Master MLDM, AIMA and 3DMT - Computer Vision course

Exam March 2020 - 2h without documents

(4 parts with a total of 23 questions accounting for 30 points + 3 bonus points, the exam will be scored for 26 points)

Part 1 (30 mns)**:** 3D reconstruction

Question 1 *(2 points)***:** Show how to map the coordinates of a 3D point in the scene to the coordinates of a 2D point in the image.

Question 2 *(1 point)*: Is it possible to recover depth from a monocular point of view? Justify

* **Question 3** *(2 points)*: Show how to map the coordinates of interest points in multiple 2D images of the same 3D scene viewed from different points of view.

Question 4 *(1 point)*: Is it possible to recover depth from a binocular point of view? Disparity

Question 5 *(2 points)*: The two images below are a stereo pair taken using parallel cameras, aligned horizontally. Describe a framework (i.e. a solution) for matching corresponding features in the two images and estimating their disparities. State clearly how your method exploits the epipolar constraint.

Epipolar lines are parallel

Find corresponding pts.

correlation

learned methods

Question 6 *(2 points)*: Explain how to obtain estimates of the 3-D positions of the objects in the scene above from stereo disparities. What information about the cameras is required for this?

$d = \frac{fb}{z}$

Question 7 *(2 points)*: If only one of the two images in the pair above is available, what image features might be used to make inferences about the 3-D structure of the scene?

Fundamental equation of the thin lens.

Part 2 (20 mns)**:** pinhole model

Question 1*(1 point)***:** What is the name of the mathematical equation of a thin lens which models the relationship between the focal f, the distance Z of a point in the scene to the camera frame and the distance z of its 2 projection in the image frame? Write this equation.

$\frac{1}{Z} + \frac{1}{z} = \frac{1}{f}$

perspective transform

Question 2 *(2 points)*: With the pinhole model the most distant objects appear smaller, why? With the pinhole model parallel lines in the scene can intersect in the image, show with a graphic why?

$x = f\frac{X}{Z}, y = f\frac{Y}{Z}$ $Z\uparrow$ $y, x\downarrow$

Question 3 *(3 points)***:** For the two images shown below estimate manually the positions of the three major vanishing points, corresponding to the building orientations. Do these vanishing points are located within the image plane? If yes, can we state that we have a weak or a strong perspective?

Both strong

Part 3 (25 mns)**:** In this part, we will consider a system consisting of a Time-of-Flight (ToF) camera rigidly connected with a regular, high-resolution RGB camera in a stereo setup. We will see how to calibrate the cameras, how to reconstruct 3D point clouds from the ToF depth data and how to match the depth information with the color information from the RGB camera. See hints in the Annex.

Question 1 *(2 points)***:** Explain the necessary process for determining the intrinsic camera parameters K_{rgb} and K_{tof} of the RGB and the ToF camera, respectively.

Camera calibration

Question 2 *(1 point)*: Explain the calibration process that is necessary as a prerequisite to relate the pixels in the ToF depth image to their corresponding pixels in the RGB image.

$\begin{pmatrix} s_x & & o_x \\ & s_y & o_y \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$

Question 3 *(2 points)*: Outline how to reconstruct a metric 3D point cloud from the ToF depth data. Assume that you are given the depth d at each pixel location $x = (x, y)^T$ in the ToF depth image. We would like to obtain a set of 3D points with coordinates $X = (X, Y, Z)^T$.

Question 4 *(1 point)*: Show how to map each 3D point to the coordinate frame of the RGB camera.

Question 5 *(1 point)*: Explain how to find the corresponding color value for each 3D point.

Part 4 (45 mns)**:** In this part, we will consider that $\mathbf{x} = (x, y, 1)^T \in \mathbb{P}^2$ and $\mathbf{x}' = (x', y', 1)^T \in \mathbb{P}^2$ are the measured projections of an unknown 3D points $\mathbf{X}$ on the two images I_1 and I_2 from two different views of the same scene. Let **F** be the fundamental matric computed between the two views. Let **P** and **P'** be the projection matrices of the two views. These two matrixes are assumed to be known exactly (or with a very good accuracy), while the measures **x** and **x'** are supposed to be inaccurate, i.e. they do not satisfy the epipolar constraint.

