Computer Vision

3D vision

Zdenek Krnoul, Miroslav Jirik

Department of Cybernetics
Faculty of Applied Sciences
University of West Bohemia

Obsah

- Introduction to 3D vision (Zdeněk Krňoul)
  - Types of tasks
  - 3D vision theory
  - Input data and recording principles
- 3D data representation (Miroslav Jiřík)
  - Current techniques

3D vision - Two types of tasks:

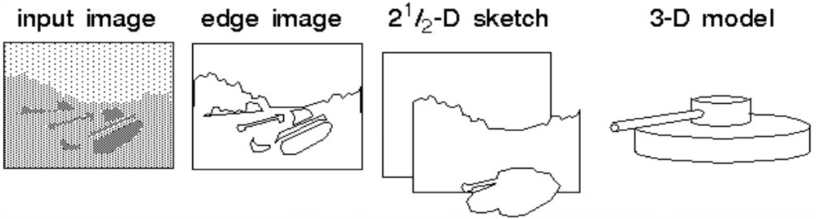
- 3D reconstruction:
  - Search for geometric and physical parameters of objects in the scene;
  - It is possible to reconstruct 3D shape of objects (3D scanners, 3D face - identification, etc.)
- 3D recognition / estimation:
  - Search and classification of objects in the scene (car or robot navigation, etc.)
  - Algorithms are used to determine position, orientation, scale, etc.
- bottom-up, top-down (model-based vision)

Note In practical tasks, we often try to avoid 3D vision and convert the task to 2D. The reason is the requirements for robustness, low cost and high efficiency of the use of computer vision systems.

3D vision - Marr's theory, 1982

4 levels of 3D scene representation:

- intensity image
- primal sketch (edge image)
- 2 1/2 dimensional sketch
- full 3D representation



The procedure is bottom-up and significantly reduces the computational complexity.

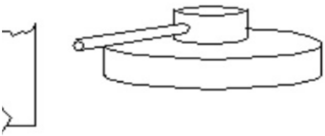
Primal sketch:

- Includes information about the sizes and directions of significant brightness changes in the image;
- Mutual geometric arrangement;
- It assumes that the lines and spots obtained in this way preserve the information needed for the later 3D representation.

2 1/2 dimensional sketch:

- It uses the information contained in the primal sketch;
- Various techniques can be integrated, collectively called ”Shape from X”:
  - Shape from Stereo vision;
  - Shape from motion (optical flow);
  - Shape from brightness;
  - Shape from texture;
- basic information about depth.

Full 3D representation of the object:



- It relies on geometric properties that **can be found** in the image;
- The geometric properties are expressed with respect to **a coordinate system** based on the shape of the object;
- Basic geometric properties:
    - center (most often center of mass)
    - total size
    - generalized symmetry axis (if any)

Input data, recording options

- Intensity image(s) - (RGB) (often multiple views of the scene from different angles are needed);
- Stereo vision - just two intensity images, calibrated cameras;
- Depth sensors - a common practice today.



Depth map, 3D point cloud

1. Algorithms based on various principles:
    - lens focusing technique;
    - stereo vision (dense stereo) - from intensity images
    - techniques based on structured light or light reflection;
    - known texture parameters, etc.
2. We often use a specialized sensor or camera, such as the well-known MS Kinect, Intel RealSense, etc.

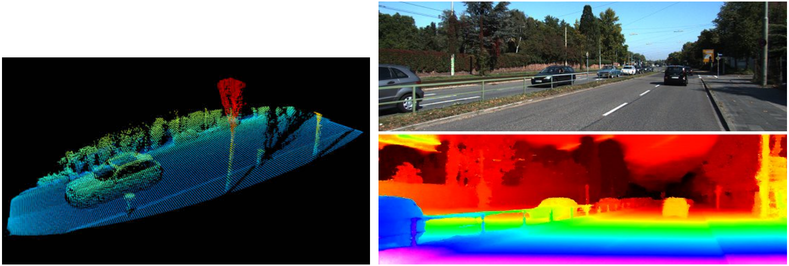


Figure: Blickfeld, KITTI dataset

3D sensors:

- **Structured light** - the scene is illuminated by an auxiliary light source;

