Lecture 1: Introduction to Robotic Vision

Wei Qi Yan

Auckland University of Technology

Table of Contents

Course Descriptor

Learning Outcomes

On successful completion of this course, students will be able to:

- Express the concepts and methods of image formation related to robotic vision, including camera properties and calibration.
- Demonstrate critical thinking in designing problem solutions for tasks in robotics, i.e., robotic vision, visual control and deep learning.
- Derive solutions for particular robotic vision and visual control tasks characterised by specifics of image data and deep learning algorithms.
- Critically evaluate the performance of robotic vision with deep learning algorithms, bench mark data, performance measures, and ways to define ground truth.
- Examine opportunities of using robotic vision as a part of complex robotic systems and applications.

Course Descriptor

Content

- Robotics
- Camera calibration
- Digital image processing
- Stereo vision
- 3D object reconstruction
- Deep learning
- Vision-based robotic control
- Supercomputing

Course Structure and Assessment

Lectures

- Lecture 1: Introduction to robotic vision
- Lecture 2: Mobile robots
- Lecture 3: Arm-type robots
- Lecture 4: Digital image processing
- Lecture 5: Stero vision and 3D object reconstruction
- Lecture 6: Guest lecture
- ......

Course Structure and Assessment

Lectures

- ......
- Lecture 7: Deep learning for robotic vision
- Lecture 8: Deep learning for robotic control
- Lecture 9: Vision-based robotic control
- Lecture 10: Supercomputing for robotics
- Lecture 11: Individual proposal online Q&A
- Lecture 12: Individual report online Q&A

Course Structure and Assessment

Course Structure

- Learning goals and objectives
- Learning assessments
- Learning environments
- Prescribed reading materials
- Availability of resources
- Relevance of course materials
- Effectiveness of course materials
- Individual feedback
- Communication systems
- Self-directed and teacher-directed learning
- Flipped classroom

Course Structure and Assessment

Assessments

- Individual proposal: 20% (attendance, 9 pages proposal, at least 9 references)
- Individual Report: 30% (attendance, 6 pages IEEE paper, at least 6 references)
- Online Exam: 50% (Quiz questions)
- To pass this course, the students need at least a **C-**
- Project topics will be available prior to the course start.
- Assessment descriptions and criteria will be provided for each assessment.

Course Structure and Assessment

Grade Map

- **A+**: Pass with distinctions (89.50∼100.00)
- **A**: Pass with distinctions (84.50∼89.49)
- **A-**: Pass with distinctions (79.50∼84.49)
- **B+**: Pass with merit (74.50∼79.49)
- **B**: Pass with merit (69.50∼74.49)
- **B-**: Pass with merit (64.50∼69.49)
- **C+**: Pass (59.50∼64.49)
- **C**: Pass (54.50∼59.49)
- **C-**: Pass (49.50∼54.49)
- **D**: Specified fail

Course Structure and Assessment

Written Proposals

- Organisation (i.e., topic, name/ID, page number, abstract, introduction, related work/literature review, methodology, timetable, 9 pages, 9+ references)
- Content (i.e., datasets, maths background, methods/ approaches, flowcharts/algorithm, demos, evaluations, comparisons, result analysis, writing grammar & typos, document format, etc.).
- Correctly referencing using APA format. (i.e., 9+ references, APA format, reference page numbers, APA citings, recent 5 years publications, etc.)

Course Structure and Assessment

Written Reports

- Organisation (i.e., topic, name/ID, page number, abstract, introduction, related work/literature review, methodology, experiments, conclusion, future work, 6 IEEE pages, 6+ references)
- Content (i.e., datasets, maths background, methods/ approaches, flowcharts/algorithm, demos, evaluations, comparisons, result analysis, writing (grammar & typos), document format, etc.).
- Correctly referencing using IEEE format. (i.e., 6+ references, using IEEE format, reference page numbers, IEEE citings, recent 5 years publications, etc.)

