EE211: Robotic Perception and Intelligence

Lecture 8 3D Vision for Navigation and Grasping

Jiankun WANG

Department of Electronic and Electrical Engineering
Southern University of Science and Technology

Undergraduate Course, Nov 2025

Outline

1. Configuration of the Visual System
2. Image Processing

Outline

Number of Cameras

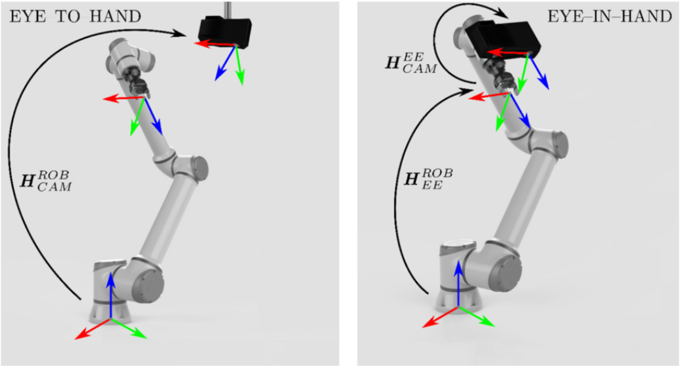
- One, or two or more cameras.
- 3D/Stereo vision: More cameras can retrieve information about depth by evaluating the object's distance with respect to the visual system.
- Humans perceive objects in 3 dimensions $\leftarrow$ the brain receives the same images from two eyes, observing the same scene from slightly different angles.
- One camera can achieve 3D vision: two images of the same object, taken from two different poses.
- With a single image, the depth can be estimated on the basis of certain geometrical characteristics of the object known in advance.
- Mono-camera systems or multi-camera systems: cheaper and easier to calibrate, but lower accuracy.

Robotics: Modeling, Planning and Control.

Placement of Cameras

- For mono-camera systems: fixed or mobile configuration.
- Eye-to-hand (fixed): camera is mounted in a fixed location.
- The camera field of view does not change during the execution of the task $\rightarrow$ the accuracy of such measurements is, in principle, constant.
- Not suitable to high dynamic scenarios, such as assembly.
- Eye-in-hand (mobile): camera is attached to the robot.
- Can be mounted both before and after the wrist.
- When the end-effector is close to the object, the accuracy becomes higher, and occlusions are virtually absent.

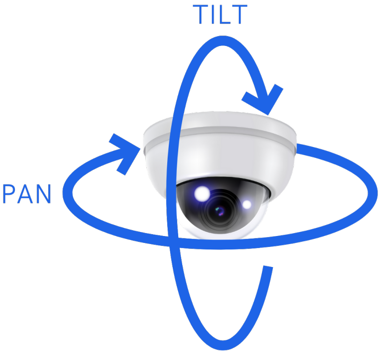
Illustrations



https://blog.zivid.com/importance-of-3d-hand-eye-calibration

Other Examples

- Hybrid configuration system.
- A separate category: robotic heads equipped with a stereo vision system mounted on motorized mechanisms that allow for yaw motion, or pan, and pitch motion, or tilt, hence the name of pan-tilt cameras.

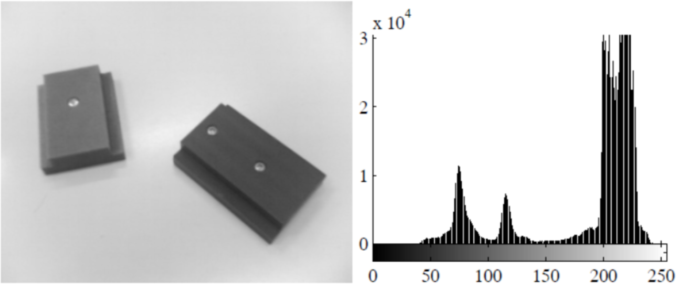


Outline

Image Processing in Robotics

- Very rich and varied and thus requires complex and computational expensive transformations before it can be used for controlling a robotic system.
- Extraction of numerical information from the image, to provide a synthetic and robust description of the objects of interest in the scene, through the so-called image feature parameters.
- Image segmentation: obtain a representation suitable for the identification of measurable features of the image.
- Image interpretation: measurement of the feature parameters of the image.

