Lec13_3D_Construction

Stereo vision 立体视觉

- is the extraction of 3D information from digital images, such as
- comparing info about a scene from two vantage points视点, 3D info extracted by examining relative positions of objects in two panels.

Multi-view representations

- A set of 2D img correspond to pics, given 3D shape from different viewpoints.
- dependent on lighting and a high number of img might be required to cover all the angles of a given 3D shape

RGB-D images

- color image that contains depth information at each pixel(ie. distance b/t camera & object)
- This distance encodes info about 3D geometry from a fixed point of view.
- img easily captured with relatively cheap hardware - many dataset of RGB-D img exists

Ultrasonic images

- created by sending pulses of ultrasound into tissue using a probe.
- ultrasound pulses **echo** off tissues with different reflection properties and **returne**d to the probe which records and displays them as an image.

Photometric stereo光度立体

- estimating surface normals of objects by observing that object under different lighting conditions.
- FACT: amount of light reflected by a surface is dependent on orientation of surface in relation to light source and observer.

Point clouds

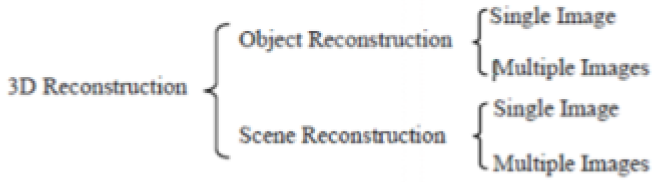
- a set a 3D vertices with coordinates x,y,z), usually captured by 3D scanners.

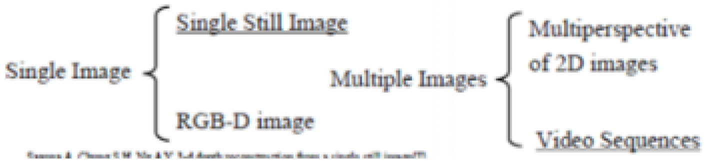
Voxel(volume element-cubic) grid representation:

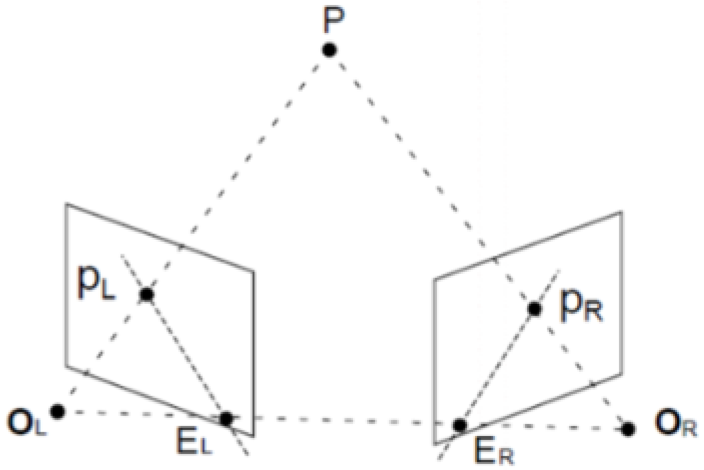
- element” and is an extension of 2D pixels to 3D.
- A voxel represents a value on a regular grid in 3D space.

Meshes网格

- a set of vertices connected to each other in order to form triangles (or sometimes quadrilaterals).
- forms a 2D surface in a 3D space
- common data structure usually employed by 3D renderers. - compute lighting effects
- Gather the good points- requires many views otherwise holes appear







Single Still Image Object Reconstruction

Challenges in deep learning methods:

1. shape complexity of objects;
2. uncertainty of objects;
3. reconstruction of fine grained objects;
4. memory requirements and calculation time;
5. training datasets;

2D image → 2D encoder(transfomer/CNN)→ latent space → 3D encoder → 3D representations (Voxel-based 3D decoder, Point cloud-based 3D decoder, Mesh-based 3D decoder

Reconstruction based on Video Sequences

Structure from Motion (SfM)

- process of reconstructing 3D structure from its projections into a series of images taken from different viewpoints.
- Incremental SfM is a sequential processing pipeline with an iterative reconstruction component.

Correspondence Search:

- correspondence search which finds scene overlap in the input images and identifies projections of same points in overlapping images.
- output: set of geometrically verified image pairs and a graph of image projections for each point.

Feature Extraction:

- invariant under radiometric and geometric Changes ← SIFT used to extract local features.

Image Matching

- tests scene overlap; searches for feature correspondences

Geometric Verification

- verifies the potentially overlapping image pairs. matching is based solely on appearance, it is not guaranteed that corresponding features actually map to the same scene point. → SfM verifies the matches by estimate a transformation maps feature points between images using projective geometry.

Incremental Reconstruction:

- input: scene graph.
- outputs estimates for registered images and the reconstructed scene structure as a set of points.

Image Registration:

- Starting from a metric reconstruction, new images can be registered to current model by solving Perspective n Point (PnP) problem using feature correspondences to triangulated points in already registered images (2D-3D correspondences).

Triangulation:

- A newly registered image must observe existing scene points. In addition, it may also increase scene coverage by extending the set of points X through triangulation.

Bundle Adjustment:

- Bundle Adjustment is the joint non-linear refinement of camera parameters and point parameters that minimizes the reprojection error.