Depth estimation from stereo images

IP final project

Bolundut Razvan-Mihai

Thechnical University of Cluj-Napoca (TUCN)
CTI Eng, Gr. 30434/1

May 31, 2023

Outline

Problem statement

STATEMENT Given two images captured from a stereo system compute an estimate of the depth of the scene.

INPUT-IMAGE *Left* and *right* images from the stereo system

INPUT-CALIB *calib.txt* file with stereo system calibration data

OUTPUT The matrix holding the depth information

Data-set

Data-set contains sets of images with the type mentioned in the Section 1 - Figures 1 and 2.

Together with the images a calibration file is provided with the structure in listing 1

Data-set

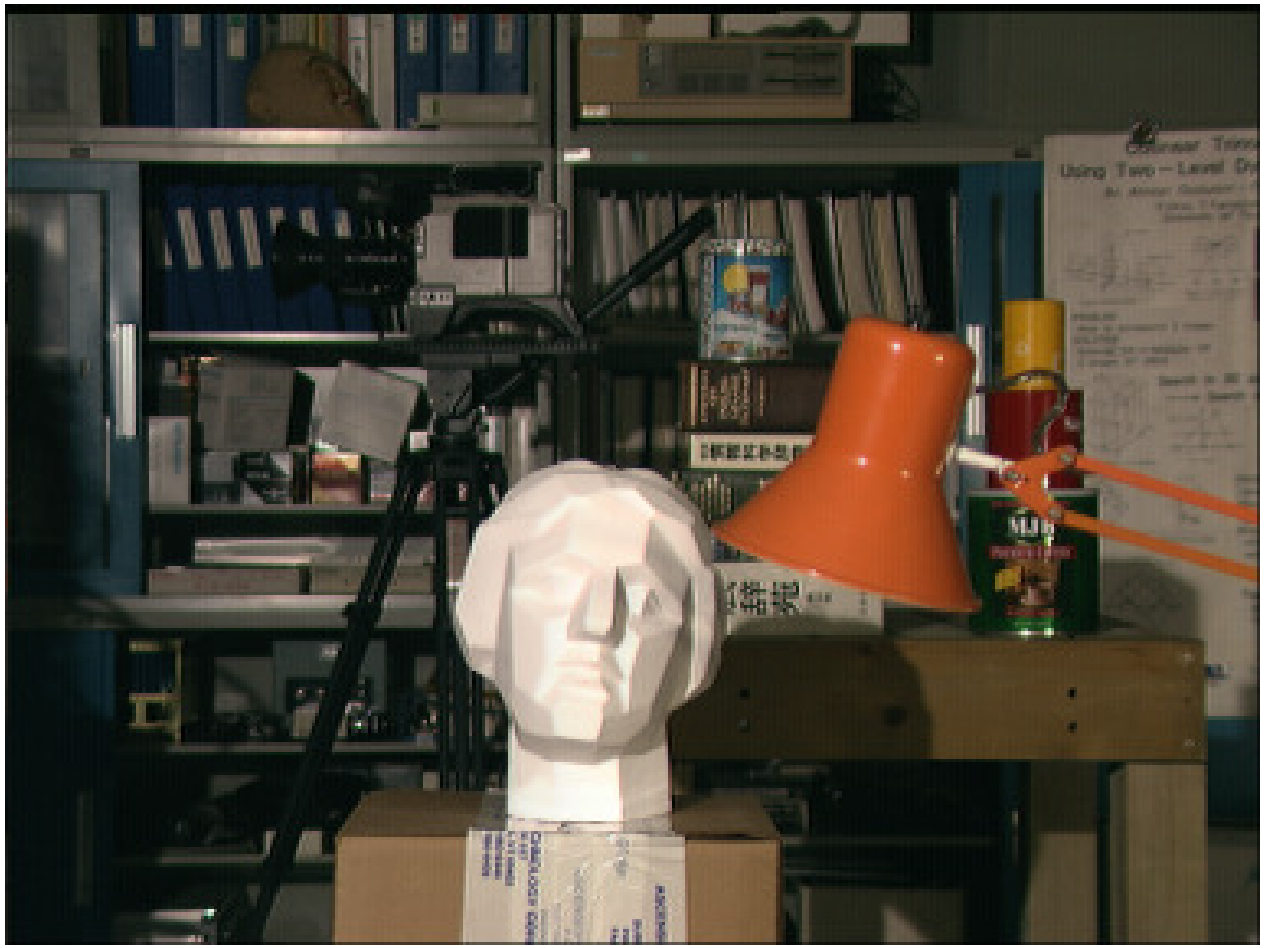

Figure: `left` image

Data-set

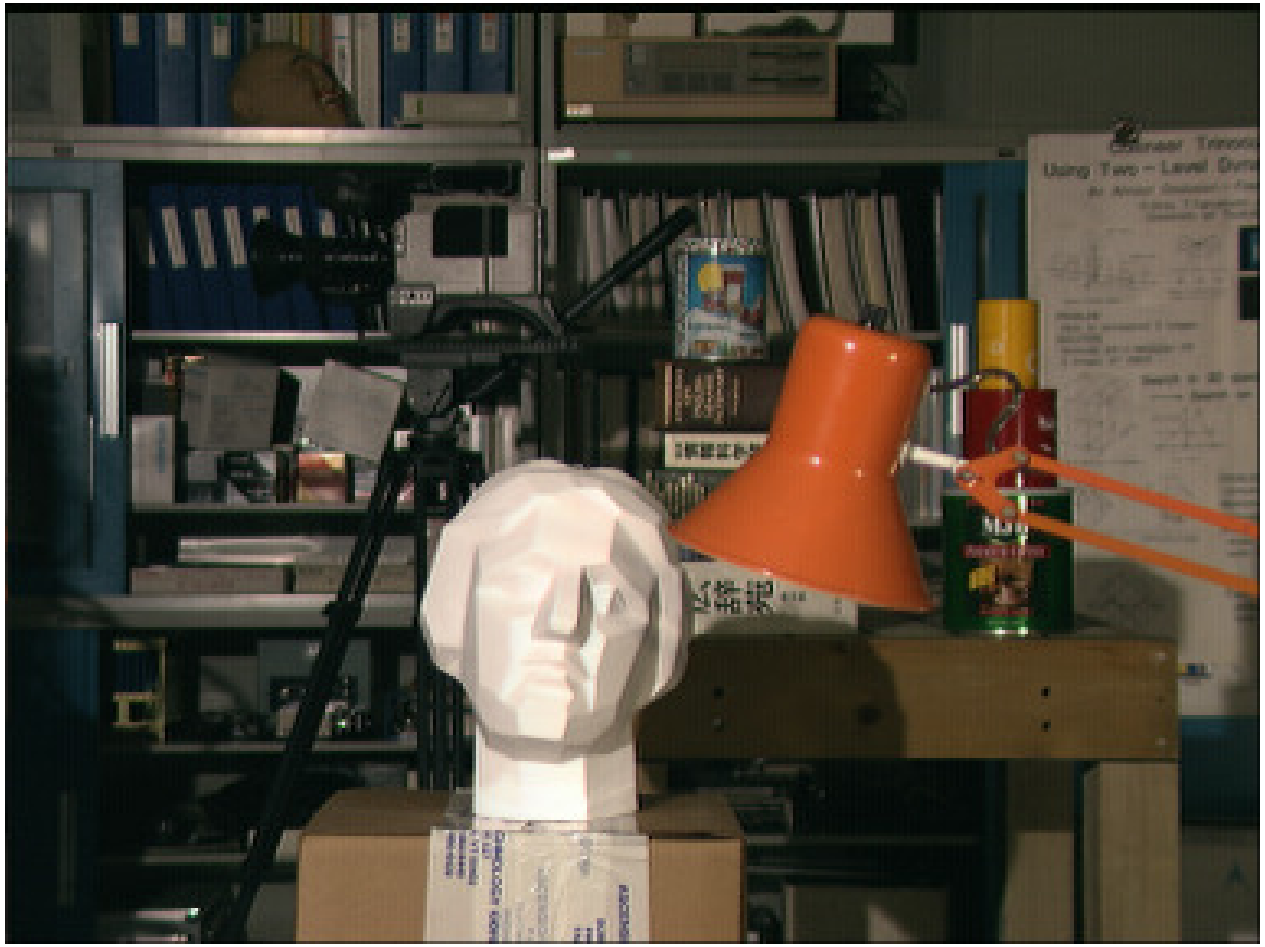

Figure: `right` image

Data-set

```
cam0=[4161.221 0 1445.577; 0 4161.221 984.686; 0 0 1]
cam1=[4161.221 0 1654.636; 0 4161.221 984.686; 0 0 1]
doffs=209.059
baseline=176.252
width=2880
height=1988
ndisp=280
isint=0
vmin=25
vmax=248
dyavg=0
dymax=0
```