In order to obtain a solution for the triangulation problem (i.e. finding the best estimate $\hat{\mathbf{X}}$ of $\mathbf{X}$), two corrected correspondences $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ have to be computed in order to minimize the geometric error:

$$C(\hat{\mathbf{x}}, \hat{\mathbf{x}}') = d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2 \text{ subject to } \hat{\mathbf{x}}'^T \mathbf{F} \hat{\mathbf{x}} = 0$$

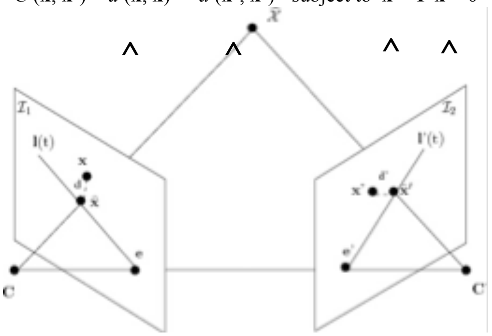


Figure 1: Unlike the measured points **x** and **x'**, the corresponding points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ satisfy the epipolar constraint. The point $\hat{\mathbf{X}}$ is chosen such that the $d^2 + d'^2$ is minimized.

An optimal solution can be found by using the following steps:

- Two rigid transforms are applied to the measurements and epipoles in order to simplify the equations.
- The epipolar lines in the first image are parameterized by a parameter *t*. An epipolar line in the image I_1 can be written as l(*t*).
- Using the fundamental matrix, the corresponding epipolar line l'(*t*) in the image I_2 is computed.
- The value of the parameter *t* that minimizes the cost function is estimated.

This solution is optimal if the Gaussian noise model can be assumed to be correct. In the following we will only address the first steps of this solution.

Question 1 *(1 point)*: What are the relationships between **x**, **x**’, **P**, **P’** and **X** ?

$x = Px$
$x' = Px^$

Question 2 *(1 point)*: What is the epipolar constraint (between **x** and **x**’) ?

x' have to be in the epipolar line $Fx = 0$

Question 3 *(1 point)*: Write down the 2D projective transformation matrices T_{3x3} and T'_{3x3} such that T_{3x3} $\mathbf{x} = \mathbf{x}_0 = (0, 0, 1)^T$ and T'_{3x3} $\mathbf{x'} = \mathbf{x'}_0 = (0, 0, 1)^T$? (Note that the two matrices correspond to 2D translations in order to put the measurements in the center of each image coordinate frame).

$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ . & . & . \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$

Question 4 *(1 point)*: Give the expression of the fundamental matrix $\mathbf{F}_0$ that relates $\mathbf{x}_0$ and $\mathbf{x'}_0$.

Question 5 *(1 point)*: Give the equations that make it possible to compute the right epipole $\mathbf{e} = (e_1, e_2, f)^T$ and the left epipole $\mathbf{e}' = (e'_1, e'_2, f')^T$.

Question 6 *(1 point - bonus)*: If we normalize **e** and **e**’ such that $e_1^2 + e_2^2 = e'^2_1 + e'^2_2 = 1$, give the expression of the rotation matrices **R** and **R**’ such that $\mathbf{R}\,\mathbf{e} = \mathbf{e}_0 = (1, 0, f)^T$ and $\mathbf{R}'\,\mathbf{e}' = \mathbf{e}'_0 = (1, 0, f')^T$ (Not that $\mathbf{e}_0$ and $\mathbf{e}'_0$ are on the horizontal axis).

Question 7 *(1 point - bonus)*: Give the general form of the matrix $\mathbf{F}_1 = \mathbf{R}'\,\mathbf{F}_0\,\mathbf{R}^T$ (i.e. give the value of each of its 3x3 coefficient).

$$\mathbf{F}_1 = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & a & b \\ \alpha_5 & c & d \end{bmatrix}$$

Question 8 *(1 point - bonus)*: Give the vector representing the parameterized epipolar line l(*t*) in the first image I_1 passing through $\mathbf{x}_1(t) = (0, t, 1)^T$ and $\mathbf{e}_0$.