Course Structure and Assessment

Course Requirements and Regulations

- Absence
- Submitting assignments
- Extensions
- Late assignments
- Plagiarism/cheating
- Turnitin software
- ChatGPT and AI software
- Returning marks
- Reconsideration of assessment

Questions?

Course Structure and Assessment

Questions?

How to pass this course?

1. Individual proposal, individual report, online quiz
2. Group proposal, individual report, online quiz
3. Individual proposal, group report, online quiz
4. None of the given options

The right answer is:___

Course Structure and Assessment

Questions?

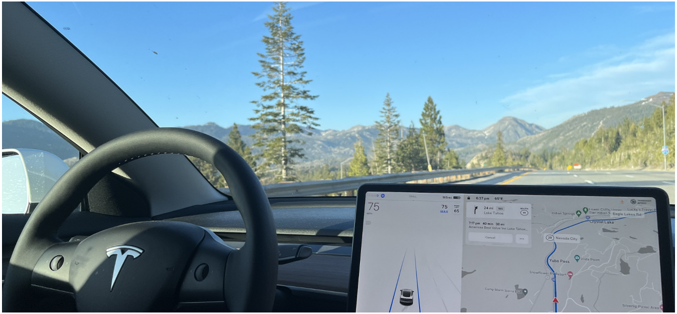
Robotics

- Robotics: Robotic vision and robotic control
- Robotic Vision: Camera calibration, image formation, image processing, stereo vision, 3D reconstruction, etc.
- Robotic Control: Machine intelligence, GA algorithm, reinforcement learning, visual servoing, etc.
- Robots: Wheeled robots, humandriod / andriod robots, mobile robots, arm-type robots, flying robots, tracked robots, etc.

Introduction to Robotic Vision

Wheeled Robot

Tesla Autopilot is an advanced driver-assistance system (ADAS) developed by Tesla.



https://bostondynamics.com/blog/electric-new-era-for-atlas/

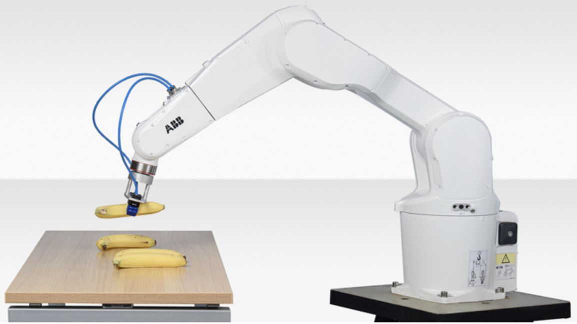
Humaniold Robot

An android (male) or gynoid (female) is a humanoid robot.

https://bostondynamics.com/blog/electric-new-era-for-atlas/

Arm-Type Robot

Cartesian/Gantry robot's arm has three prismatic joints, whose axes are coincident with a Cartesian coordinator.



https://new.abb.com/products/robotics/robots/articulated-robots/irb-1200-hygienic

Terminology of Robots

- *Number of Axes*: Roll, pitch, and yaw control are required for full control of the end manipulator.
- *Degrees of Freedom* (DOF): Number of points a robot can be directionally controlled around.
- *Working Envelope*: Region of space a robot can encompass.
- *Kinematics*: Arrangement and types of joints.
- *Payload*: Amount that can be lifted and carried.
- *Speed*: Angular or linear movement velocity
- *Acceleration*: Limits maximum speed over short distances.
- *Power Source*: Electric motors or hydraulics are typically used.

https://new.abb.com/products/robotics/robots/articulated-robots/irb-1200-hygienic

Introduction to Robotic Vision

Subjects of Robotic Vision

- Robotic mapping
- Robotic navigation
- Human-robot interaction
- Imitation learning
- Pose recognition
- Gesture recognition
- Emotion recognition
- ......

https://new.abb.com/products/robotics/robots/articulated-robots/irb-1200-hygienic

Questions?

Questions?

What's the relationship between DOF and the number of Axes in robotics?

