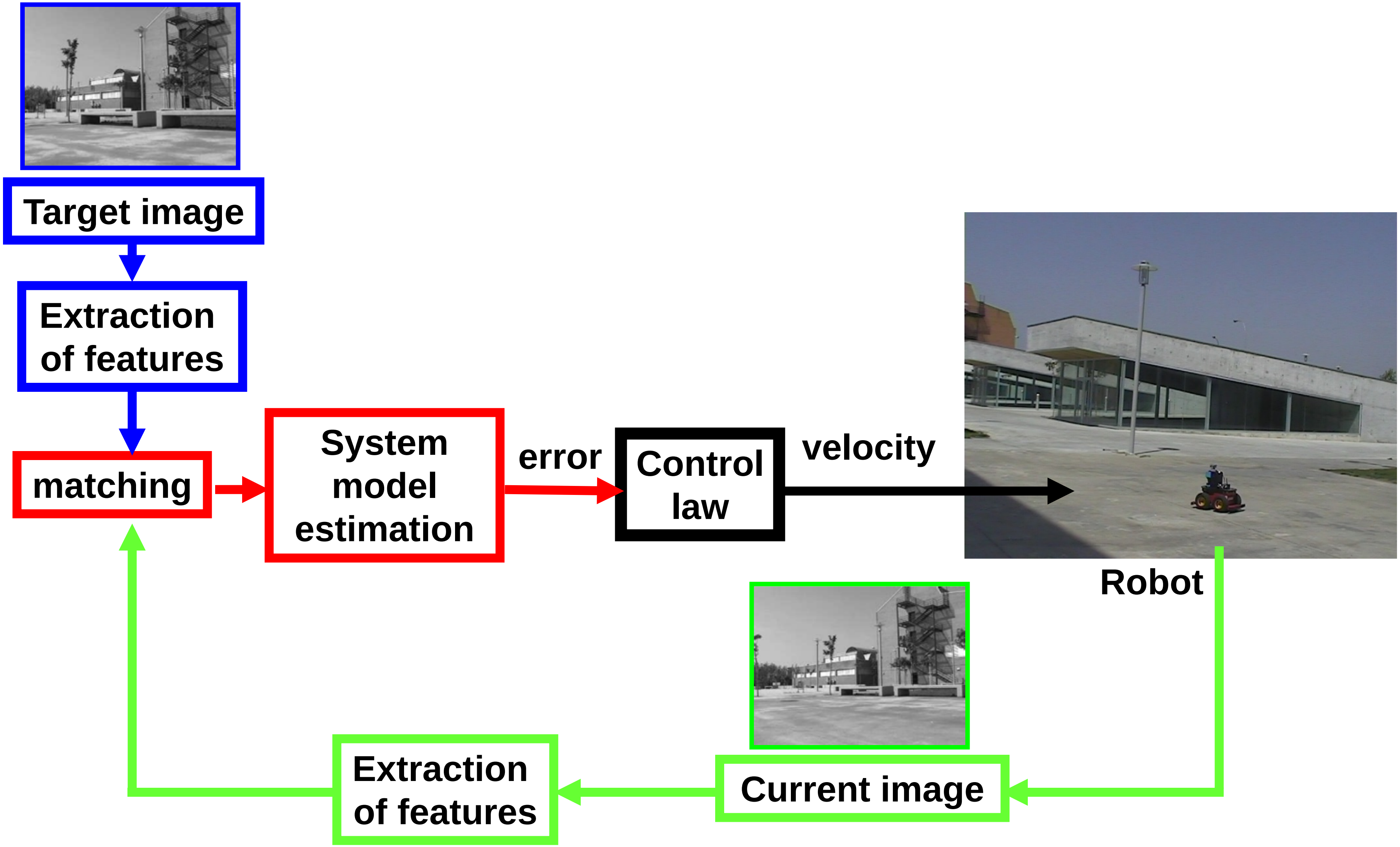


Introduction: Visual Servoing

Open loop

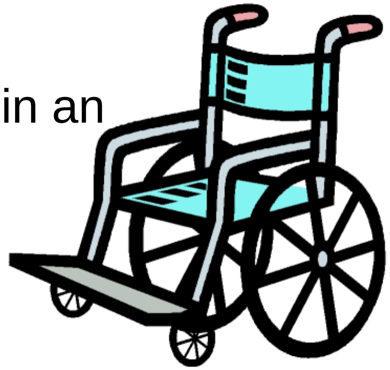
Closed loop

Introduction: Visual Servoing



Introduction: Visual Servoing

- Visual servoing in the market
  - Robotic systems have contributed substantially to increasing the accuracy and speed of automated processes mainly in the manufacturing industry.
  - They generally require a detailed description of the workspace and the objects handled.
    - It requires considerable human and financial resources
  - There is a lack in the perception capability of current robotic systems in terms of:
    - Unknown or dynamic environments, undefined locations, calibration errors, etc.
    - The irruption of visual control responds to these challenges
  - Visual servoing:
    - The objective is the positioning of a robot's end-effector in an unstructured environment.



Index

- Introduction

- **Robotic manipulation**
  - Basics
  - Spatial localization
  - Kinematic model

- Vision systems
- Classification of visual control systems
  - Position-Based Visual Servoing
  - Image-Based Visual Servoing
    - Stability analysis
  - Hybrid Visual Servoing
- Conclusion and bibliography

Robotic manipulation

- Types of robots
  - Mobile robots
    - Terrestrial
      - Wheeled vehicles, crawlers, ...
      - Humanoids, legged
    - Aerial
      - Drones, quadrotors, ...
    - Submarines
  - Robotic manipulators
    - Industrial robots
    - <u>Collaborative robots</u>



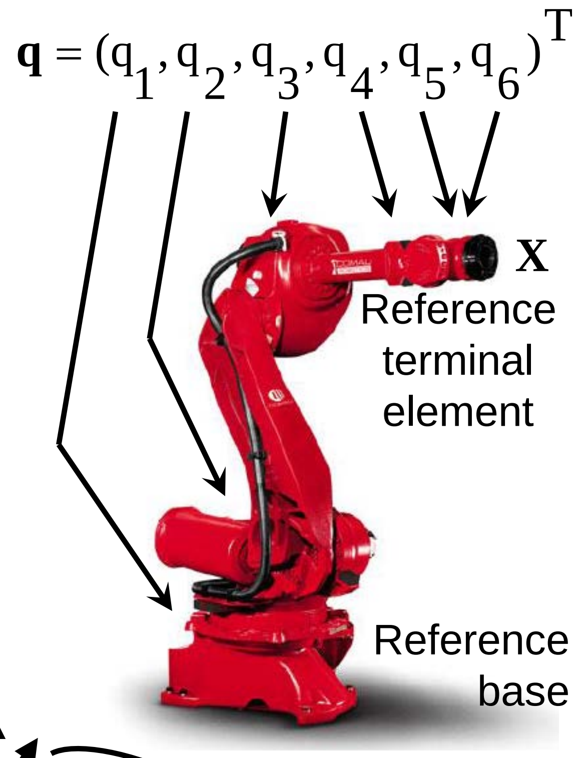






Robotic manipulation: Spatial localization

- Geometric model to locate the terminal element (position and orientation).
- Joint coordinates: $\mathbf{q} = (q_1, q_2, \ldots, q_n)^T$
  - With n the number of links and joints.
- Coordinates in the task space, operational or Cartesian. $\mathbf{X} = (x, y, z, rx, ry, rz)^T$
- Reference systems and transformations
  - Reference system associated to each link of the robot.
  - Obtain location from other relative locations

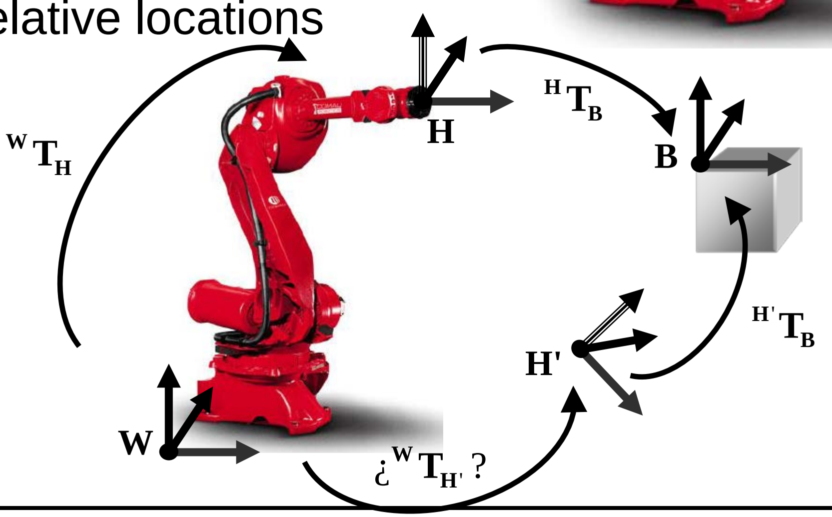


$$ {}^{W}\mathbf{T}_{H} = \begin{pmatrix} \mathbf{R} & \mathbf{p} \\ \mathbf{0} & 1 \end{pmatrix} $$

$$ ¿\,{}^{W}\mathbf{T}_{H'}\,? $$

$$ {}^{W}\mathbf{T}_{H} \cdot {}^{H}\mathbf{T}_{B} = {}^{W}\mathbf{T}_{H'} \cdot {}^{H'}\mathbf{T}_{B} $$

$$ {}^{W}\mathbf{T}_{H'} = {}^{W}\mathbf{T}_{H} \cdot {}^{H}\mathbf{T}_{B} \cdot \left({}^{H'}\mathbf{T}_{B}\right)^{-1} $$