Annex

Lecture Outline

1. Introduction and Motivation
2. Principles of ToF Imaging
3. Computer Vision with ToF Cameras
4. Case Studies
5. Other Range Imaging Techniques
6. Human Body Tracking with the Kinect

Computer Vision with ToF Cameras

Geometric Calibration of ToF Cameras

- Standard optics used in commercial ToF cameras
- Use ToF amplitude image for calibration
- Standard calibration procedure for camera intrinsics
  - $f_x = fm_x$: focal length in terms of pixel dimensions (x)
  - $f_y = fm_y$: focal length in terms of pixel dimensions (y)
  - c_x: principal point (x)
  - c_y: principal point (y)
  - Lens distortion parameters
- Typical approach:
  - checkerboard calibration pattern
  - World-to-image point correspondences
  - Linear estimation of intrinsic/extrinsic parameters
  - Non-linear optimization

Computer Aided Medical Procedures | Technische Universität München

Computer Vision with ToF Cameras

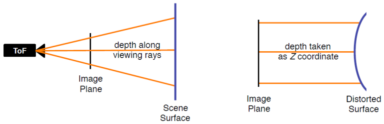
Geometric Calibration of ToF Cameras

- Standard optics used in commercial ToF cameras
- Use ToF amplitude image for calibration
- Standard calibration procedure for camera intrinsics
  - $f_x = fm_x$: focal length in terms of pixel dimensions (x)
  - $f_y = fm_y$: focal length in terms of pixel dimensions (y)
  - c_x: principal point (x)
  - c_y: principal point (y)
  - Lens distortion parameters
- Typical approach:
  - checkerboard calibration pattern
  - World-to-image point correspondences
  - Linear estimation of intrinsic/extrinsic parameters
  - Non-linear optimization

Computer Vision with ToF Cameras

Extraction of Metric 3D Geometry from ToF Data

- Simply taking measured depth *d* as *Z* coordinate is not sufficient
- Depth is measured along rays from camera center through image plane



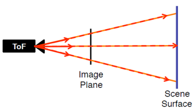
Computer Vision with ToF Cameras

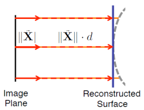
Extraction of Metric 3D Geometry from ToF Data

- Ray from camera center into 3D scene:

$$\begin{pmatrix} (x-c_x)Z/f_x \\ (y-c_y)Z/f_y \\ Z \end{pmatrix}^{\top} \rightarrow \begin{pmatrix} (x-c_x)/f_x \\ (y-c_y)/f_y \\ 1 \end{pmatrix}^{\top} =: \tilde{\mathbf{X}}$$

- Normalize to unit length (keep only direction), multiply with depth: $\mathbf{X} = \|\tilde{\mathbf{X}}\| \cdot d$

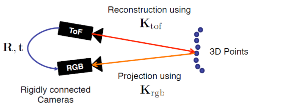




Computer Vision with ToF Cameras

Combining ToF with Other Cameras

- Additional, complementary information (e.g. color)
- Higher-resolution information (e.g. for superresolution)
- Example: combination with a high-resolution RGB camera
- Approach: Stereo calibration techniques, giving $\mathbf{R}, \mathbf{t}$ and $\mathbf{K}_{\text{tof}}, \mathbf{K}_{\text{rgb}}$



(1)

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} \xrightarrow[M1]{} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \xrightarrow[M2]{} \begin{pmatrix} x \\ y \end{pmatrix} \xrightarrow[M3]{} \begin{pmatrix} u \\ v \end{pmatrix}$$

$$M1 = [R \quad T] \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

$$M2 = \begin{pmatrix} f/z & & 0 & \\ & f/z & 0 & \\ & & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix}$$

$$M3 = \begin{pmatrix} s_x & & o_x \\ & s_y & o_y \\ & & 1 \end{pmatrix}$$

So,

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} f_x & & o_x \\ & f_y & o_y \\ & & 1 \end{pmatrix} [R \quad T] \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix}$$