Exam for Computer Vision and Imaging (Extended)

ID 1803086

3rd June 2021

After inserting your student ID and the module title in the preamble, write your answers below.

1. (a) i.

$$\frac{1}{1e^{-2}} - \frac{1}{-1} = \frac{1}{f}$$

$$f = 101^{-1} = \frac{1}{101} = 0.\dot{0}09\dot{9}m$$

ii.

$$\frac{\frac{1}{101}}{1} = \frac{d}{1.6}$$

$$1.6 \times \frac{1}{101} = d$$

$$d = \frac{8}{505} = 0.0\dot{1}58\dot{4}m$$

(b) Max pooling

i. 2 × 2 kernel stride 1

$$\begin{bmatrix} 9 & 9 & 8 & 3 & 6 \\ 9 & 9 & 3 & 5 & 5 \\ 3 & 4 & 4 & 6 & 6 \\ 7 & 7 & 9 & 9 & 6 \\ 7 & 9 & 9 & 9 & 3 \end{bmatrix}$$

ii. 2 × 2 kernel stride 2

$$\begin{bmatrix} 9 & 8 & 6 \\ 3 & 4 & 6 \\ 7 & 9 & 3 \end{bmatrix}$$

iii. 3 × 3 kernel stride 1

$$\begin{bmatrix} 9 & 9 & 8 & 6 \\ 9 & 9 & 6 & 6 \\ 7 & 9 & 9 & 9 \\ 9 & 9 & 9 & 9 \end{bmatrix}$$

iv. 3 × 3 kernel size stride 3

$$\begin{bmatrix} 9 & 6 \\ 9 & 9 \end{bmatrix}$$

Average pooling

i. 2 × 2 kernel stride 1

$$\begin{bmatrix} 5 & 5.25 & 3.5 & 1.5 & 2.25 \\ 3.25 & 3.25 & 1.75 & 2.25 & 2.25 \\ 1.25 & 2 & 2 & 3 & 3 \\ 2.75 & 3 & 3.5 & 4.75 & 2.5 \\ 2.75 & 4.5 & 6.5 & 5.5 & 2 \end{bmatrix}$$

ii. 2 × 2 kernel stride 2

$$\begin{bmatrix} 5 & 3.5 & 2.25 \\ 1.25 & 2 & 3 \\ 2.75 & 6.5 & 2 \end{bmatrix}$$

iii. 3 × 3 kernel stride 1

$$\begin{bmatrix} 3.89 & 3.22 & 2.56 & 2.22 \\ 2.56 & 2.44 & 2.67 & 2.11 \\ 2.33 & 2.78 & 3.44 & 2.89 \\ 2.89 & 4.56 & 4.67 & 3.56 \end{bmatrix}$$

iv. 3 × 3 kernel size stride 3

$$\begin{bmatrix} 3.89 & 2.22 \\ 2.89 & 2.56 \end{bmatrix}$$

Min Pooling

i. 2 × 2 stride 1

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \end{bmatrix}$$

ii. 2 × 2 stride 2

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

iii. 3 × 3 stride 1

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

iv. 3×3 stride 3

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Comparison

The min and max pooling operations act to pick out the local minima and maxima in each *window* respectively. These can be good for reducing the size of the image but maintaining large features. The larger the kernel size relative to the image size and the higher the stride, the more information that is lost.

The average pooling can be viewed as a form of smoothing tool, similar to the mean filter used in image pre-processing. By amalgamating information from each pixel in the *window* less information is lost, but the image can lose focus, becoming more blurry.

All operations reduce the size of the input matrix, leading to these operations often being used for downsampling images.

(c) i. By projecting a pattern of known geometry, one can compare the position of each point in the projected point cloud with its expected position to map the surface of the object the cloud has been projected onto.

This is analogous to us using an image of a grid of known geometry to calibrate the intrinsic parameters of an unknown camera.

ii. The other methods of depth measuring we have covered include:

1. Time of flight
2. Structured light imaging
3. Stereophotogrammetry

And we can assume the following conditions were required by Apple to create a viable product:

1. Small form-factor
2. Accuracy - for security and ease of use reasons
3. Low/ fast compute time
4. Works in a variety of lighting scenarios (night/ day/ etc.)

Of these 4, all are covered to some extent by Structured light imaging (SLI) when compared with the other approaches covered.