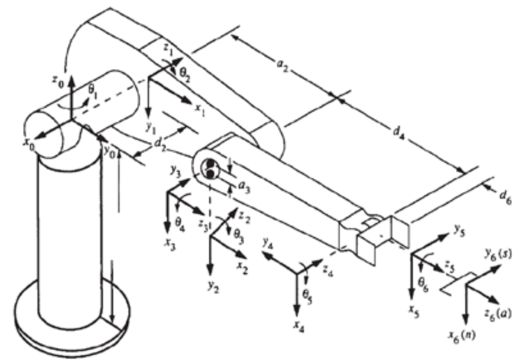


Robotic manipulation: Kinematic model

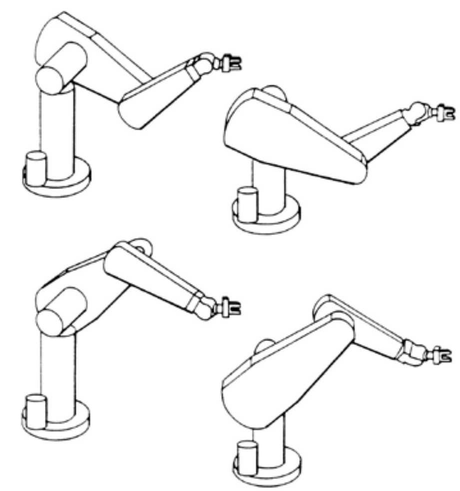
- **<u>Direct</u>** geometric model: $\mathbf{X} = \mathrm{f}(\mathbf{q})$
  - Motion analysis (compute $\mathbf{X}$ from $\mathbf{q}$):
  - Denavit-Hartenberg (1955) reference assignment.
    - Kinematic chain of links connected with joints.
    - Association of a reference system to each link

$$ {}^{0}\mathbf{T}_{n} = {}^{0}\mathbf{T}_{1}\ {}^{1}\mathbf{T}_{2} \cdots {}^{n-1}\mathbf{T}_{n} $$

    - Normalized representation of transformations between consecutive links.



- **<u>Inverse</u>** geometrical model: $\mathbf{q} = \mathrm{f}^{-1}(\mathbf{X})$
  - Motion control (control $\mathbf{q}$ to obtain $\mathbf{X}$):
  - No systematic method to obtain it and in general no unique solution.
  - System with n unknowns and 12 possible nonlinear equations (at most 6 independent unknowns).



Robotic manipulation: Kinematic model

- Differential model

Velocities of the joints		Velocity of the robot hand
$\dot{\mathbf{q}} = (\dot{q}_1, \dot{q}_2, \cdots, \dot{q}_n)$	Direct Jacobian → ← Inverse Jacobian	$\mathbf{V} = (v_x, v_y, v_z, \omega_x, \omega_y, \omega_z)$

- Direct Jacobian:
  - Hand velocities as a function of joint velocities.

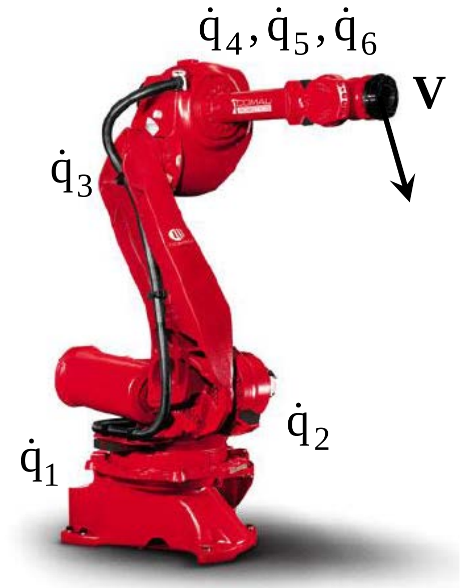
$$\mathbf{V} = \mathbf{J}(\mathbf{q}) \cdot \dot{\mathbf{q}}$$

- Jacobian matrix

$$\mathbf{X} = f(\mathbf{q})$$

$$\mathbf{J}(\mathbf{q}) \equiv \frac{\partial f(\mathbf{q})}{\partial \mathbf{q}}$$

$$\begin{pmatrix} v_x \\ v_y \\ v_z \\ \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} = \mathbf{J}(\mathbf{q}) \cdot \dot{\mathbf{q}} = \begin{bmatrix} \frac{\partial f_x}{\partial q_1} & \cdots & \frac{\partial f_x}{\partial q_n} \\ \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots \end{bmatrix} \cdot \begin{pmatrix} \dot{q}_1 \\ \vdots \\ \dot{q}_n \end{pmatrix}$$



- Inverse Jacobian: Joint velocities as a function of hand velocities.

$$\mathbf{V} = \mathbf{J}(\mathbf{q}) \cdot \dot{\mathbf{q}} \quad \Rightarrow \quad \dot{\mathbf{q}} = \mathbf{J}^{+}(\mathbf{q}) \cdot \mathbf{V}$$

Index

- Introduction
- Robotic manipulation

- **Vision systems**
  - **Why vision?**
  - **Camera model**

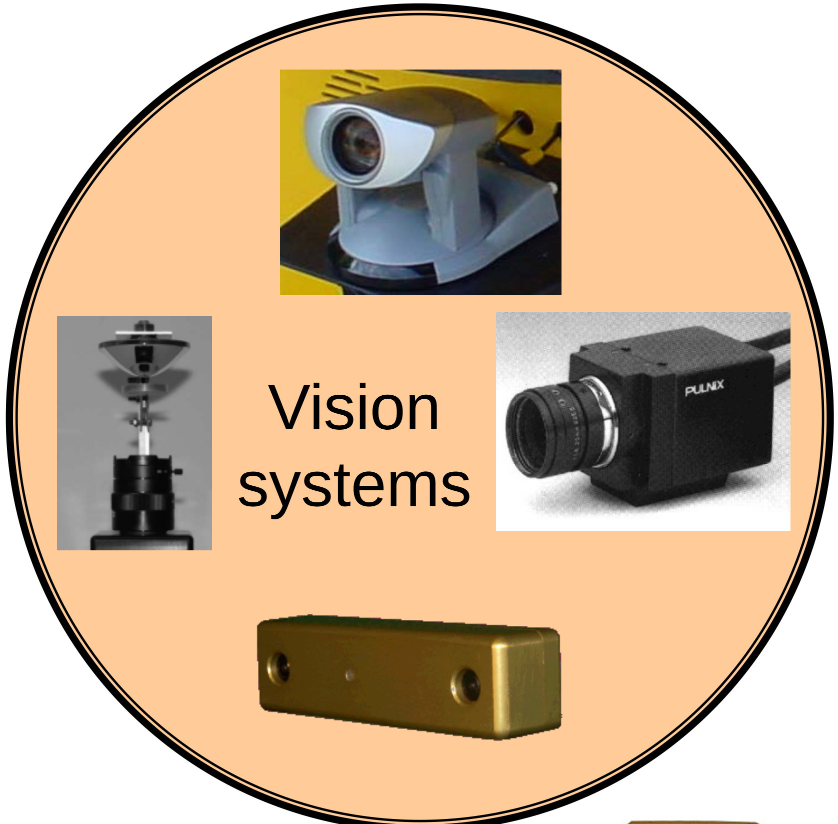
- Classification of visual control systems
- Position-Based Visual Servoing
- Image-Based Visual Servoing
- Stability analysis
- Hybrid Visual Servoing
- Conclusion and bibliography

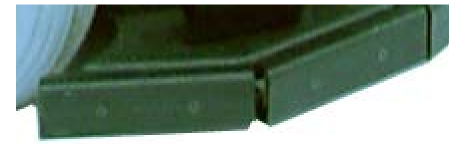
Vision systems

- Sensors

Ultrasounds

Laser





Contact sensor

3D Laser

Odometer

Infrared

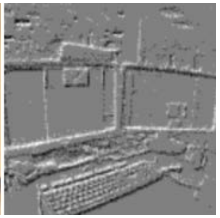
GPS

IMU

Vision systems

- **Why vision?**
  - Great richness of information
  - Real-time processing
  - Versatile perception system
  - Miniaturization prospects
  - Wide variety of camera types
  - Low cost