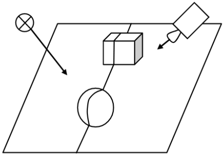
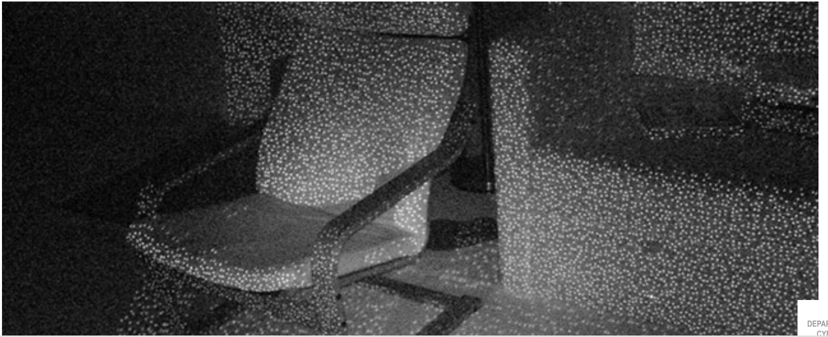


Figure: Common principle of devices in obtaining a depth map.

- **Laser** - similar principle as radars and sonars.

The principle of structured lighting:

- Scene / surface illuminated eg by a narrow strip of light or a unique pattern (MS Kinect I - infrared) and the scene is captured from a different angle, (shadow problem);
- Necessary calibration of the projector and camera (often built in the device);
- We need an algorithm for decoding the projected pattern;
- Use the stereo vision principle, see later;
- Good accuracy is obtained for one-color non-glossy surfaces, with Kinect I, it interferes with other IR radiation (eg the sun).



Moiré stripes:

- Usually three or four phase shifted sine patterns are projected onto the surface
- the frequency of the patterns corresponding to the camera-pixel grid (image sensor);
- Capturing a single image for each pattern;
- Deformed images captured by the camera are used to calculate the phase map, which contains information about the depth of the surface;
- High precision is obtained for one-color non-glossy surfaces.

(a) (b) (c) (d)

Note There are many other methods based on a similar principle.

Laser sensors:

- Operating in the near-infrared range (850nm) invisible to the human eyes;
- Source typically a sinusoid or square wave;
- Measures the phase shift between the sent and received signal (eg 3D spot - lidar, or as an image - ToF camera - MS Kinect II)

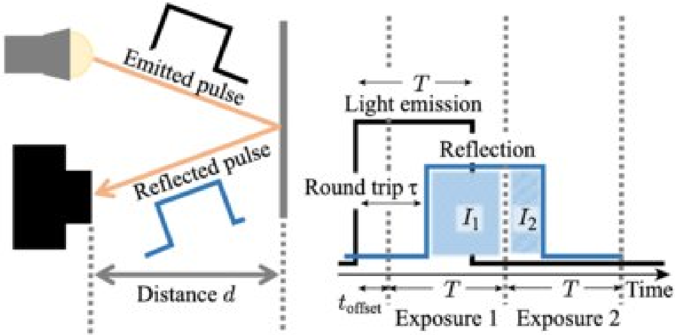


Figure: Taken from [KOT+17]

Variable lens focus:

-

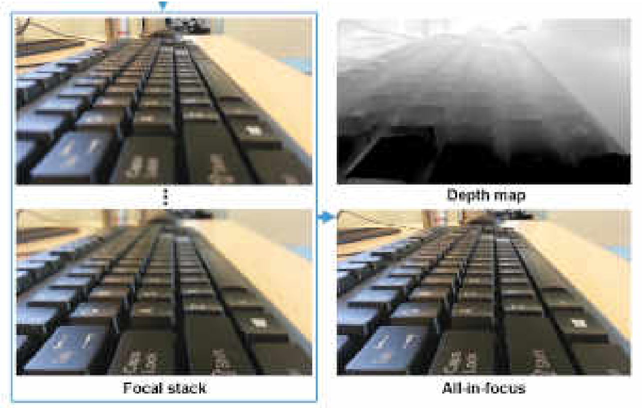
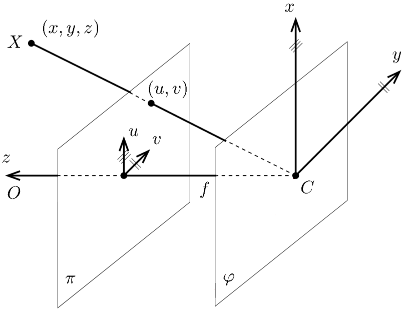


