Structured Light Vision Systems Using a Robust Laser Stripe Segmentation Method

Zhankun Luo
supervisor: Prof. Lizhe Tan

Department of Electrical and Computer Engineering
Purdue University Northwest

April 28, 2021

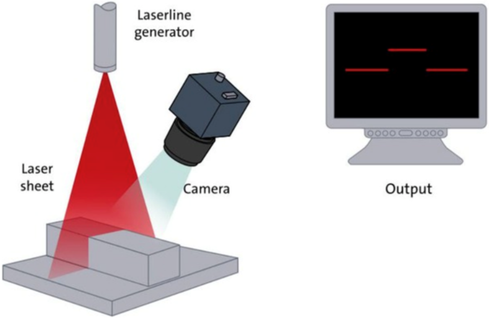
PURDUE UNIVERSITY® NORTHWEST™

Agenda

- Introduction
- Motivation & Objective
- Theory of multiple cameras
- Segmentation using neural networks
- Post processing
- Experiment result and comparison
- Conclusion

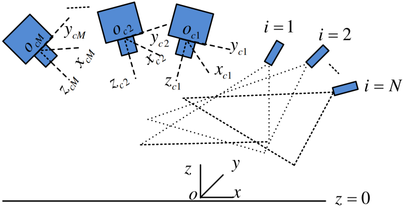
Introduction

- Configuration
  - Laser emitter
    structured light
  - Camera
    capture image
  - Computer
    post process

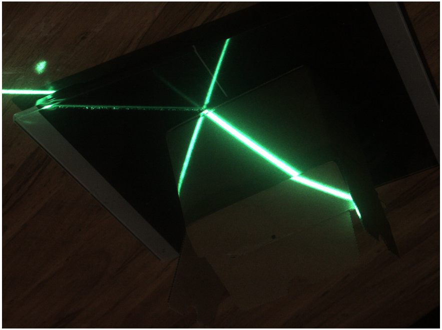


Motivation & Objective

- Motivation
  - Theory of multiple cameras
  - Segmentation using neural networks
- Objective
  - Height measurement
  - Eliminate reflected light noise in the background



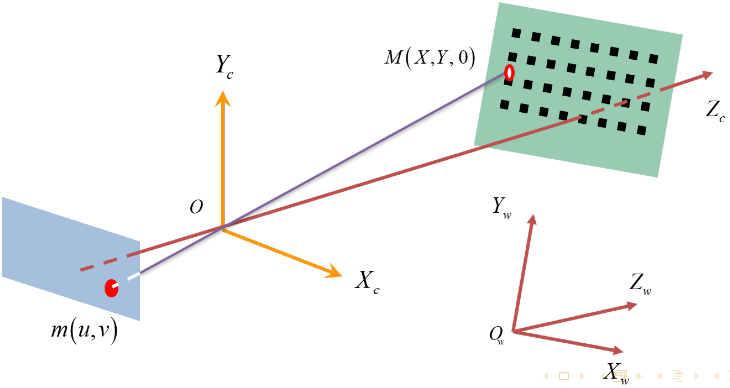
(a) The schematic of multiple shots, lasers



(b) Experiment configuration

Theory of multiple cameras

- Coordinate system
  - World system
    $M = [X, Y, Z]^T, \tilde{M} = [X, Y, Z, 1]^T$
  - Camera system
    $M_c = [X_c, Y_c, Z_c]^T$
  - Pixel system
    $m = [u, v]^T, \tilde{m} = [u, v, 1]^T$



Theory of multiple cameras

- step 1: Zhang's camera calibration
  - Basic equation
    $s\ \tilde{m} = A\ [R \quad t]\ \tilde{M}$
  - Known variables
    - pixel coordinates $\tilde{m} = [u, v, 1]^T$
    - world coordinates $\tilde{M} = [X, Y, 0, 1]^T$ on plane $Z = 0$
  - Other variable
    - depth to pinhole $s = Z_c$
  - Calibrated parameters
    - camera intrinsic matrix $A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$
      (u_0, v_0) the coordinates of the principal point
      α, β the scale factors in image u and v axes
      γ the parameter describing the skewness of the two image axes
    - extrinsic matrix $[R \quad t] = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \end{bmatrix}$
      R rotation matrix
      t translation