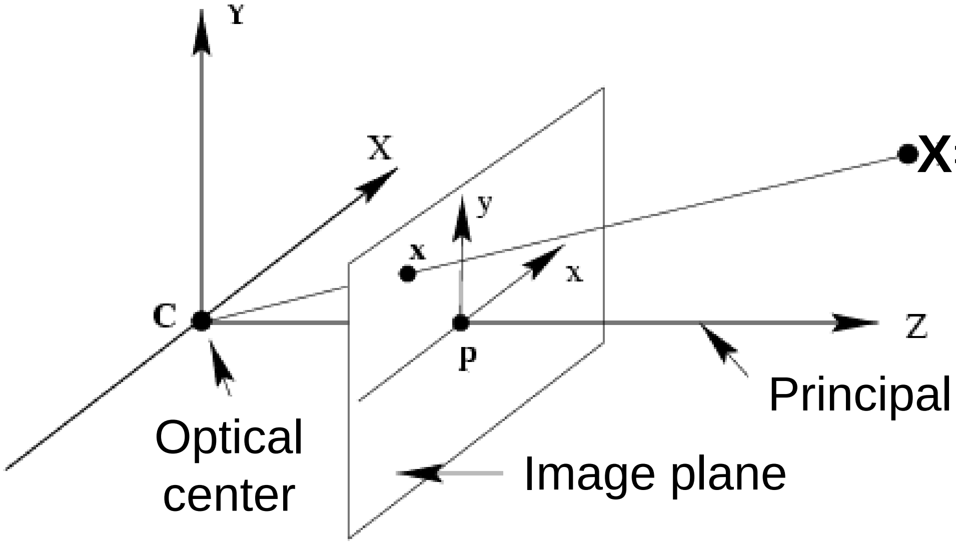


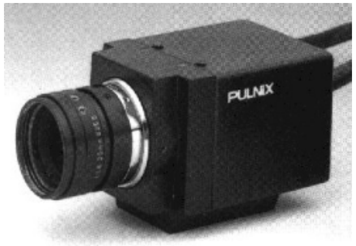






Vision systems: Camera model

- Pinhole camera
  - Captures a set of rays passing through the **center of projection** (center of the camera, focal point).
  - Image is formed in the **image plane**
  - **Pinhole** model as approximation of the actual camera





Intrinsic parameters | Perspective Projection | Extrinsic parameters

- Perspective projection:

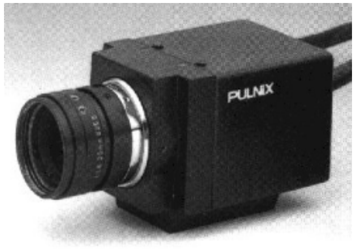
$$\begin{bmatrix} u \\ v \\ s \end{bmatrix} = \lambda \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{CW} & \mathbf{t}_{CW} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Index

Classification of visual servoing systems

Robot

+



Vision system

+

Control law?

=

Visual servoing

Classification of visual servoing systems

Categories:

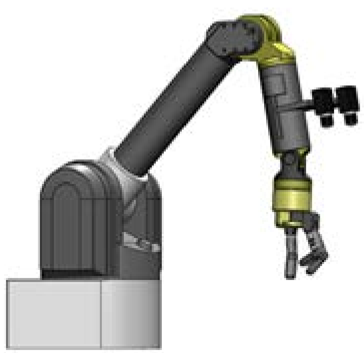
- **According to the configuration of the camera and the robot**
  1. **Camera in hand**
  2. **Camera to hand**
  3. **Multiple cameras**
- According to the hierarchical structure of the vision system and the joint control of the robot
  1. Open loop control: perception and action
  2. Systems of dynamic perception and action
  3. Visual direct control
- According to the definition of the error signal
  1. Position-based visual servoing
  2. Image-based visual servoing
  3. Hybrid visual servoing

Classification of visual servoing systems

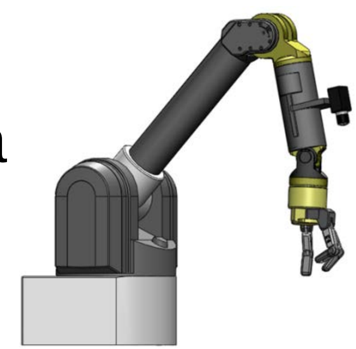
- Depending on the camera and robot configuration

1. <u>Camera in hand</u>:
   - Advantages
     - The camera can observe the movement of the terminal element with fixed resolution and without occlusions.
   - Disadvantages
     - Variable camera position.
     - Field of view can change radically with robot movement.

Stereo cameras in hand



Camera in hand

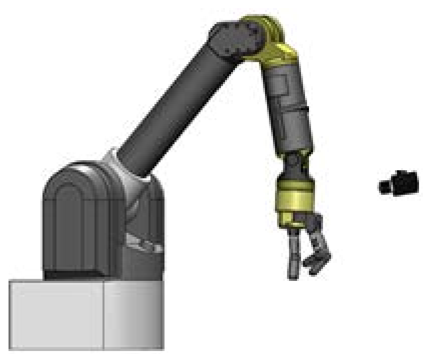


- Depending on the camera and robot configuration

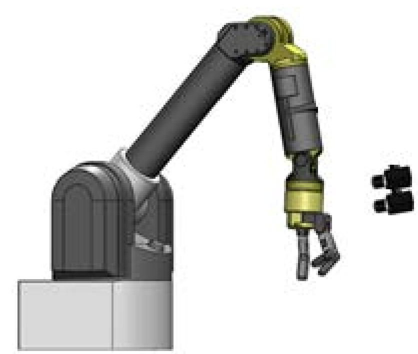
  2. Camera to hand:

     - Advantages
       - Field of view remains constant and independent of robot movement.
       - Constant camera position. Offline calibration.
     - Disadvantages
       - The robot may cause occlusions in the camera field of view during its movement.

Camera to hand

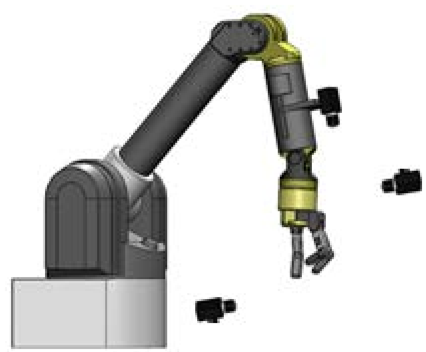


Stereo cameras to hand



- Depending on the camera and robot configuration
  3. Multiples cameras:
     - Advantages
       - Wide combined field of view minimizing occlusions.
     - Disadvantages
       - Additional processing cost
       - Calibration of multiple cameras
       - Synchronization of cameras

Multiples cameras



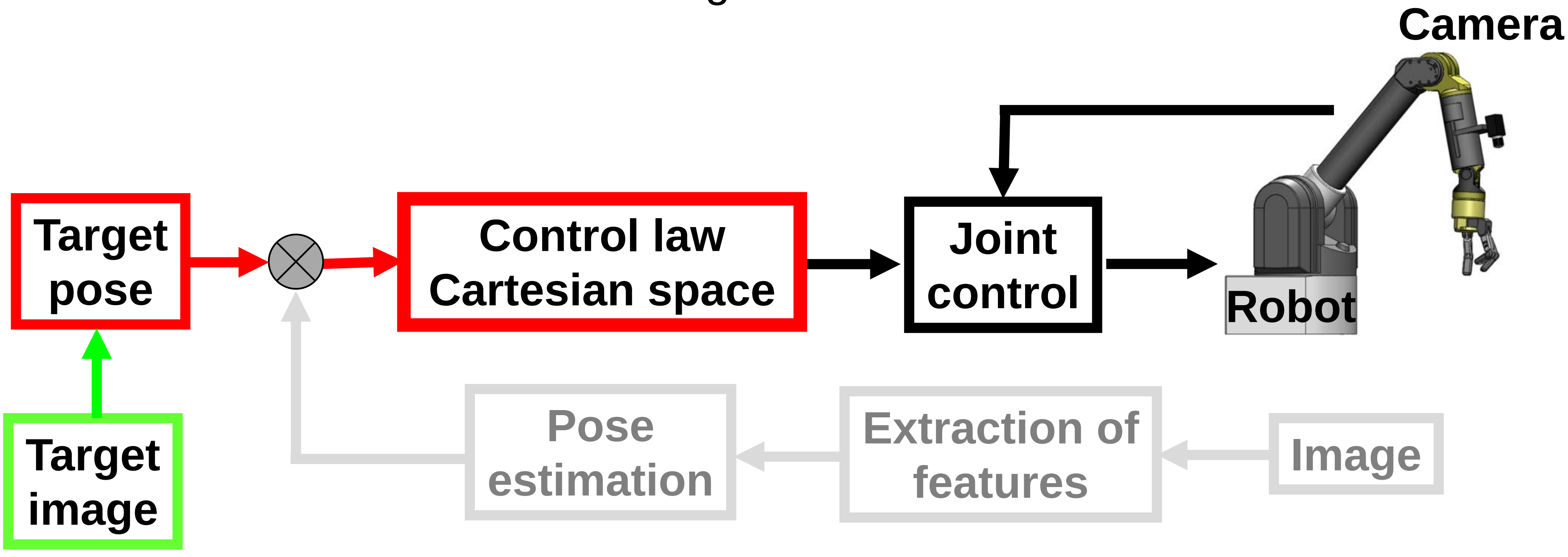
Classification of visual servoing systems

