Robot Vision:
Stereo Matching
Depth cameras

Ass.Prof. Friedrich Fraundorfer

SS 2018
Outline

- Geometric relations for stereo matching
- Dense matching process
- Census Transform
- Semiglobal matching
- Depth cameras
  - Coded light
  - Kinect style depth cameras
  - TOF cameras
Dense matching

- SfM only gives sparse 3D data
- Only feature points (e.g. SURF) are triangulated – for most pixel no 3D data is computed
- Dense image matching computes a 3D point for every pixel in the image (1MP image leads to 1 million 3D points)
- Dense matching algorithms need camera poses as prerequisite
Geometric relation

- Stereo normal case
- Depth $Z$ [m] can be computed from disparity $d$ [pixel]

$$d = f \frac{B}{Z}$$
Rectification

- Image transformation to simplify the correspondence search
  - Makes all epipolar lines parallel
  - Image x-axis parallel to epipolar line
  - Corresponds to parallel camera configuration

Before rectification

„Stereo normal case“
Dense matching process

- Estimate disparity (height) for all pixels in image left.
  - Evaluate correspondence measure for every possible pixel location on the line (e.g. NCC, SAD)

- Disparity $d$: Offset between pixel $p$ in the left image and its corresponding pixel $q$ in the right image.

$I_l$  +  $I_r$  =  $D$
Census Transform

- A popular block matching cost
- Good robustness to image changes (e.g. brightness)
- Matching cost is computed by comparing bit strings using the Hamming distance (efficient)
- Bit strings encode if a pixel within a window is greater or less than the central pixel

\[
\begin{array}{ccc}
89 & 63 & 72 \\
67 & 55 & 64 \\
58 & 51 & 49 \\
\end{array}
\]

\[00000011\]
Dense matching process

- Reference image
- Matching image
- Epipolar line
- Matching cost
- Depth
Dense matching process

Matching image

Matching cost

Depth

Cost column
Dense matching process
Semiglobal matching

Dense Cost Computation

Cost Aggregation

LR Check, Filtering

Disparity images $D(p), D(q)$

Disparity Selection

Depth cameras - Overview

- Depth cameras or RGBD cameras directly output an RGB image and a depth image

- Principles:
  - Stereo cameras with onboard processing
    - DJI Guidance, Roboception, Perceptin
  - Structured Light
    - Coded light – Projector-camera system
    - Random patterns – Stereo system with active lighting
    - Kinect-style methods – Projector-camera system with fixed random projection
  - TOF cameras – time of flight principle

- Huge importance for mobile robotics
Coded Light

- A projector-camera system where a projector outputs stripe patterns (e.g. binary pattern)
- The pattern solves the correspondence problem in stereo matching
- Projector and camera need to be calibrated
- The stripes are coded and encode directly a unique position of a corresponding pixel in the projector.
Coded light

sequence of dark/light pattern defines a code word for a position and thus a unique position in the projector

pattern changed over time
Coded light
Coded light
Coded light
Coded light
Coded light
Coded light
Random patterns

- Stereo camera system cannot measure depth in textureless/homogeneous areas
- Solution: Project random pattern as texture to ease stereo matching
- Typically this is done in infrared spectrum such that it is not visible for users
- Such a system works in the dark as well

- Example: Intel Realsense
- Standard stereo system (2 calibrated cameras)
  - 1 IR projector for stripe-type pattern
- Works outdoors as well, then the pattern is not visible due to strong sunlight (then it just works like standard stereo matching)
Kinect style method

- Kinect is a projector-camera system with onboard depth processing
- Projects a **known** static IR-dot pattern
- Depth is computed from a combination of depth from stereo and depth from focus
- The system also contains an RGB camera
- Sensors is often called a RGBD sensor
Time-of-flight cameras

- Does not work on the stereo (triangulation) principle but with time-of-flight principle

- Principle:
  - Sends out NIR light (no spatial coding)
  - Sensor array measures response
  - Distance is measured by measuring time between emitting and receiving the light (pulsed or continuous wave method)

- Typically do not provide synchronized color image but a reflectance image

- Example: PMD Flex (224 x 171px resolution)
Time-of-flight cameras – Continuous wave method

- Camera emits NIR light where amplitude is a sine wave.
- Phase shift is measured between emitted and received light
- Phase shift can be converted into distance

\[ d = \frac{c}{4\pi f} \Delta \phi \]