Theory of multiple cameras

- step 2: Laser plane calibration (stage 1)
  - Basic equations
    $s\ \tilde{m} = A\ M_c, \quad \pi_0^T \begin{bmatrix} M_c \\ 1 \end{bmatrix} = 0$
  - Known variables
    - pixel coordinates $\tilde{m} = [u, v, 1]^T$
    - camera intrinsic matrix A
    - checkerboard plane $(Z = 0)$ $\pi_0 = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}^{-T} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$

    $$Z = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tilde{M} = 0 \Longleftrightarrow \pi_0^T \begin{bmatrix} M_c \\ 1 \end{bmatrix} = 0$$
  - Calculated camera coordinates
    - depth to pinhole $s = Z_c$
    - camera coordinates $M_c = [X_c, Y_c, Z_c]^T$ for points on intersection of the checkerboard plane & the laser plane

Theory of multiple cameras

- step 2: Laser plane calibration (stage 2)
  - Basic equations
    $\pi^T \begin{bmatrix} M_c \\ 1 \end{bmatrix} = 0$
  - Known variables
    - camera coordinates $M_c = [X_c, Y_c, Z_c]^T$ for points on intersection of the checkerboard plane & the laser plane
  - Method
    - convert M_c for multiple camera systems into the same camera system
    - estimate π with the least-squares method
  - Calibrated parameters
    - laser plane $\pi = [a, b, c, -1]^T$

Theory of multiple cameras

- step 3: Height measurement
  - Basic equations
    $s\ \tilde{m}' = A\ M_c', \quad \pi^T \begin{bmatrix} M_c' \\ 1 \end{bmatrix} = 0, \quad M_c' = [R \quad t]\ \tilde{M}' = RM' + t$
  - Known variables
    - pixel coordinates $\tilde{m}' = [u', v', 1]^T$ for points on intersection of the object surface & the laser plane
    - laser plane $\pi = [a, b, c, -1]^T$
    - camera intrinsic matrix A
    - extrinsic matrix $[R \quad t]$
  - Calculated coordinates
    - camera coordinates $M_c' = [X_c', Y_c', Z_c']^T$ for points on intersection of the object surface & the laser plane
    - world coordinates $M' = [X', Y', Z']^T$ for points on intersection of the object surface & the laser plane
    - "Height" $H = \frac{1}{N} \sum Z'$ average for all the points on intersection of the object surface & the laser plane

- Segmentation using neural networks
  - Objective
    eliminate the noise caused by reflection and scattering of light
  - U-Net
    - encoder and decoder
    - usage: remove the background noises and keep laser stripes

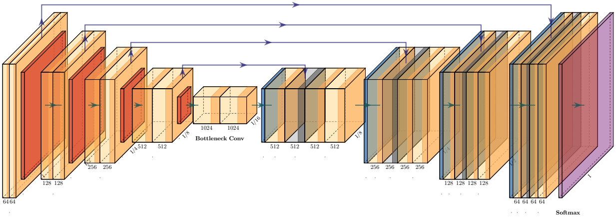
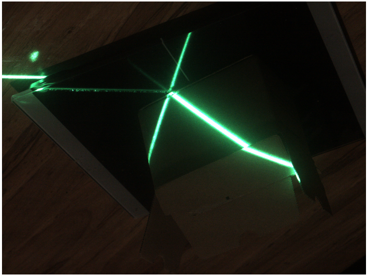


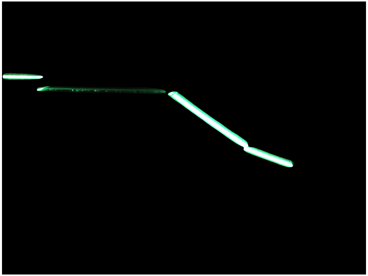
Figure: U-Net architecture

Segmentation using neural networks

- Segmentation using neural networks
  - Dataset
    - 500 images with the reflective light and the scattering light
    - 90% of dataset (450 images) as the training subset
    - 10% of dataset (50 images) as the validation subset during training



(a) Image with reflective noise



(b) Extracted stripe with U-Net

Segmentation using neural networks

- Segmentation using neural networks
  - Training procedure
    - learning rate scheduler: ReduceLROnPlateau strategy
      reduce the learning rate $\eta \leftarrow 0.1 \cdot \eta$ once learning stagnates
      (dice coefficient on validation subset stops increasing)
    - loss function: cross entropy
      $J(\boldsymbol{w}) \equiv -\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{y}_i^T \log\left(\hat{\boldsymbol{y}}_i\right) + (\vec{1} - \boldsymbol{y}_i)^T \log\left(\vec{1} - \hat{\boldsymbol{y}}_i\right)\right)$
      $\hat{\boldsymbol{y}}_i$ i-th predicted mask by U-Net, $\boldsymbol{y}_i$ i-th ground truth mask
    - optimizer: RMSprop
      with momentum to escape from the local minimum of neural networks

