Computer Vision -- Lab 2

1. OBJECTIVES

This laboratory aims to provide further exposure to advanced image processing and computer vision techniques. In this laboratory you will:

(a) Explore different image segmentation methods.

(b) Experiment with pixel sum-of-squares difference (SSD) to find a template match within a larger image. Estimate disparity maps via SSD computation.

2. INTRODUCTION

2.1. Image Segmentation

Image segmentation aims to segment interested image regions from the image background. Different image segmentation approaches have been developed, including image thresholding that determines a local/global threshold for image background-foreground segmentation, filter bank that designs multiple filters for multi-class textured image segmentation (together with k-mean clustering), as well as deep neural networks that learn filters and classifiers for semantic image segmentation.

2.2. 3D Stereo Vision

The pixel intensity sum-of-squares difference (SSD) between a small image patch T and a large image I at the (x, y) location is given by

$$S(x,y) = \sum_{j=0}^{M} \sum_{k=0}^{N} \big(I(x+j, y+k) - T(j,k)\big)^2$$

By computing the SSD at *all locations* of the image I, we can generate a SSD image $S(x, y)$. This allows the location in the image that best matches the image patch T to be found.

Evaluating the above equation directly can be very time consuming. An alternative approach is to decompose the equation into 3 parts

$$S(x,y) = \sum_{j=0}^{M} \sum_{k=0}^{N} \big(I(x+j, y+k)\big)^2 + \sum_{j=0}^{M} \sum_{k=0}^{N} \big(T(j,k)\big)^2 - 2\sum_{j=0}^{M} \sum_{k=0}^{N} I(x+j, y+k) \cdot T(j,k)$$

The second term is constant with respect to different parts of the image, while the first and third terms can be expressed as convolution which can be efficiently computed via FFTs.

In stereo vision, the goal is to compute the relative depth of a 3D point from the stereo cameras. If rectified images (i.e. where projected 2D points from the same 3D point have the same y coordinate in different images -- they lie on the same scanline) are used, we can express depth in terms of disparity. Disparity is the difference between the x coordinate of the projections in two images and is inversely proportional to depth.

Given two images, stereo vision allows the computation of a *disparity map*. A disparity map is simply an array containing the disparities or depths that are associated with every pixel in one image (typically the left image).

3. EXPERIMENTS

3.1. Image Segmentation

This is a document image segmentation task targeting to write codes to segment the text from the image background. When document images suffer from different types of degradation as in the provided document examples (i.e., document01.bmp, document02.bmp, document03.bmp, and document04.bmp), thresholding becomes more challenging which often causes problems while performing optical character recognition.

a) For each provided document image, apply Otsu's global thresholding algorithm to perform global thresholding for text segmentation. Quantitatively evaluate the segmentation results by computing the sum of difference image between the segmented binary image and the corresponding ground truth (e.g., by sum(sum(difference image)) in Matlab). Analyse and discuss the obtained thresholding results as well.

b) For each provided document image, apply Niblack's local thresholding algorithm to perform local thresholding for text segmentation. Adjust the parameter in Niblack's thresholding algorithm to study how to obtain the best quantitative segmentation results. Analyse and discuss the obtained thresholding results as well.

c) Discuss and implement any possible improvements of results as obtained with Niblack's local thresholding algorithm. Evaluate, analyse and discuss the improvements.

3.2. 3D Stereo Vision

This is a substantial section as you will need to write codes to compute disparity *maps* for pairs of rectified stereo images P_l and P_r. The disparity map is inversely proportional to the depth map which gives the distance of points in the scene from the camera.

<u>Estimating Disparity Maps</u>

The overview of the algorithm is:

for *each* pixel in P_l,

i. Extract a template comprising the $11 * 11$ neighbourhood region around that pixel.

ii. Using the template, carry out SSD matching in P_r, but *only along the same scanline*. The disparity is given by

$$d(x_l, y_l) = x_l - \hat{x}_r$$

where x_l and y_l are the relevant pixel coordinates in P_l, and $\hat{x}_r$ is the SSD matched pixel's *x*-coordinate in P_r. You should constrain your horizontal search to small values of disparity (<15). If using Matlab, you may use functions *conv2*, *ones*, *rot90*, etc., to compute the SSD matching between the template and the input image.

iii. Input the disparity into the disparity map with the same P_l pixel coordinates.

a) Implement the disparity map algorithm by a function which takes two arguments of left and right images, and 2 arguments specifying the template dimensions. The function should return the disparity map. Try to minimize the use of for loops.

b) Download the synthetic stereo pair images of 'corridorl.jpg' and 'corridorr.jpg' and convert both to grayscale.

c) Run your algorithm on the two images to obtain a disparity map D, and check the results via

```
>> imshow(-D, [-15 15]);
```

The results should show the nearer points as bright and the further points as dark. The expected quality of the image should be similar to `corridor_disp.jpg' which you can view for reference.

Comment on how the quality of the disparities computed varies with the corresponding local image structure.

d) Rerun your algorithm on the real images of 'triclops-i2l.jpg' and triclops-i2r.jpg'. You may refer to 'triclops-id.jpg' for expected quality. How does the image structure of the stereo images affect the accuracy of the estimated disparities.