Example: Gray Levels



- Black-and-white image and corresponding gray-level histogram.
- Gray levels are quantized from 0 to 255.
- Three main peaks — from left to right — corresponding to the darkest object, the lightest object, and the background.

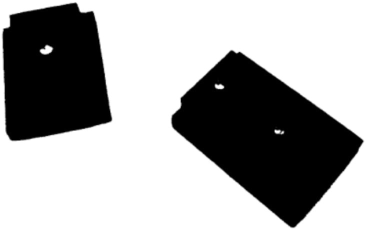
Image Segmentation

- Grouping process: the image is divided into a certain number of groups, referred to as segments.
- Region-based segmentation: grouping sets of pixels sharing common features into two-dimensional connected areas.
- Boundary-based segmentation: identifying the pixels corresponding to object contours and isolating them from the rest of the image.
- Complementary: a boundary can be achieved by isolating the contours of a region and, conversely, a region can be achieved simply by considering the set of pixels contained within a closed boundary.

Region-based Segmentation

- Idea: obtaining connected regions by continuous merging of initially small groups of adjacent pixels into larger ones.
- Binary segmentation or image binarization, separating one or more objects present in the image from the background by comparing the gray level of each pixel with a threshold I.
- Light objects against a dark background, dark objects against a light background.
- Binary image: pixels of objects, set light intensity to 0; pixels of background, set to 1.
- The choice of the threshold.

Region-based Segmentation



- Gray-level histogram: contains clearly distinguishable minimum and maximum values, corresponding to the gray levels of the objects and of the background.
- Dark objects against a light background: the closest minimum to the left, where the threshold $I = 152$.

Boundary-based Segmentation

- Grouping many single local edges, which are sets of pixels where the light intensity changes abruptly: (1) identify local edges; (2) construct short-curve segments by edge linking; (3) obtain the boundaries by joining these curve segments.
- Edge detection is essentially a filtering process and can often be implemented via hardware.
- Boundary detection is a higher level task usually requiring more sophisticated software.
- Using the most effective edge detector to simplify the boundary detection process.

Gradient-based Edge Detection

- Intensity of a pixel: $I(X_I, Y_I)$.
- Gradient of $I(X_I, Y_I)$: measure the rate of change of the gray level, will have large magnitude close to these transitional boundary areas.
- Idea: grouping the pixels where the magnitude of the gradient is greater than a threshold.
- Approximate the derivative along directions X_I and Y_I with the first differences:

$$\Delta_1 = I(X_I + 1, Y_I) - I(X_I, Y_I)$$
$$\Delta_2 = I(X_I, Y_I + 1) - I(X_I, Y_I)$$

Gradient-based Edge Detection

- Roberts operator: the first differences computed along the diagonals of a (2×2) square of pixels.

$$
\begin{aligned}
\Delta_1 &= I(X_I + 1, Y_I + 1) - I(X_I, Y_I) \\
\Delta_2 &= I(X_I, Y_I + 1) - I(X_I + 1, Y_I)
\end{aligned}
$$

- Sobel operator: defined on a (3×3) square of pixels.

$$
\begin{aligned}
\Delta_1 &= I(X_I + 1, Y_I - 1) + 2I(X_I + 1, Y_I) + I(X_I + 1, Y_I + 1) \\
&= -I(X_I - 1, Y_I - 1) - 2I(X_I - 1, Y_I) - I(X_I - 1, Y_I + 1) \\
\Delta_2 &= I(X_I - 1, Y_I + 1) + 2I(X_I, Y_I + 1) + I(X_I + 1, Y_I + 1) \\
&= -I(X_I - 1, Y_I - 1) - 2I(X_I, Y_I - 1) - I(X_I + 1, Y_I - 1)
\end{aligned}
$$

Gradient-based Edge Detection

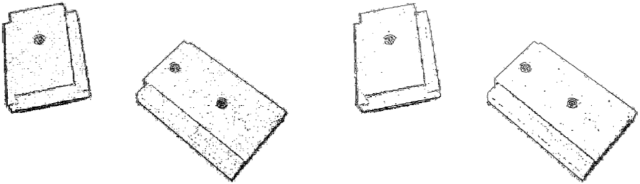
- Approximated magnitude, or norm, of gradient $G(X_I, Y_I)$ can be evaluated using one of the following two expressions:

$$G(X_I, Y_I) = \sqrt{\Delta_1^2 + \Delta_2^2}$$

$$G(X_I, Y_I) = |\Delta_1| + |\Delta_2|$$

- Direction $\theta(X_I, Y_I)$ with the relationship

$$\theta(X_I, Y_I) = \arctan(\Delta_2, \Delta_1)$$



- Contours of image obtained using Roberts (left) and Sobel (right) operators, and binarization.
- thresholds: $l = 0.02$, and $l = 0.0146$.