Categories:

- According to the configuration of the camera and the robot
  1. Camera in hand
  2. Camera to hand
  3. Multiple cameras
- **According to the hierarchical structure of the vision system and the joint control of the robot**
  1. **Open loop control: perception and action**
  2. **Systems of dynamic perception and action**
  3. **Visual direct control**
- According to the definition of the error signal
  1. Position-based visual servoing
  2. Image-based visual servoing
  3. Hybrid visual servoing

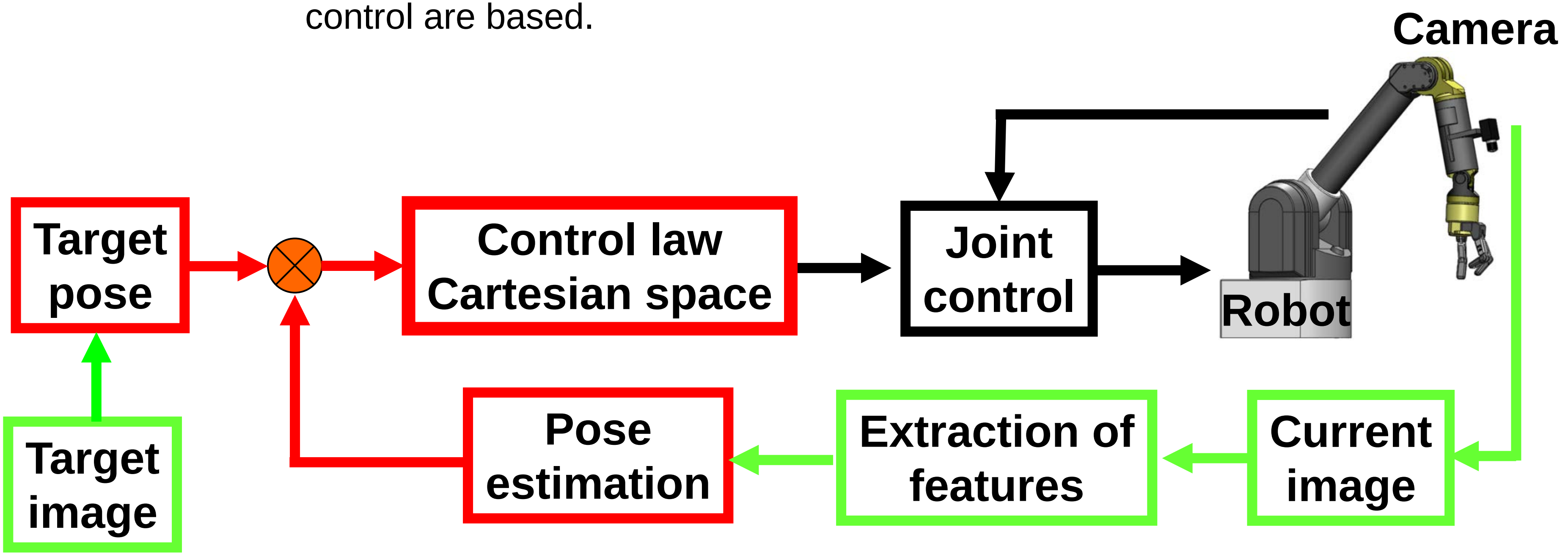
Classification of visual servoing systems

- According to the hierarchical structure of the vision system and the joint control of the robot
  1. <u>Open loop control: perception and action</u>
     - The extraction of information from the image and control of the robot are two sequential tasks
     - The robot executes the task assuming that the working environment is unchanged.



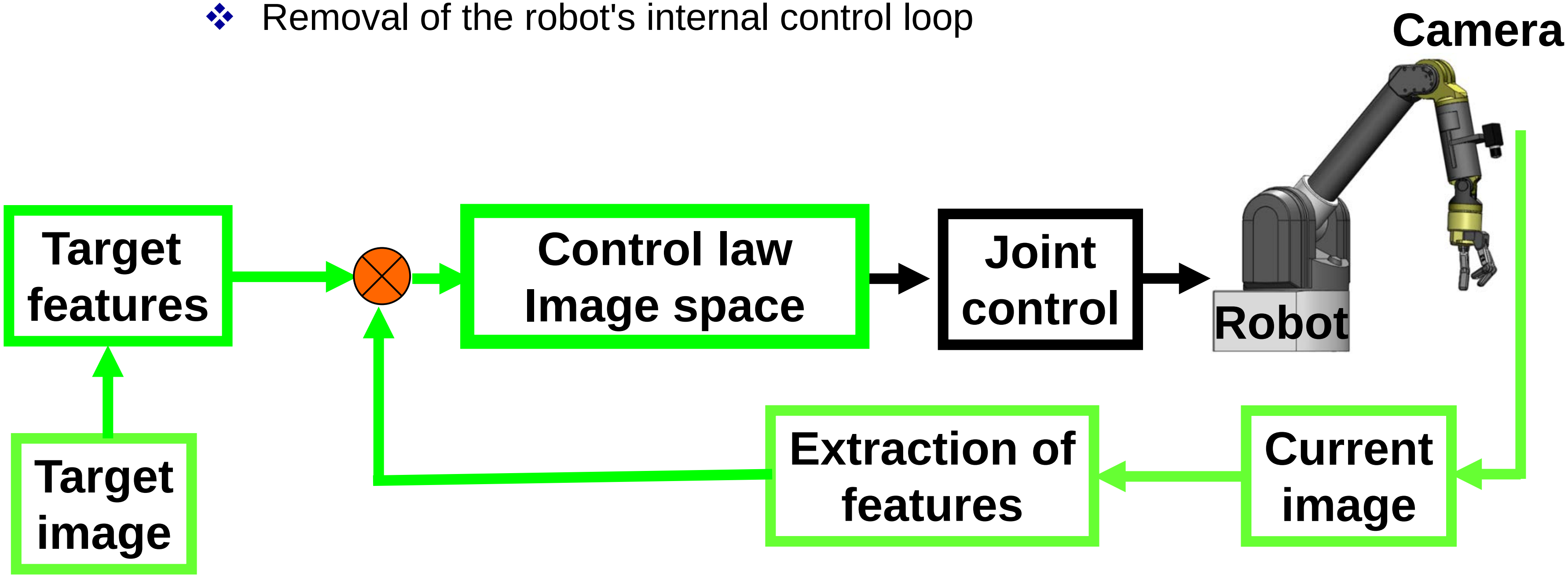
Classification of visual servoing systems

- According to the hierarchical structure of the vision system and the joint control of the robot
  2. <u>Systems of dynamic perception and action</u>
     - Two control loops at different frequencies:
       - Internal servomotor control loop at high frequency.
       - External loop of visual perception at a lower frequency.
       - This is the category on which most implementations in the field of visual control are based.



Classification of visual servoing systems

- According to the hierarchical structure of the vision system and the joint control of the robot
  3. Visual direct control
     - A control loop
       - Visual perception at high frequency
       - Eliminate the robot controller, replacing it with a controller that only uses vision to stabilize the mechanism
       - Removal of the robot's internal control loop



Classification of visual servoing systems

- According to the hierarchical structure of the vision system and the joint control of the robot
- Traditionally there has been a clear predominance of *look and move* (perception and action) systems over direct servo control systems:
  - The slow sampling rate available for vision and the use of complex, nonlinear dynamics by direct servo control systems result in a complex control problem.
  - Look-and-move control methods separate the kinematic peculiarities of the mechanism from visual control, whereas for direct servo control there is a strong dynamic coupling.
  - Internal joint control loops with a high sampling rate exhibit good dynamics.
  - Many controllers have specialized mechanisms to deal with kinematic singularities, thus simplifying system design.