1. The number of axes is greater than DOF
2. The number of axes is less than DOF
3. The number of axes is euqal to DOF
4. None of the given options

The right answer is:___

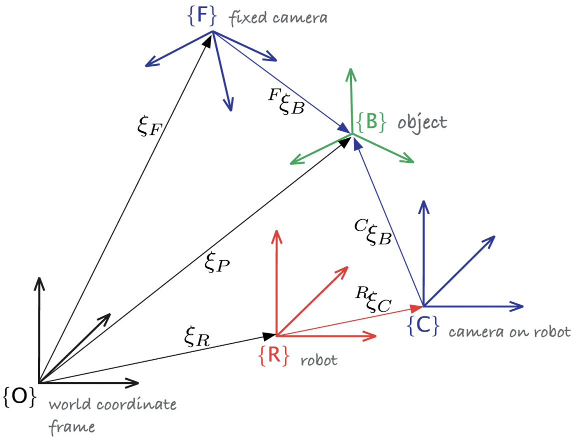
Questions?

Representing Position and Orientation

- A fundamental requirement in robotics and computer vision is to represent the position and orientation of objects in an environment.
- We describe the position and orientation of the object by using its coordinate frame.
- A coordinate frame, or Cartesian coordinate system, is a set of orthogonal axes which intersect at a point known as the origin.
- The position and orientation of a coordinate frame is known as its pose and is shown graphically as a set of coordinate axes.
- An important characteristic of relative poses is that they can be composed or compounded.

Representing Position and Orientation

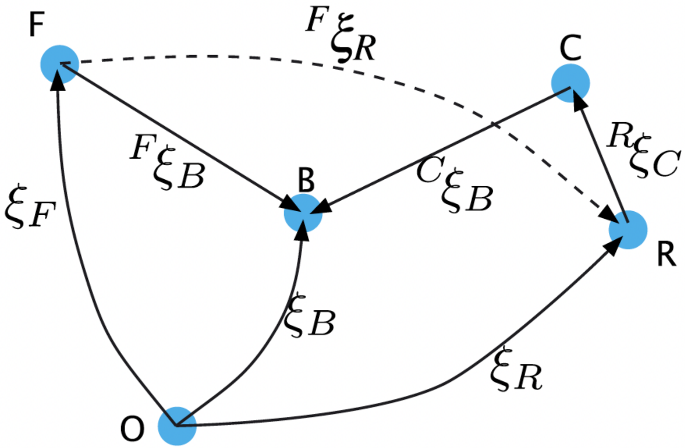
An important characteristic of relative poses is that they can be composed or compounded.



Multiple 3D coordinate frames and relative poses

Representing Position and Orientation

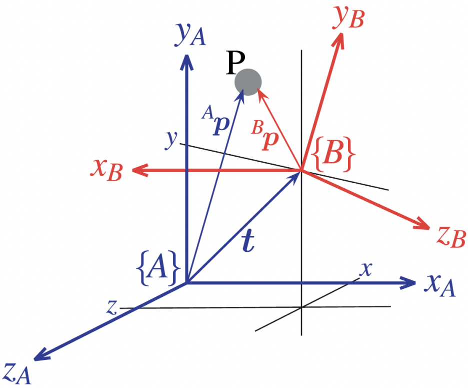
An alternative representation of the spatial relationships is a directed graph.



Multiple 3D coordinate frames and relative poses

Representing Position and Orientation

Combining translation and orientation



Two 3D coordinate frames {A} and {B}. {B} is rotated and translated with respect to {A}.