Image Interpretation

- The process of calculating the image feature parameters from the segments.
- Moments: widely used in visual servoing.
- Definition of moment $m_{i,j}$ of a region $\mathcal{R}$ of an image:

$$m_{i,j} = \sum_{X_I, Y_I \in \mathcal{R}}^{max} I(X_I, Y_I) X_I^i Y_I^j$$

- For binary images, by assuming the light intensity equal to one for all the points of region $\mathcal{R}$, the others are zero, then

$$m_{i,j} = \sum_{X_I, Y_I \in \mathcal{R}}^{max} X_I^i Y_I^j$$

- Moments $m_{0,0}$ coincides with the area of the region, computed in terms of the total number of pixels of region $\mathcal{R}$.

$$m_{i,j} = \sum_{X_I, Y_I \in \mathcal{R}}^{max} X_I^i Y_I^j$$

- Coordinates of the so-called centroid of the region.

$$\overline{x} = \frac{m_{1,0}}{m_{0,0}} \quad \overline{y} = \frac{m_{0,1}}{m_{0,0}}$$

- Central moments:

$$\mu_{i,j} = \sum_{X_I, Y_I \in \mathcal{R}} (X_I - \overline{x})^i (Y_I - \overline{y})^j$$

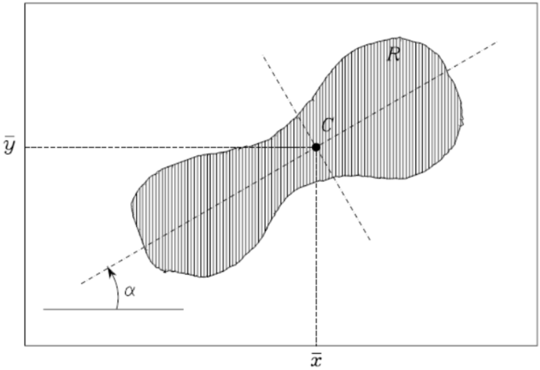
Image Interpretation*

- The central moments of second order $\mu_{2,0}$ and $\mu_{0,2}$ have the meaning of inertia moments with respect to axes X_I and Y_I respectively, while $\mu_{1,1}$ is an inertia product.
- Moment $m_{0,0}$ corresponds to the mass of the body and the centroid corresponds to the centre of mass.

$$\mathcal{I} = \begin{bmatrix} \mu_{2,0} & \mu_{1,1} \\ \mu_{1,1} & \mu_{0,2} \end{bmatrix}$$

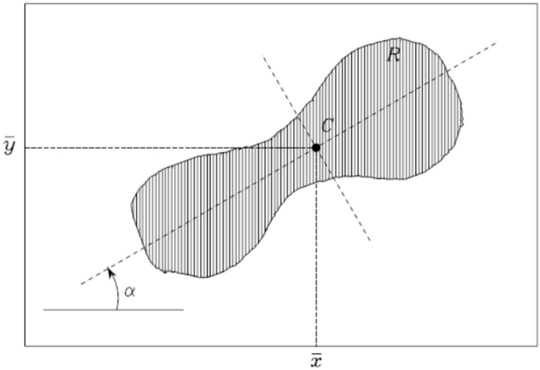
- has the meaning of inertia tensor relative to the centre of mass.

Image Interpretation*



- The eigenvalues of matrix $\mathcal{I}$ define the principal moments of inertia, termed principal moments of the region and the corresponding eigenvectors define the principal axes of inertia, termed principal axes of the region.

Image Interpretation*



- $\mathcal{R}$ is asymmetric $\rightarrow$ principal moments are different $\rightarrow$ characterize the orientation of $\mathcal{R}$:

$$\alpha = \frac{1}{2}\tan^{-1}\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}$$

Q & A

1. Configuration of the Visual System

2. Image Processing