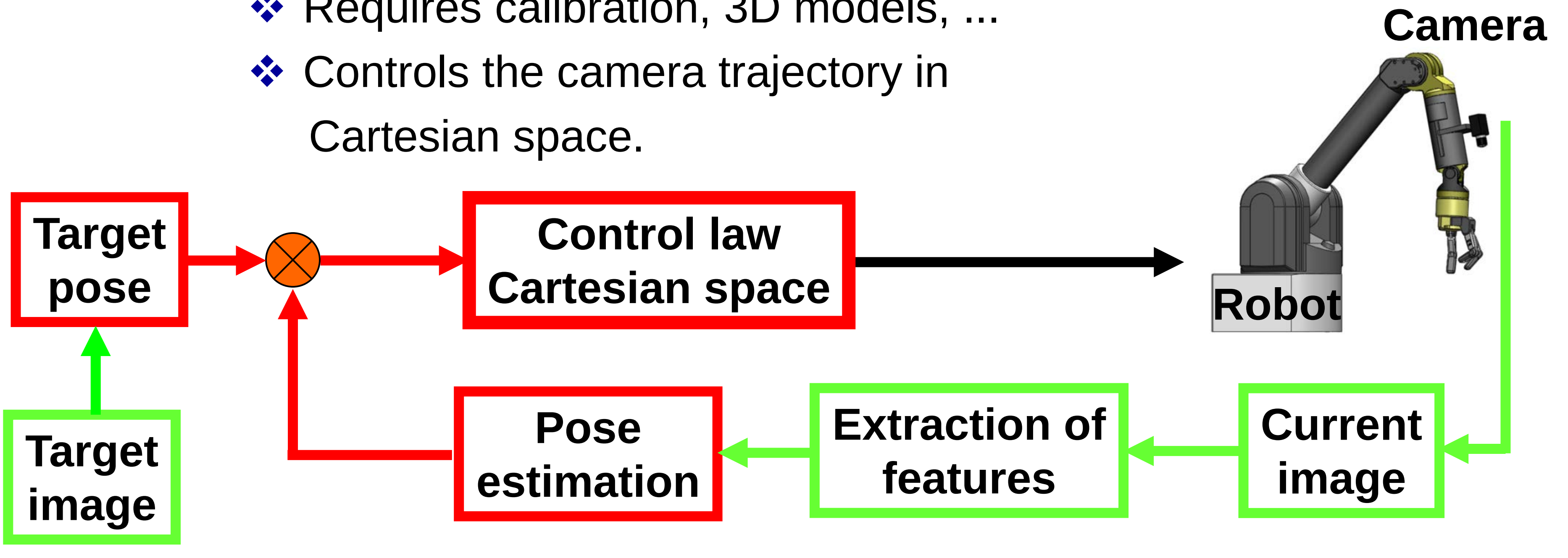
Classification of visual servoing systems

Categories:

- According to the configuration of the camera and the robot
  1. Camera in hand
  2. Camera to hand
  3. Multiple cameras
- According to the hierarchical structure of the vision system and the joint control of the robot
  1. Open loop control: perception and action
  2. Systems of dynamic perception and action
  3. Visual direct control
- **According to the definition of the error signal**
  1. **Position-based visual servoing**
  2. **Image-based visual servoing**
  3. **Hybrid visual servoing**

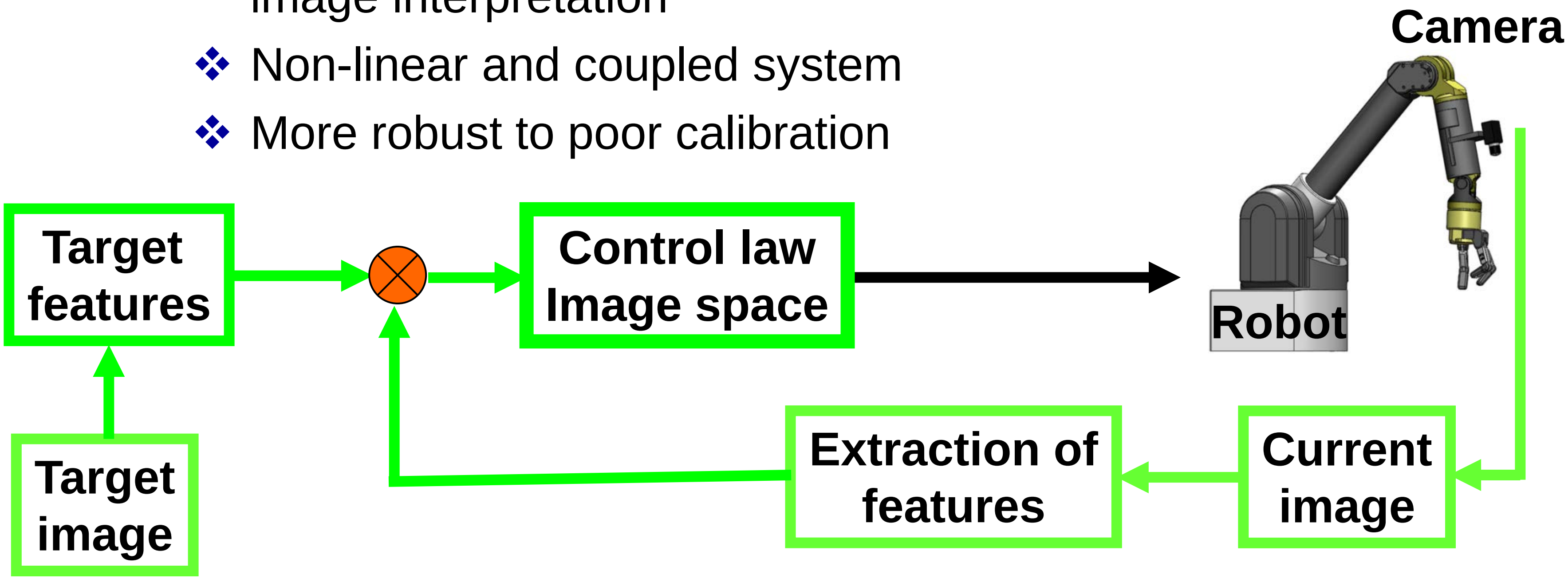
Classification of visual servoing systems

- According to the definition of the error signal
  1. Position-based visual servoing
     - Estimation of position and orientation of the target
     - The task is defined in 3D space
       - Feedback of current position of the terminal element.
       - Requires calibration, 3D models, ...
       - Controls the camera trajectory in Cartesian space.



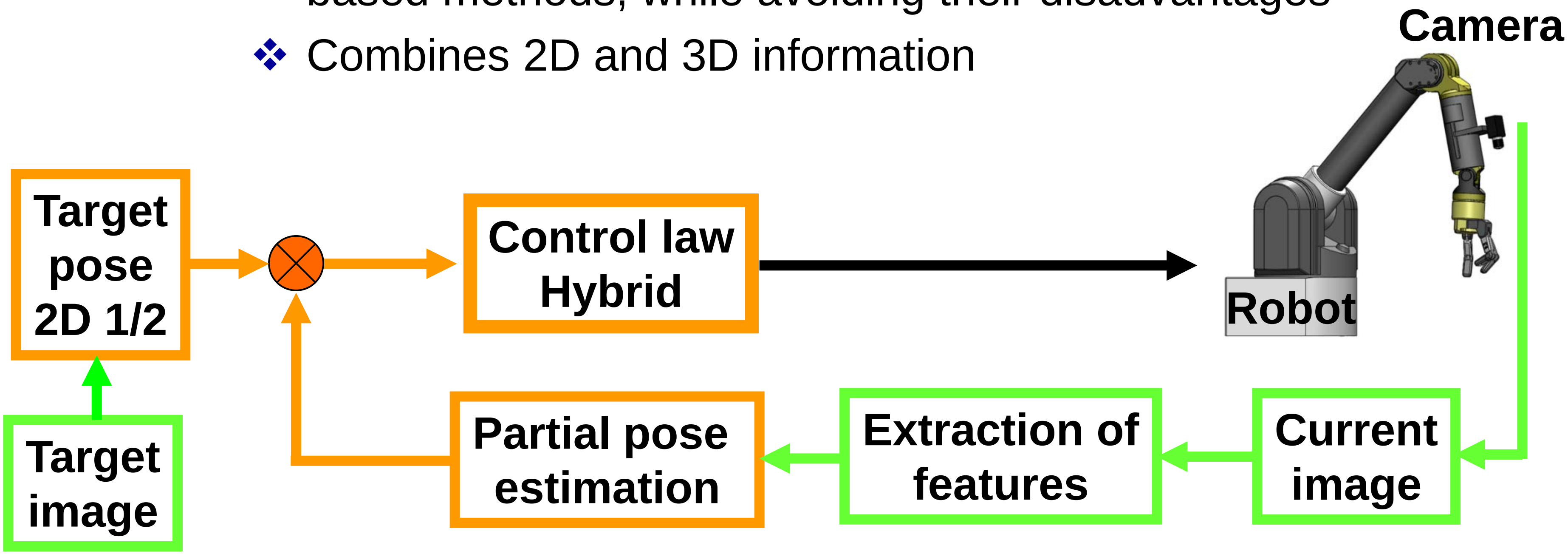
Classification of visual servoing systems

- According to the definition of the error signal
  2. Image-based visual servoing
     - The target is defined in image coordinates (pixel).
     - Task is defined in the image
       - Can reduce computational delay, eliminating the need for image interpretation
       - Non-linear and coupled system
       - More robust to poor calibration



Classification of visual servoing systems

- According to the definition of the error signal
  3. Hybrid visual servoing
     - Partially position-based and image-based
       - Partial decoupling in the interaction matrix
       - Presents the advantages of the position-based and image-based methods, while avoiding their disadvantages
       - Combines 2D and 3D information