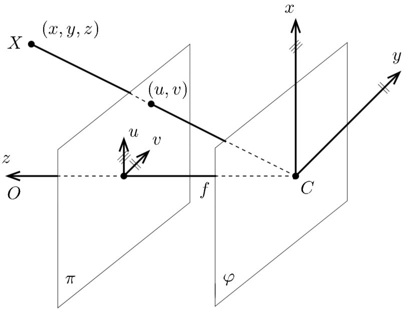
Figure: For mobile phone, taken from [SHS15]

Stereo vision:

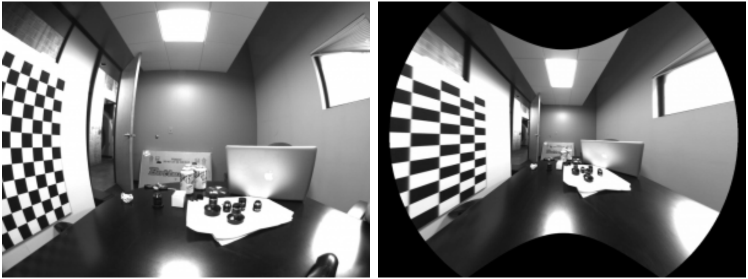
- Uses general **perspective camera model** - projection of $3D$ space into $2D$ image plane;
- This is always a **Center projection**.



Note special case ... the center of the projection lies at infinity → **affine camera** *and it is a generalization of the so-called parallel projection*

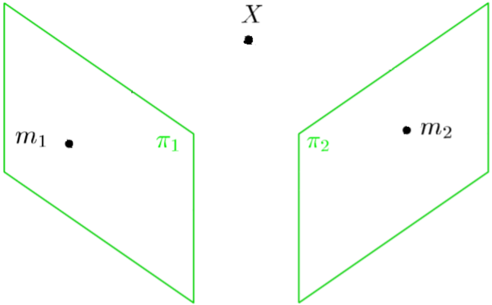


- Perspective camera has 11 degrees of freedom (DOFs): 1x focal distance in pixels + 1x pixel aspect ratio + 1x axis skew + 2x origin of coordinates + 3x shift + 3x camera rotation = 11 DOFs;
- We will use the camera calibration algorithm (linear in homogeneous coordinates)
- Theoretically 5.5 space points are enough

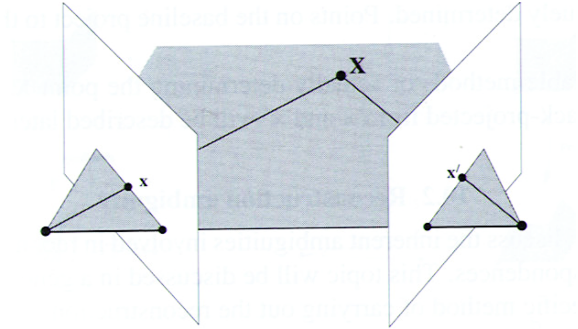


- In practice, it is necessary to correct the coordinates of the image - radial distortion
- The effect is caused by an imperfection of the center projection, which is physically done through the lens

- In stereo vision, we have two views of the same scene from different directions
  - Views can be captured simultaneously at one time (eg two cameras) (Stereovision, Binocular disparity)
  - Or, movement of the camera in front of the static object (eg robot movement) (Egomotion Estimation)
  - Or, the movement of the object in front of the static camera (eg I turn the apple in front of the camera) (Structure from Motion - SFM)
- These tasks are dual and lead to the same solution



Stereo vision - 3D reconstruction



- If we know both cameras calibrated;
- and at the same time we have a corresponding pair;
- then we can determine the 3D coordinates of an unknown point X.