Listing 1: calib.txt

Implementation

The implementation steps to get the disparity matrix are:

1. Transform color images to gray-scale
2. Transform gray-scale images with *census* transform [Zab05]
3. Compute correlation matrix between the 2 images
4. Compute the disparity from the correlation
5. Filter with median filter disparity to get gray-scale output

CENSUS transform

$$census_k = census_{k-1} << 1;$$
$$\text{census}_k = census_k \vee (img_{\{i-u+\frac{s}{2}, j-v+\frac{s}{2}\}} < img_{\{i,j\}});$$

$$\begin{bmatrix} 127 & 50 & 127 \\ 127 & 100 & 10 \\ 0 & 127 & 99 \end{bmatrix} \rightarrow \{0, 1, 0, 0, 0, 1, 1, 0, 1\}$$

Correlation matrix

- 3D matrix
- $corr_{\{i,j,k\}} = hamm(img^L_{\{i,j\}}, img^R_{\{i,k\}})$
  - $hamm(str^1, str^2) = \sum_0^{len(str^1)} eqs(str^1_i, str^2_i)$
  - $eqs(\alpha, \beta) = \begin{cases} 0 & \alpha \neq \beta \\ 1 & \alpha = \beta \end{cases}$

Disparity computation

- minimum distance (hamming distance) between correlations
- computed on the depth of the correlation matrix
- $j - v_{min}$[1] $> k >$ v_{max}[2]

$$\left[\left[\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{1,4} & x_{1,5} & x_{1,6} \\ x_{1,7} & x_{1,8} & x_{1,9} \end{bmatrix} \begin{matrix} X & X \\ & \\ X & X \\ X & X \end{matrix}\right] \left[\begin{bmatrix} x_{2,1} & x_{2,2} & x_{2,3} \\ x_{2,4} & x_{2,5} & x_{2,6} \\ x_{2,7} & x_{2,8} & x_{2,9} \end{bmatrix} \begin{matrix} X & X \\ & \\ X & X \\ X & X \end{matrix}\right]\right]$$

[1] Listing 1

[2] Listing 1

Disparity computation

- $disp_{\{i,j\}} = \min_{j-v_{min}>k>v_{max}} (\sum_{u=0}^{s} \sum_{v=0}^{s} corr_{\{i-u+\frac{s}{2}, j-v+\frac{s}{2}, k\}})$