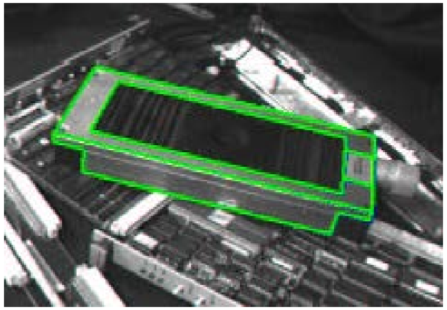
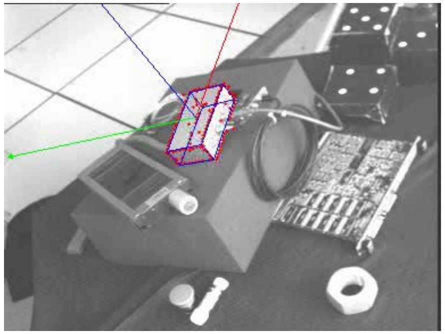
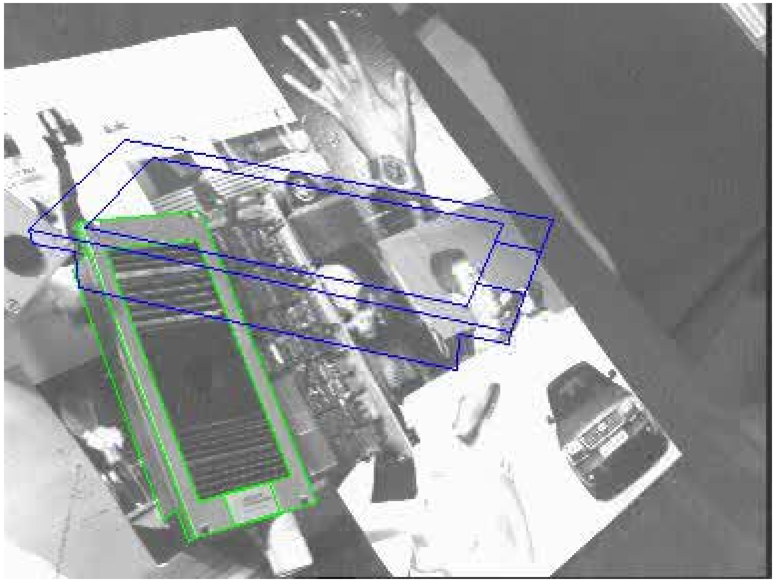
Index

Position-Based Visual Servoing

- Position-based visual servoing
  - Vision data is used to build a 3D representation of the environment.
    - The 3D position of the manipulator relative to its desired position is estimated from the acquired images.
      - Based on 3D model
      - Two-view geometry
      - Stereo view
    - Control law that eliminates the error in 3D space.
  - Related bibliography :
    - [Taylor et al. 1985], [Zhang et al. 1990], [Carlsson 1991], [Tonko et al. 1997], [Basri et al. 1998], [Martinet & Gallice 1999], [Taylor & Ostrowski 2000], [Liang & Pears 2002], [Benhimane et al. 2005].

Position-Based Visual Servoing: Model

- 3D localization estimation: Model based
  - The objective of the control is to move the robot camera until the 3D model projection of the object corresponds to the observed image.

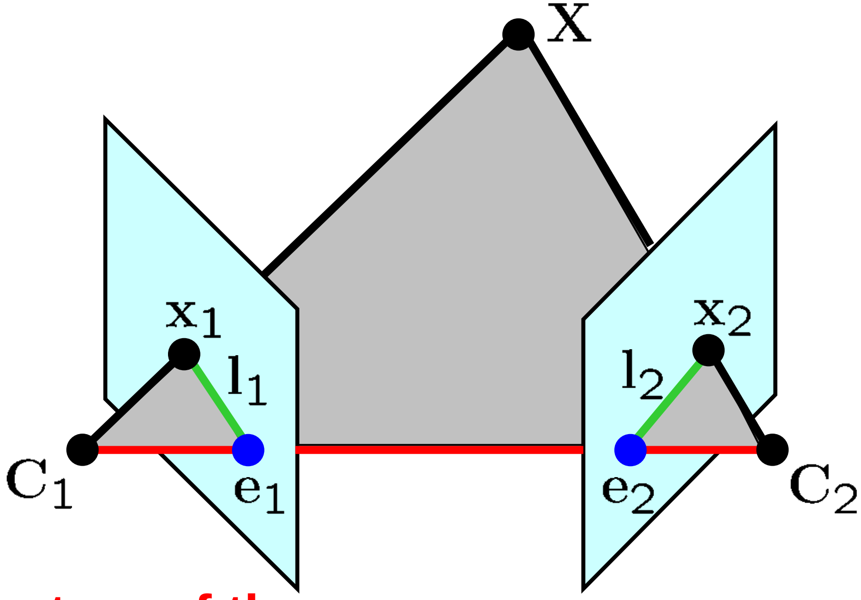




  - It is required to know
    - 3D model of the object (CAD)
    - Intrinsic camera calibration
    - Matching of features in the 3D model with features of its projection on the image
  - Dementhon's algorithm for estimating object location

- 3D pose estimation: Epipolar Geometry
  - Intrinsic geometry between two views. It depends only on the internal parameters of the cameras and their relative position.
  - The fundamental matrix encapsulates this intrinsic geometry.

$$\mathbf{l}_2 = \mathbf{F}\,\mathbf{x}_1$$

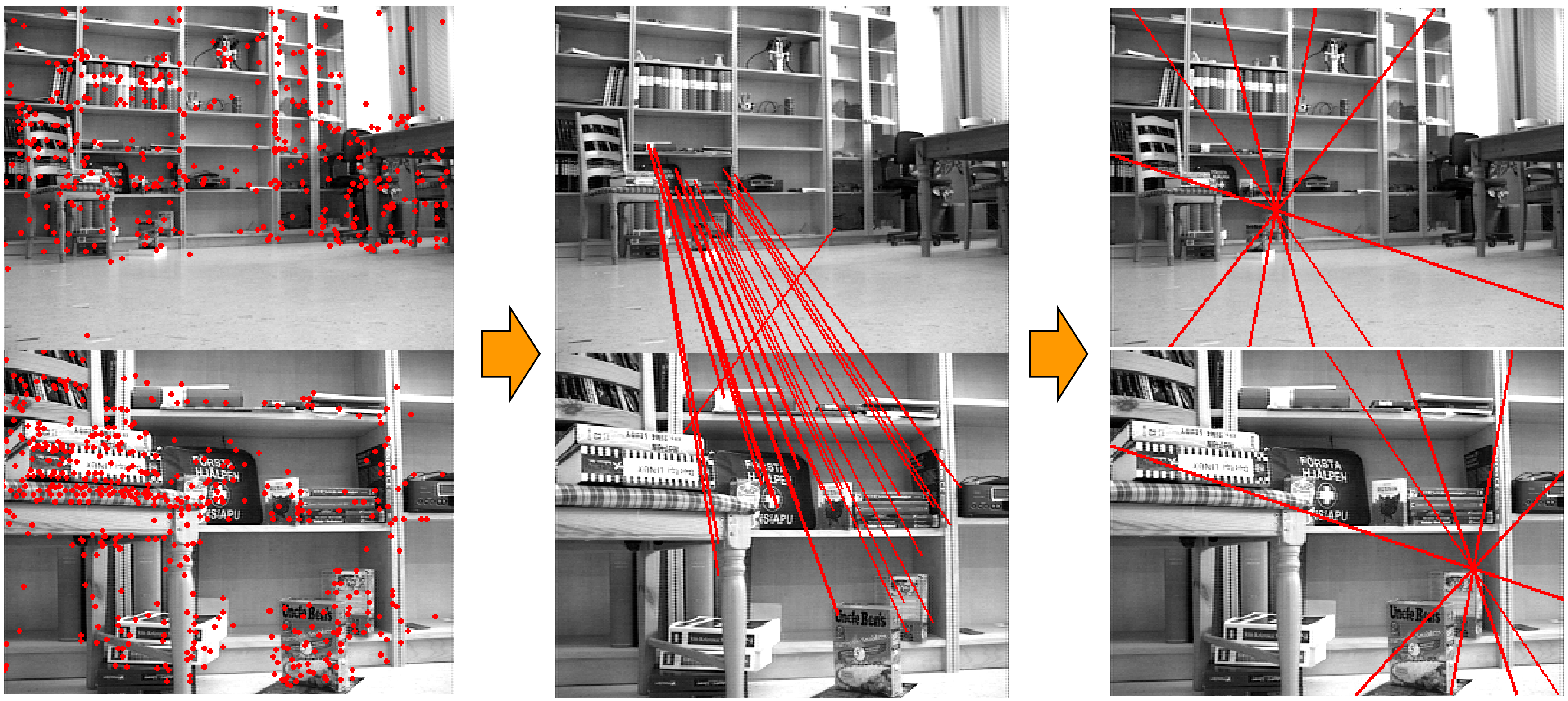
$$\mathbf{x}_2^T \mathbf{F}\,\mathbf{x}_1 = 0$$



- **Baseline: Line joining the centers of the cameras.**
- **Epipole: Intersection between the baseline and the image.**
- **Epipolar plane: Plane containing the baseline.**
- **Epipolar line: Intersection of an epipolar plane with the image.**

Position-Based Visual Servoing: Epipolar geometry

- Epipolar geometry estimation
  - Automatic image feature extraction and matching (SIFT, Harris points,...)
  - Robust fundamental matrix estimation (RANSAC, MAPSAC,...)



