Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

INFORMATION ENGINEERING DEPARTMENT

COMPUTER ENGINEERING - AI & ROBOTICS

3D Data Processing 2023/2024

Francesco Crisci 2076739

Contents

Chapter 1

Intoduction concepts

1.1 Sensors

A point cloud is a data structure used to represent a collection of **multi-dimensional points** and is commonly used to represent three-dimensional data. The points usually represent the X,Y and Z geometric coordinates of a sampled surface. Each point can hold additional information: RGB colors, intensity values and so on.

1.1.1 Proximity sensors

An emitter transmits a light wavelength. Light eventually reflected back from a nearby object, e.g. perceived by a photoresistor.

1.1.2 Sonar

Senses the range of an obstacle from the round-trip time of an ultrasound pulse. The narrower the cone, the higher the angular resolution of the sensor.

1.1.3 LiDAR

It stands for Light Detection And Ranging. We have two types: **Continuous wave**, and **Pulse-based**: it measures directly the time-of-flight, i.e. the round-trip-time of a pulse of light. Needs very short laser pulses and high temporal accuracy.
Solid-state LiDAR uses non-mechanical elements to deflect the laser beams. Usually an optical phase array (OPA) of closely spaced optical antennas.

1.1.4 Radar

Radio Detection and Ranging are similar to LiDARs, but emit radio waves. **FMCW**(Frequency-Modulated Continuous-Wave) radars emit a signal where the frequency varies linearly over time. The difference between the frequency of the signal sent and that received is linearly related to the distance from the object that generated the reflected signal.
Lower angular and range resolution/accuracy w.r.t. LiDARs, but more robust to environmental sonditions cush as snow, fog, ... and able to compute the relative velocity between the sensor and the objects. In **imaging radar**, multiple FMCW transceivers are cascaded together tpo generate a dense 3D point cloud of the surrounding environment.

1.1.5 Time-of-flight Cameras

Meeting point between LiDARs and digital cameras. IR light emitters illuminate the scene, sensed by a 2D imaging sensor sensible to the projected light.

1.1.6 Stereo Cameras

Two or more usually identical, rigidly mounted cameras, framing the same scene from different point of view. Knowing the rigid body transformation between the two cameras and the pixel onto which a 3D point is projected in both cameras, it is possible to estimate the depth of the 3D point by usinbg triangulation techniques.
Active stereo cameras: To create visual saliency also in homogeneous surfaces. Couple the digital cameras with a dense pattern projector that projects into the scene some visible textured pattern.
Structured Light Camreas: SL cameras employ the same operating principle of stereo cameras, but one of the two cameras is replaced with a light projector that illuminates the scene with a textured visual pattern.

The pattern is known and each pattern patch represents a unique binary matrix.
Time-Multiplexing coding: Projecting a series of light patterns so that every pixel is encoded with the codeword identified by the pattern sequence. Most common structure of the patterns is a sequence of stripes in(de)creasing their width by the time.

1.1.7 RGB-D Cameras

RGB-D cameras are sensor ensembles composed by a ToF or a SL camera rigidly coupled in the same chassis with a color camera. Each pixel of the RGB image is associated with a corresponding pixel in depth map.

1.2 Linear Algebra and Rigid-Body Transofrmationas

When dealing with 3D data, we often need to change the coordinate frame, move sensors, move objects in front of sensros, etc...
Many 3D sensors are based on ptojective geometry.
Euclidean Space: E^N, the set of n-dimensional points that satisfy the axioms of Euclidean geometry. Relationships can be expressed in terms of angles and distances. No special point, no additions between points, only differences (**vectors**).
Linear Vector Space: A set V of objects over the field F is a vector space if its elements are closed under **scalar multiplication** and **vector summation**. Given two vectors $u, v \in V$ and $\alpha, \beta \in \mathbb{F}$, we have $\alpha u + \beta v \in V$.
Euclidean Vector Space: a linear space denoted by $\mathbb{R}^N$ is called **real vector space**, each element defined by an ordered tuple. Any point in E^N can be identified with a vecotr in R^N.
Dot Product: Given two vector u,v in R^N, we define the canonical inner product, or dot product, as:

$$< u, v >= u^T v = u_1 v_1 + ... + u_n v_n$$

Distances and Angles: dot product is used to measure distances: Euclidean norm or L2-norm is just the square root of the dot product. Dot product is also used to measure angles.
Subspaces: a subset $w \subseteq V$ of a vector space V is called a subspace if the zero vector belongs to W. given a set of vectors S, the spanned subspace is :

$$span(S) = \left\{ \sum_{i=1}^{n} \alpha_i v_i | v_i \in S, \alpha_i \in \mathbb{R} \right\}$$

S is linearly independent if $\sum_{i=1}^{n} \alpha_i b_i = 0 \Rightarrow \alpha_i = 0 \forall i = 1, ..., m$.
Basis: A seto of vectors S is said to be a basis for V if is linearly independent and it spans the whole space V. There are other notions from linear algebra, but it is the fifth year seeing this stuff, so i'm gonna skip them. Feel free to look at the slides.

1.2.1 Rigid Objects

A non-deformable objecy O in 3D can be modeled by a set of points in a Euclidean vector space. We assume that the coordinate frame of this space represents an inertial frame we call the **world** frame. If the objecy starts to move, its points move accordingly, i.e., they change their coordinates p(t) w.r.t. the world frame over time. Distancces between any of these two points u(t), q(t) must be preserved over time as the object moves:

$$||u(t_1) - q(t_1)|| = ||u(t_2) - q(t_2)||, \forall t_1, t_2 \in \mathbb{R}$$

A transformation that preserves distances is called a Euclidean transformation. Not enough to describe a rigid body movement. Consider reflection: it does not preserve, generally, the cross product.
Individual objects points cannot move relative to each other. Attach a coordinate frame to the object. Rigidbody motion w.r.t. the world frame can be described by the motion of the origin of the object frame and the rotation of the object frame.
Object pose is defined by translational and rotational part. The translational part is a vector t between the origin of the world frame and that of the object frame. The rotational part is a relative orientation R of the object frame axes w.r.t. fixed world frame axes. In practice, we apply R first, then add the translation t.
Suppose the object is performing a pure rotation around its origin o. We may always assume that the origin of the world frame is the center of rotation o.

1.2.2 Rotation Matrix

All objects coordinates p_o are related to R^3 standard basis in the object frame:

$$p_o = [x_o \; y_o \; z_o]^T = x_o e_{o,1} + y_o eo, 2 + z_o eo, 3$$

To represent the positon of any point w.r.t. world, we just need to present $e_{o,1}, e_{o,2}, e_{o,3} : T(e_{o.,1} = r_1), T(e_{o.,2} = r_2), T(e_{o.,3} = r_3)$ More math on the slides, including the slides "FROM OBJECTS TO CAMERA".

1.3 Homographies