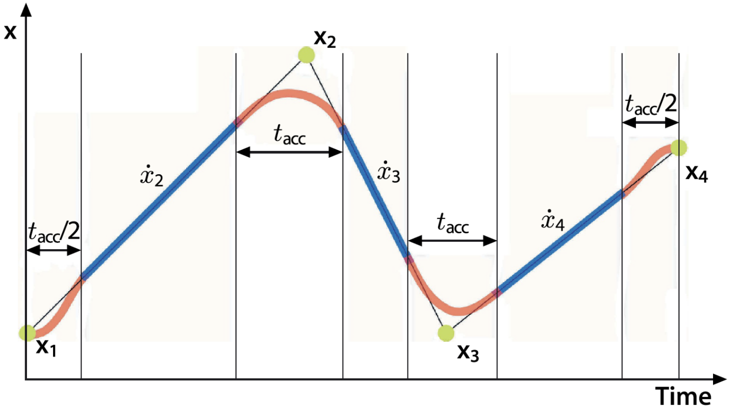
Time and Motion

- The object pose is varying as a function of time.
- Trajectory, the temporal sequence of poses, smoothly changes from an initial pose to a final pose.
- The rate of change of pose, its temporal derivative relates to mechanics such as linear velocity and angular velocity.
- Given measurements from linear velocity and angular velocity sensors, we estimate the pose for a moving object.
- As the velocity of the linear segment increases, its duration decreases and ultimately its duration would be zero.
- In fact, too high or too low a value for the maximum velocity will result in an infeasible trajectory.

Trajectory

- A path is a spatial construct – a locus in space that leads from an initial pose to a final pose.
- A trajectory is a path with specified timing.
- An important characteristic of a trajectory is that is smooth – position and orientation vary smoothly with time.
- The trajectory has defined boundary conditions for position, velocity and acceleration.
- Smoothness means that its first few temporal derivatives are continuous.
- Polynomials are simple to compute and can easily provide the required smoothness and boundary conditions.
- There is often a need to move smoothly along a path through one or more intermediate or via points without stopping.

Trajectory



The multisegment trajectory: The motion comprises linear motion segments with polynomial blends.

Questions?

Questions?

In robotics, polynomials are simple to compute and can easily provide the required:

1. smoothness and boundary conditions.
2. smoothness only.
3. boundary conditions only.
4. None of the given options

The right answer is: ___

Questions?

The Chronicle

- 2024: OpenAI Sora, YOLOv9, YOLOv10
- 2023: YOLOv8, Diffusion Transformer(DiT), DALL · E
- 2022: ChatGPT, YOLOv7 & YOLOv6
- 2021: Vision Transformer (ViT)
- 2020: YOLOv4 & YOLOv5, GPT-3
- 2019: 2018 ACM Turing Award
- 2018: YOLOv3 & Mask R-CNN
- 2017: CapsNets & YOLO9000
- 2016: You Only Look Once (YOLO)
- 2015: ResNet, GoogLeNet, SSD, & Fast / Faster R-CNN
- 2014: GANs & VGG
- 2013: Region-Based CNN(R-CNN)
- 2012: AlexNet (ImageNet)
- 1997: Long Short-Term Memory (LSTM)

The Chronicle

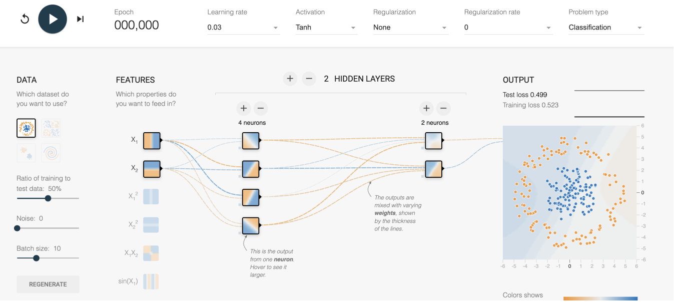
- ...
- 1995: Support Vector Machine (SVM)
- 1986: Restricted Boltzmann Machine (RBM)
- 1986: Iterative Dichotomiser 3 (ID3) for generating decision trees.
- 1974: Multilayer Perceptron (MLP)
- 1970: Automatic Differentiation (AD, e.g., Chain Rule)
- 1969: XOR Logic Function
- 1960: Least Mean Squares (LMS)
- 1959: Machine Learning for the Game of Checkers.
- 1957: Perceptron (IBM 704).
- 1946: ENIAC (Electronic Numerical Integrator and Computer)