Feature extraction — Point matching — Epipolar geometry estimation

Position-Based Visual Servoing: Epipolar geometry

- Essential Matrix = Calibrated Fundamental Matrix

$$\mathbf{P}_1 = [\mathbf{I};\mathbf{0}]; \quad \mathbf{P}_2 = [\mathbf{R}_{21};\mathbf{t}_{21}] \Rightarrow \quad \mathbf{x}_2{}^T \mathbf{E}_{21}\mathbf{x}_1 = 0$$

$$\mathbf{E}_{21} = [\mathbf{t}_{21}]_\times \mathbf{R}_{21}$$

$$\text{since} \quad \mathbf{F}_{21} = \mathbf{K}_2{}^{-T}[\mathbf{t}_{21}]_\times \mathbf{R}_{21}\mathbf{K}_1{}^{-1} \Rightarrow \mathbf{E}_{21} = \mathbf{K}_2{}^T \mathbf{F}_{21}\mathbf{K}_1$$

$$\mathbf{p_2}^\mathbf{T}\mathbf{F_{21}p_1} = 0 \quad \text{with} \quad \mathbf{p_1} = \mathbf{K}_1\mathbf{x}_1; \;\; \mathbf{p_2} = \mathbf{K}_2\mathbf{x}_2$$

- Decomposition of the essential matrix to obtain rotation and translation (up to scale factor).
- Algorithm:

$$\mathbf{E} = \mathbf{U}\cdot\mathbf{S}\cdot\mathbf{V}^T$$

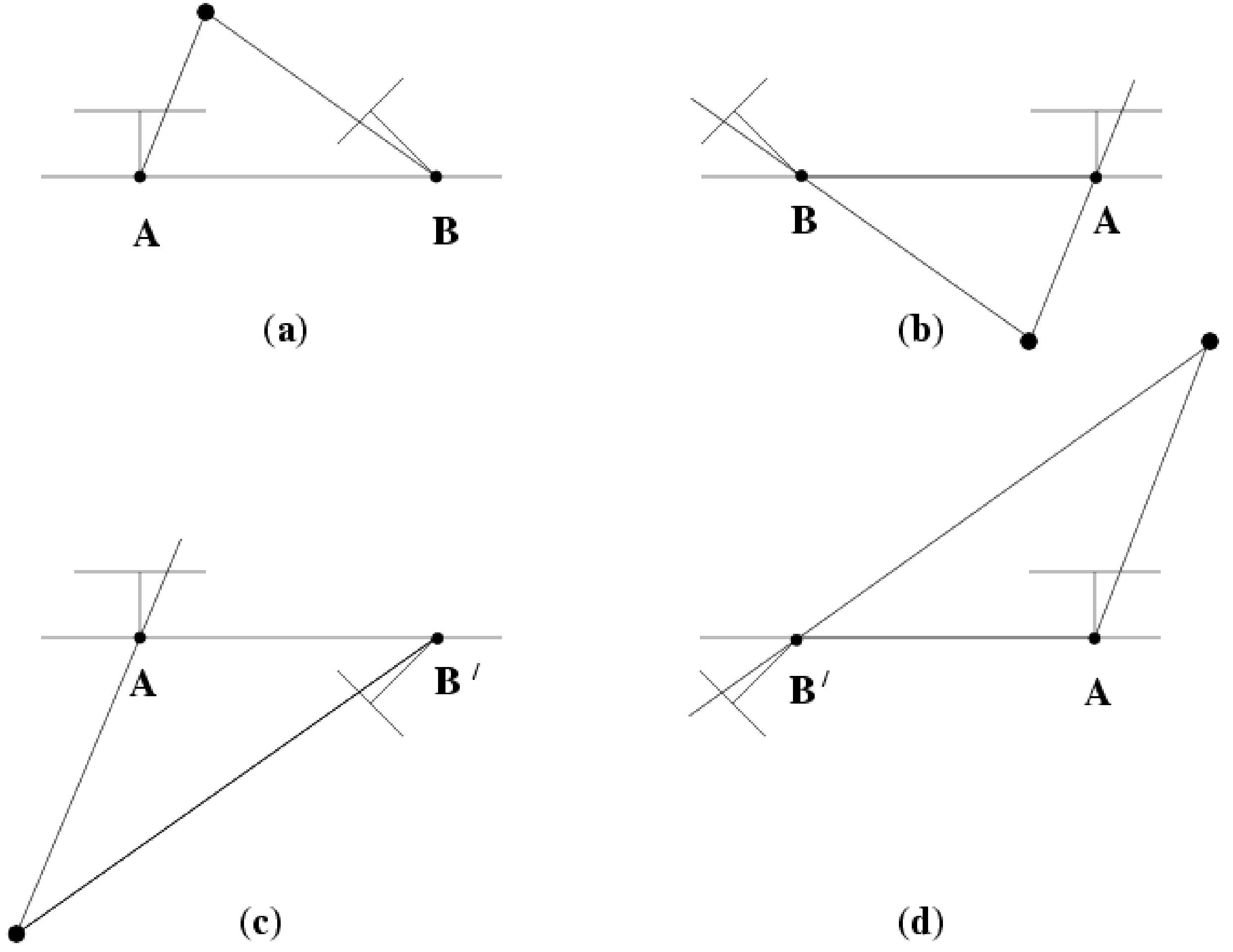
$$\mathbf{t} = \mathbf{U}\cdot(0,0,1)^T$$

$$\mathbf{R}_1 = \mathbf{U}\cdot\mathbf{W}\cdot\mathbf{V}^T \qquad \text{With } \mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_2 = \mathbf{U}\cdot\mathbf{W}^T\cdot\mathbf{V}^T$$

Position-Based Visual Servoing: Epipolar geometry

- The essential matrix decomposition yields four theoretical solutions: $(\mathbf{t},\mathbf{R}_1),(\mathbf{t},\mathbf{R}_2),(-\mathbf{t},\mathbf{R}_1),(-\mathbf{t},\mathbf{R}_2)$



Position-Based Visual Servoing: Homography

- Homography:
  - Two images can be geometrically related by a homography induced by a plane of the scene:

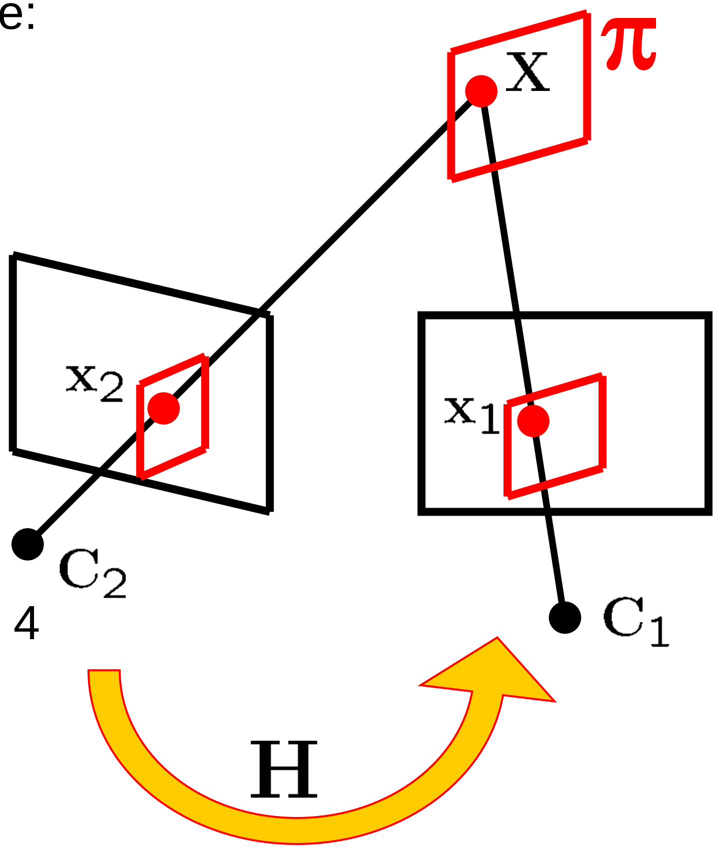
$$\mathbf{H} \in \Re^{3\times 3}$$

  - Corresponding points related by a homography:

$$\mathbf{x}_2 = \mathbf{H}\mathbf{x}_1$$

  - Homography can be computed from at least 4 corresponding points of the plane in the two images.