The hardware required for SLI is more flexible than both of the other approaches, allowing for it to be more compactly constructed and embedded in a small form-factor device such as a smartphone. This approach requires a dot-matrix projection camera as well as a standard and, ideally, and infrared camera for capturing and mapping the point cloud.

The accuracy of SLI is high, working well with featureless surfaces (unlike stereophotogrammetry) and doing so with more accuracy than time of flight. It is particularly effective for smaller objects, such as, the human face.

The compute time of SLI is higher than time of flight, but lower than Stereophotogrammetry, along with the cheaper and smaller equipment this is a feasible tradeoff with mobile processing increasing hugely over the past few years.

Due to Apple's use of an infrared camera along with the point cloud being projected using light in this spectrum, the system works almost independently of external lighting factors, satisfying this condition.

(d) i. The random noise generated by the radiation could be similar to that of salt and pepper noise.

There are a number of noise removal methods, but the ones which work best with random noise include the median filter. The size for the kernel is related to the size of the sensor.

Other noise removal methods include the mean and max filters. The median filter outperforms these due to the fact it essentially attempts to interpolate pixel values based on the surrounding neighbourhood inside the kernel window.

ii. The Canny edge detector operates as follows:

- Filter image using derivative of Gaussian filter (DoG) filter.
  This stage attempts to smooth edges and suppress noise in the image, the reduce the number of false-positives in the final result
- Calculate the gradient magnitude and orientation
  In this stage useful information about the features of the image are calculated for use in latter steps to find edges more accurately.
- Apply non-maximal suppression
  Here we use the gradient magnitude and orientation calculated in the previous step in order to thin edges and remove possible false-positives. Edge thinning reduces the size of wide ridges to a single pixel in width.
- Apply Hysteresis thresholding
  In this form of thresholding, we utilise 2 thresholds: a higher and a lower.
  Initially we globally apply the higher threshold before attempting to *trace* possible edges by applying the lower threshold in areas surrounding strong edges, following these contours until values fall below the lower threshold.
  This form of thresholding accurately detects strong edges as well as weaker contours connected to these same edges.

It is useful to apply a Hough transform to the result of this as it produced concrete algebraic representations for the lines detected in the image. These can be further utilised when analysing the structure of the image.

iii. There are a number of challenges facing the robot when attempting a task such as this:

- Varying illumination, even within a single frame of reference.
- Different possible viewpoints leading to different relative poses/ shapes of the object.
- In a situation such as the one described, it is likely that there is a lot of debris surrounding the object, perhaps even entirely obscuring it.
- Intra-class variation

There are two main approaches to problems such as this: classical and *AI*. However, there are multiple pre-processing steps which can be applied in both cases.

Edge detection through the Canny detector described above as well as noise removal via median filtering can reduce the number of false-positives in our detection stage.

Image normalisation for lighting can also alleviate the issue of varying illumination.

Classical

In a classical approach we could utilise a process such as a distance transform using Chamfer distance. In this approach we could construct a mask or multiple masks of the object in question. Then for every n^{th} frame, the robot could construct a *distance transform* matrix onto which each mask could be applied until all masks have been tried or matching edges are detected in the frame.

By using multiple masks we could cover all possible variations of object, solving the intra-class variation concern.

If the object is detected, the operator can be alerted and they can respond.

AI

A deep learning approach such as a convolutional neural network could be a more robust approach especially when considering intra-class variation. The object's appearance could have been degraded by being at the scene of a nuclear meltdown so robustness in this area is very important.

By training on a large training set of images of the object and its possible variations our model can incorporate the pre-processing into the network, possibly leading to speedup as the CNN could learn more efficient methods for noise removal in the specific scenario.

Statement of good academic conduct
By submitting this assignment, I understand that I am agreeing to the following statement of good academic conduct.

- I confirm that this assignment is my own work and I have not worked with others in preparing this assignment.
- I confirm this assignment was written by me and is in my own words, except for any materials from published or other sources which are clearly indicated and acknowledged as such by appropriate referencing.
- I confirm that this work is not copied from any other person's work (published or unpublished), web site, book or other source, and has not previously been submitted for assessment either at the University of Birmingham or elsewhere.
- I confirm that I have not asked, or paid, others to prepare any part of this work for me.
- I confirm that I have read and understood the University regulations on plagiarism.