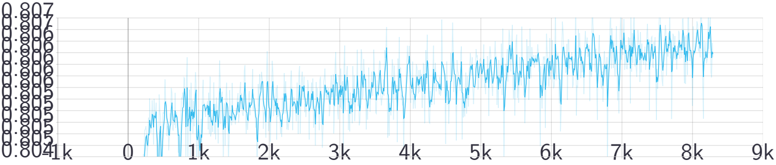
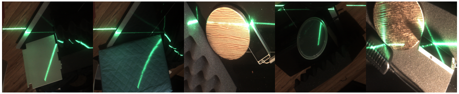


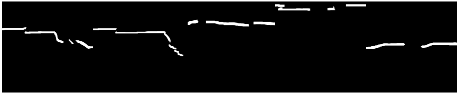
Figure: The dice coefficient on test dataset during training.

- Segmentation using neural networks
  - Results of U-Net
    - dice coefficient = 0.8108 Intersection over Union (IoU) = 0.6900



(a) The images used for training

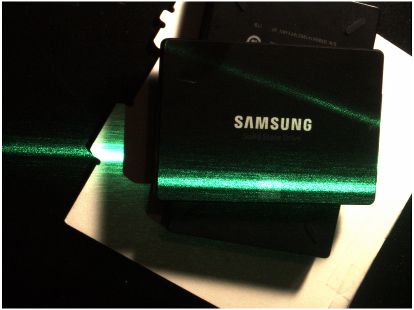
(b) The predicted masks



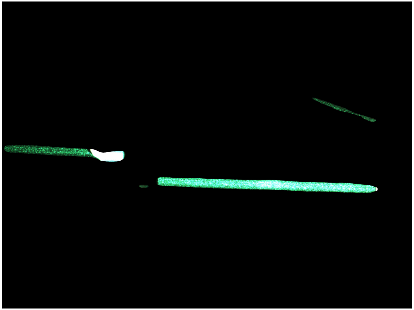
(c) The ground truth masks

Post processing

- Post processing
  - issue of segmentation
    - there could still be reflective noises in the background



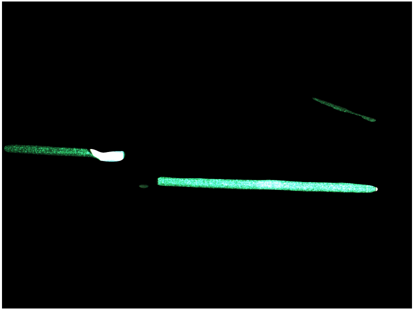
(a) The captured image



(b) The masked image

Post processing

- Post processing
  - Auto contrast enhancement
  - RGB to Grayscale
  - Erosion and dilation
    remove the small noise in the background

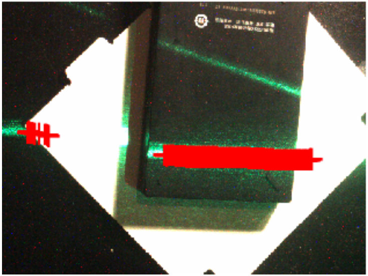


(a) The masked image

(b) The image with post processing

- Experiment result
  - without segmentation using neural networks
  - with segmentation using neural networks

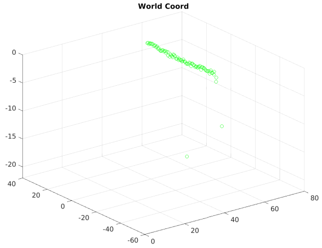
(a) Points extraction without U-Net segmentation method



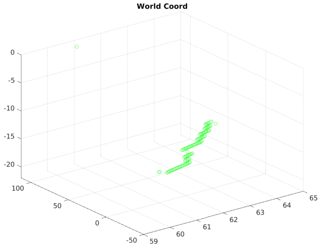
(b) Points extraction with U-Net segmentation method

Experiment result and evaluation

- Experiment result
  - without segmentation using neural networks
  - with segmentation using neural networks



(a) Cloud points without U-Net segmentation method



(b) Cloud points with U-Net segmentation method

Conclusion

- Conclusion
  - Framework for multiple shots and multiple laser emitters are derived
  - Laser stripe segmentation method based on U-Net is further developed to tackle the problem cased by the refection and scattering of light in complex environment

	Without segmentation	**With segmentation**
Simple operation	✗	✓
Luminance effect	✗	✓
Accurate	✓	✓

  - Our experiments demonstrate that the system with multiple cameras and U-Net laser stripe extraction method strengthens the stability of system