$$\mathbf{H} = \mathbf{K}_2\left(\mathbf{R} - \mathbf{t}\,\mathbf{n}^T / d\right)\mathbf{K}_1^{-1}$$



Position-Based Visual Servoing: Homography

- Homography decomposition:

$$H_c = K^{-1} \cdot H \cdot K$$

$$[U, S, V] := \mathrm{svd}(H_c)$$

$$Hc = Hc / S_{22}$$

$$V := [v_1, v_2, v_3]$$

$$(S)^2 := diag[\lambda_1, \lambda_2, \lambda_3]$$

$$\alpha = \sqrt{\frac{\lambda_3 - \lambda_2}{\lambda_3 - \lambda_1}}, \beta = \sqrt{\frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_1}}$$

$$\omega_1 = \alpha \cdot v_1 + \beta \cdot v_3$$

$$\omega_2 = \alpha \cdot v_1 - \beta \cdot v_3$$

$$U_1 = [\omega_1, v_2, (\omega_1 \times v_2)]$$

$$U_2 = [\omega_2, v_2, (\omega_2 \times v_2)]$$

$$W_1 = [H_c \cdot \omega_1, \quad H_c \cdot v_2, \quad (H_c \cdot \omega_1) \times (H_c \cdot v_2)]$$

$$W_2 = [H_c \cdot \omega_2, \quad H_c \cdot v_2, \quad (H_c \cdot \omega_2) \times (H_c \cdot v_2)]$$

$$\text{Solutions } i \ (i = 1,2):$$

$$R_i = W_i \cdot U_i^T$$

$$n_i = \omega_i \times v_2$$

$$t_i = (I_{3x3} - R_i^T \cdot H_c) \cdot n_i \cdot d$$

d: Distance to the plane

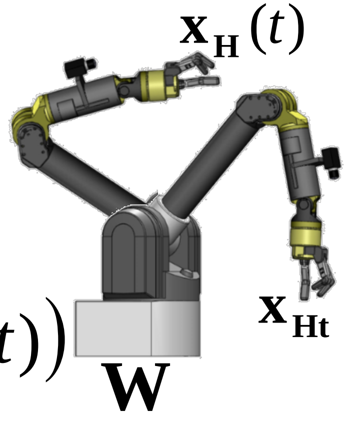
Control law

$${}^{\mathbf{W}}\mathbf{T}_{\mathbf{H}} \in \Re^{4\times 4}$$

$$\mathbf{x}_{\mathbf{H}} \in \Re^{6}$$

- Given the location of the terminal element: ${}^{\mathbf{W}}\mathbf{T}_{\mathbf{H}}(\mathbf{x}_{\mathbf{H}})$
- The goal is to reach location: ${}^{\mathbf{W}}\mathbf{T}_{\mathbf{Ht}}(\mathbf{x}_{\mathbf{Ht}})$
- The task of matching the terminal element to a desired location can be defined as.

$$\varepsilon\left(\mathbf{x}_{\mathbf{H}}(t)\right) = \mathbf{x}_{\mathbf{Ht}} - \mathbf{x}_{\mathbf{H}}(t)$$



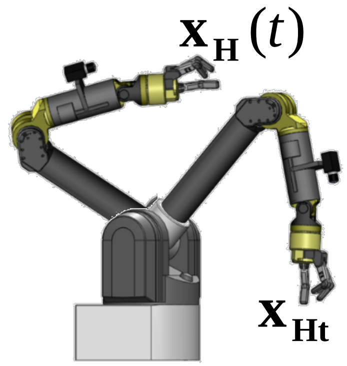
- The simplest control law that assures zero error with exponential decay is

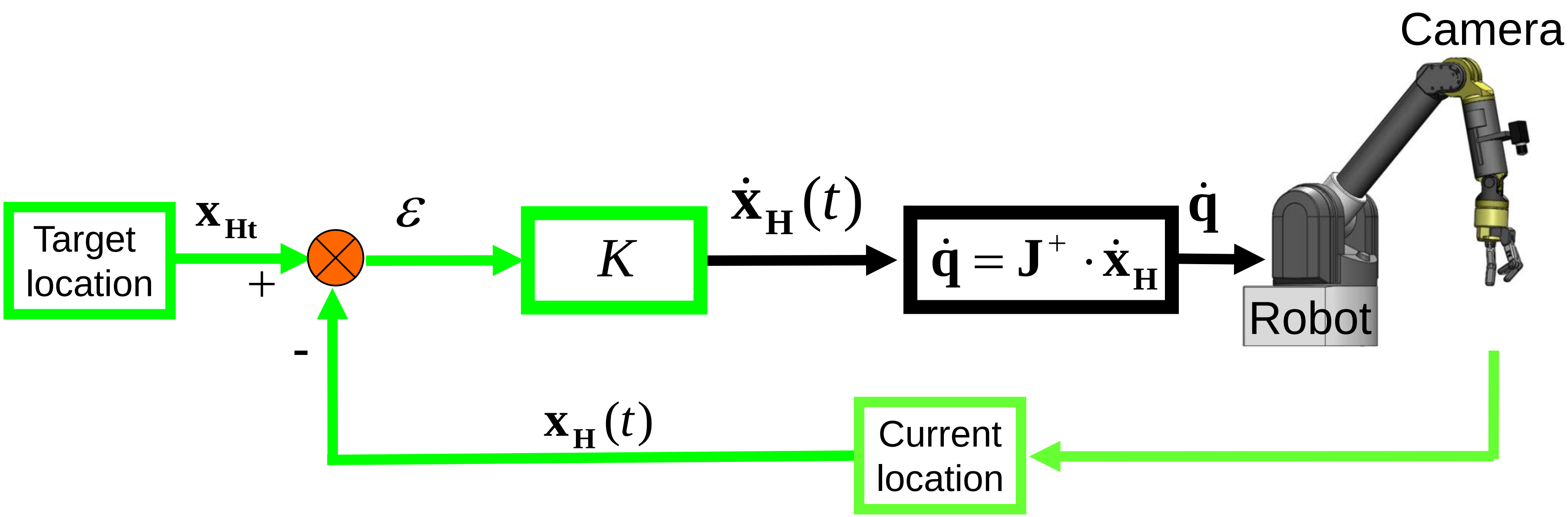
$$\dot{\mathbf{x}}_{\mathbf{H}}(t) = K \cdot \varepsilon\left(\mathbf{x}_{\mathbf{H}}(t)\right) = K \cdot \left(\mathbf{x}_{\mathbf{Ht}} - \mathbf{x}_{\mathbf{H}}(t)\right)$$

  - Controller gain ***K***
  - The output of the controller is the desired speed of the terminal element to reach the target.
  - The inverse Jacobian allows to calculate the joint velocities required to perform the task:

$$\dot{\mathbf{q}} = \mathbf{J}^{+}(\mathbf{q}) \cdot \dot{\mathbf{x}}_{\mathbf{H}}(\mathbf{t})$$

- Control law


$\mathbf{x}_{\mathbf{H}}(t)$
$\mathbf{x}_{\mathbf{Ht}}$


Camera
Target location
$\mathbf{x}_{\mathbf{Ht}}$
+
-
ε
K
$\dot{\mathbf{x}}_{\mathbf{H}}(t)$
$\dot{\mathbf{q}} = \mathbf{J}^{+} \cdot \dot{\mathbf{x}}_{\mathbf{H}}$
$\dot{\mathbf{q}}$
Robot
$\mathbf{x}_{\mathbf{H}}(t)$
Current location

Position-Based Visual Servoing: Control law

- Control law for tracking $\mathbf{x}_{\mathbf{Ht}}(t)$
  - If the desired configuration varies over time the error will not converge to zero.
  - Use proportional and integral control (PI).
    - Estimate the error variation at instant k: $\mathbf{I}_{\mathrm{k}}$

$$\mathbf{I}_{\mathrm{k+1}} = \mathbf{I}_{\mathrm{k}} + \mu \cdot \mathbf{e}_{\mathrm{k}} = \mu \cdot \sum_{j=0}^{k} \mathbf{e}_{\mathrm{j}} \qquad \text{with} \qquad \mathbf{I}_{0} = 0$$

    - Efficient for tracking a target at constant speed.

$$\mathbf{I}_{\mathrm{k+1}} = \mathbf{I}_{\mathrm{k}} \qquad \text{if} \qquad \mathbf{e}_{\mathrm{k}} = 0$$

Position-Based Visual Servoing

- Conclusions
  - It needs a priori information (calibration). Therefore unsuitable for unstructured environments.
  - Can be unstable if the initial position of the camera is far away from the desired position.
  - Directly controls the camera trajectory in Cartesian space. Straight trajectory in 3D space.
  - Decoupled control and no singular configurations.
  - No control in the image. Points can leave the field of view:

Index

- Introduction
- Robotic manipulation
- Vision systems
- Classification of visual control systems
- Position-Based Visual Servoing

- Image-Based Visual Servoing
- Stability analysis
- Hybrid Visual Servoing
- Conclusion and bibliography