Results

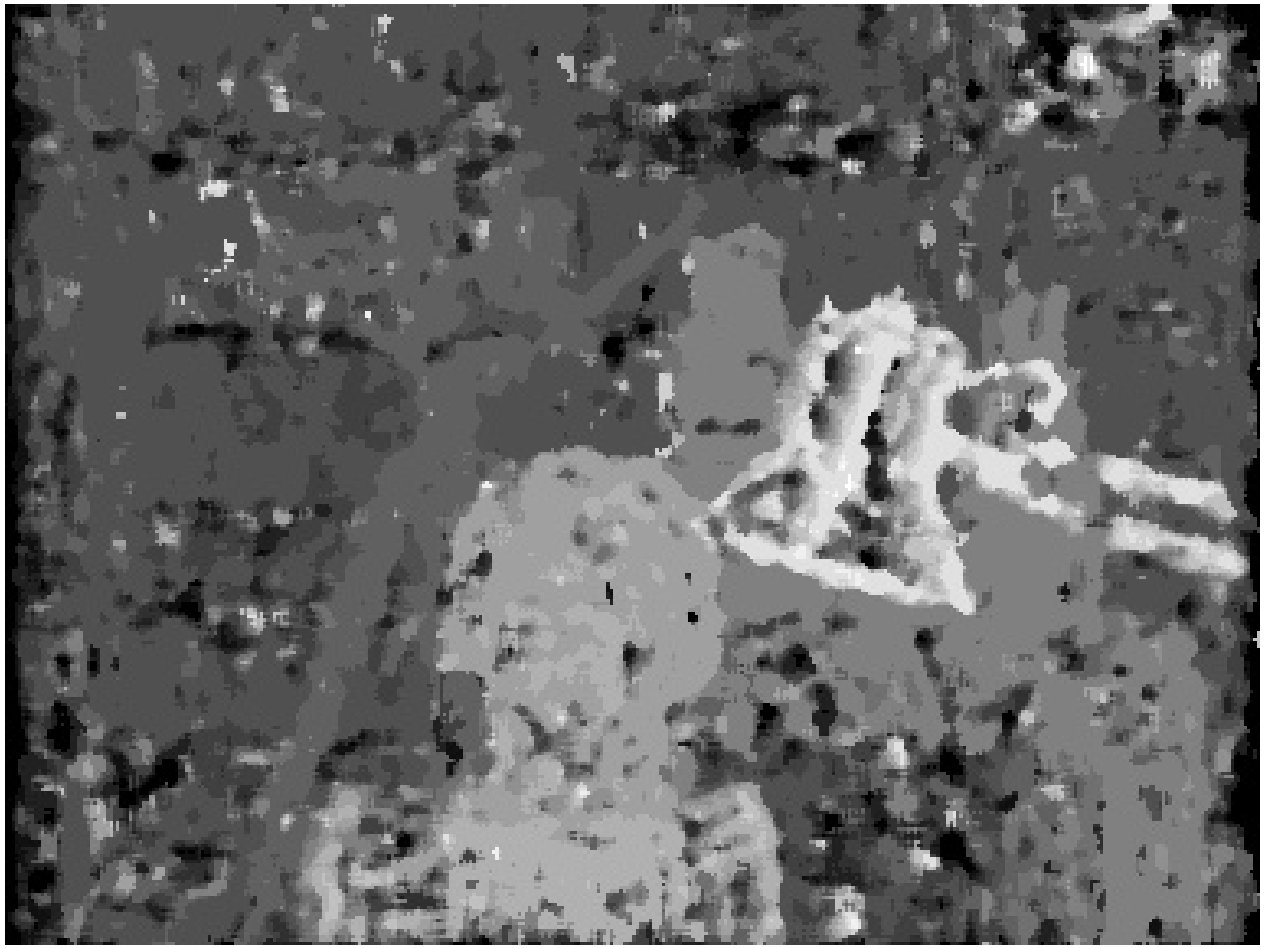

Figure: disparity matrix

Results

- In order validate results a ground truth is provided [Sch]
- The large disparity difference stems from the method

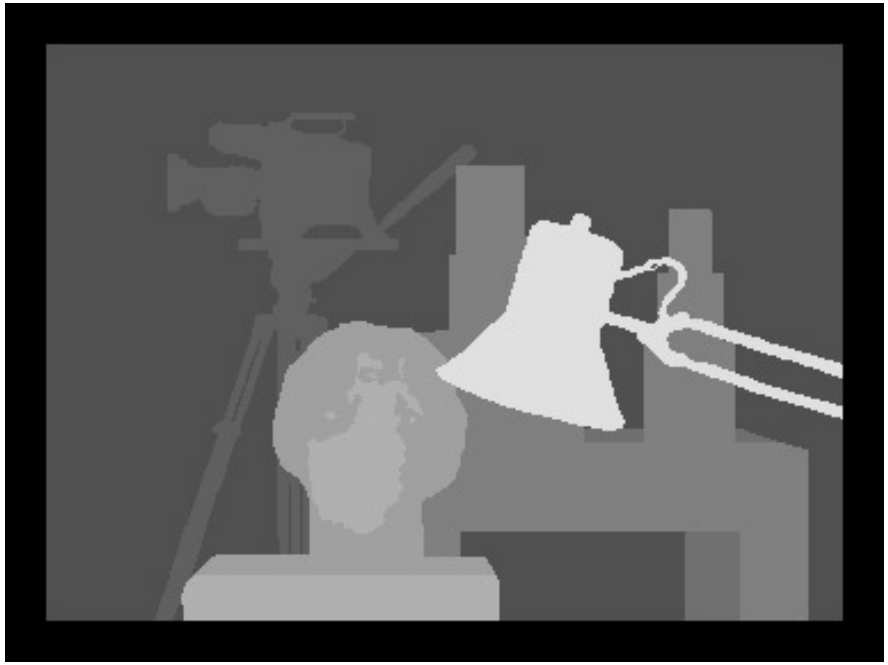

Figure: ground truth

Bibliography

[Zab05] Ramin Zabih. "Nonparametric Lo cal Transforms for Computing Visual Correspondence)". In: *LNCS* 801 (2005).

[Sch] Daniel Scharstein. *2001 Stereo datasets with ground truth*. URL: https://vision.middlebury.edu/stereo/data/scenes2001/. (accessed: 01-05-2023).