Introduction to Deep Learning

Deep Learning

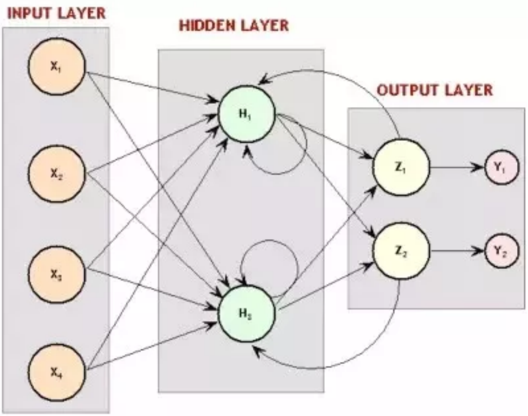
- CNNs (ConvNets) and RNNs (LSTM/GRU)
- Transformer models
- Region-Based CNN (R-CNN)
- Fast R-CNN and Faster R-CNN
- YOLO and SSD models
- AlexNet, ResNet, GoogLeNet, etc.
- Generative Adversarial Networks (GANs)
- Autoencoder
- Reinforcement Learning
- Transfer Learning
- Ensemble Learning
- ……

Deep Learning Playground



Deep Learning Playground: https://playground.tensorflow.org

RNNs: Recurrent Neural Networks



LeCun, Y., Bengio, Y. (1995) Convolutional networks for images, speech, and time series. The Handbook of Brain Theory and Neural Networks, pp. 255-258, MIT Press, USA.

RNNs: Recurrent Neural Networks

- RNNs process an input sequence maintaining in the hidden units that implicitly contains information about the history of all the past elements of the sequence.
- RNNs can be seen as very deep feedforward networks in which all the layers share the same weights.
- Turing machines (e.g., FSM) and memory networks are being employed for the tasks that would normally require reasoning and symbol manipulations.

LeCun, Y., Bengio, Y., Hinton, G. (2015) Deep learning, Nature, 521, pages 436 - 444.

RNNs: Recurrent Neural Networks

- Most NLP systems relied on gated RNNs, such as LSTMs and gated recurrent units (GRUs), with added attention mechanisms.
- RNNs (LSTM, GRU, etc) have been firmly established approaches in sequence modelling and transduction problems such as language modelling and machine translation.
- RNN models typically factor computations along the symbol positions of the input and output sequences.
- This nature of RNNs precludes parallelisation within training examples.

https://en.wikipedia.org/wiki/Transformer_(machine_learning_model)

Transformer Models

- Transformer is based solely on attention mechanisms, dispensing with recurrence and convolutions entirely.
- Transformers are the current state-of-the-art type of model for dealing with sequences, e.g., in text processing, machine translation, etc.
- Transformers were introduced in 2017 by Google Brain for NLP problems, replacing RNN models (LSTM).
- Transformer models are trained with large datasets.
- Transformer models can be fine-tuned for specific tasks.

https://en.wikipedia.org/wiki/Transformer_(machine_learning_model)

Transformer Models

- Transformer is a deep learning model that adopts the mechanism of self-attention, deferentially weighting the significance of each part of the input data.
- Like RNNs, Transformers were designed to handle sequential input data.
- Unlike RNNs, Transformers do not necessarily process the data in order. The attention mechanism provides context for any position in the input sequence.
- Transformer allows for more parallelisation than RNNs, therefore reduces training times.

https://en.wikipedia.org/wiki/Transformer_(machine_learning_model)

Questions?

Questions?

The popular deep learning model at present is:

1. Transformers
2. LSTM
3. GRU
4. None of the given options

The right answer is:___

Questions?

Learning Objectives

- To get organized for the course and to provide an overview of the core issues.
- Demonstrate critical thinking in designing problem solutions for tasks in robotics.