Shape from texture:

- Regular pattern (known)
- 3D shape from expected projection

Shape from Shading - SFS:

- Assumption of known light source position
- Lambert surface (reflects light in all directions the same)

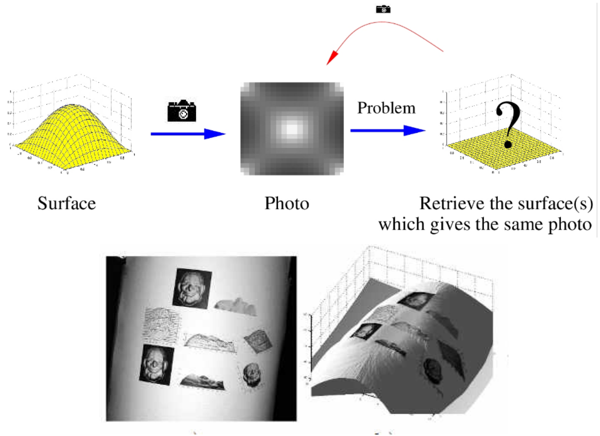


Figure: taken from [PF05]

3D object models

Can be divided into:

- Descriptive - fully describe the shape of the object, defined reflectivity and lighting, from such a model it is possible to create a synthetic intensity image and a depth map for any place of observation;
- Discriminative - only to distinguish objects of several classes.

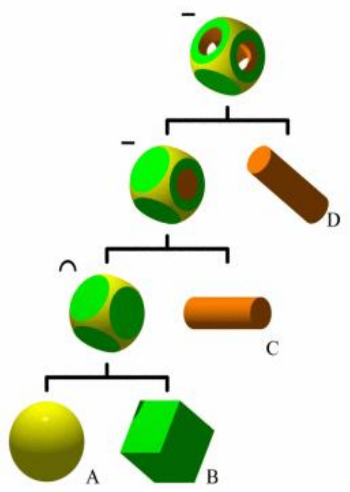
Types of representation used:

Wireframe

- a graph whose vertices correspond to 3D points (often object corners), edges correspond to boundaries (discontinuities of normals to the surface). Not suitable for describing objects with curved surfaces;

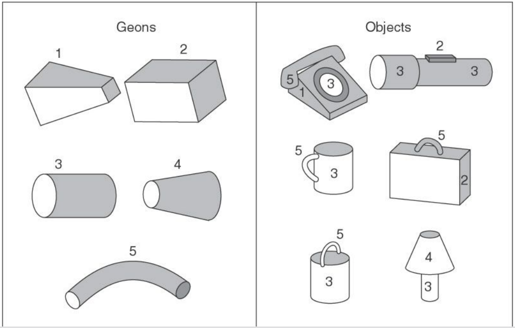
CSG (constructive solid geometry) model

- uses a **set of simple 3D objects**, such as a cone, cylinder, sphere, cube, block, as a basis;
- combines them in a certain position, magnification and orientation using simple set operations, such as intersection, unification, difference, etc;
- The model is represented by **a tree**; the sheets correspond to individual elementary bodies, the higher nodes represent set operations;
- representation is computationally very demanding.



GEON (geometrical ions)

- representation describes the qualitative properties of objects;
- 3D objects are composed of several GEONs;
- Designed 36 basic GEONs;
- Each GEON is characterized by four quality features:
  - border - straight / curved
  - symmetry - center / axis / none
  - resizing - constant / increasing / increasing and decreasing
  - axis - straight / curved



VOXEL (volumetrix element)

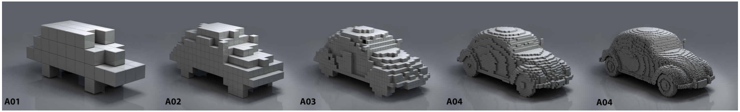
- 

Figure: http://www.bilderzucht.de/blog/3d-pixel-voxel/

References I

Kazuya Kitano, Takanori Okamoto, Kenichiro Tanaka, Takahito Aoto, Hiroyuki Kubo, Takuya Funatomi, and Yasuhiro Mukaigawa.
Recovering temporal psf using tof camera with delayed light emission.
IPSJ Transactions on Computer Vision and Applications, 9, 12 2017.

Emmanuel Prados and Olivier Faugeras.
Shape from shading.
In *in Mathematical Models in Computer Vision: The Handbook*. Springer, 2005.

Supasorn Suwajanakorn, Carlos Hernandez, and Steven M. Seitz.
Depth from focus with your mobile phone.
In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.