

# Exercise 1: Basic Image Processing

Due: 25.10.2018

Please note that Exercise 1 is long and may require all five sessions to complete. Don't leave it all to the last day and don't be afraid to ask for help from the assistants!

Please answer the theoretical parts of the exercise on this answer sheet and label it with your group number. In case of insufficient space, please use your own paper and label it accordingly.

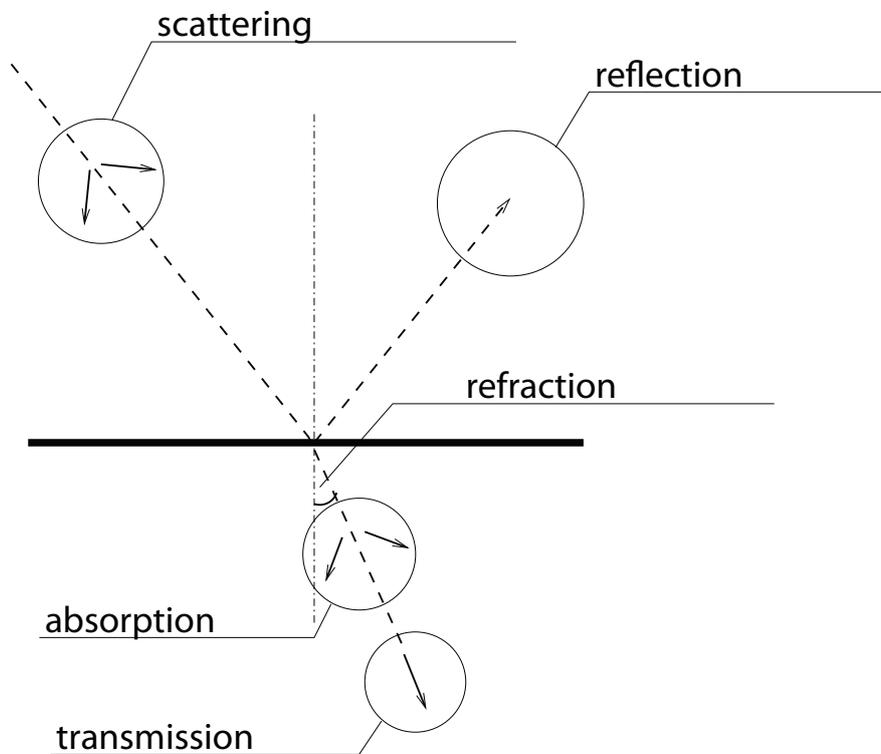
## 1 Image Acquisition

Group: .....

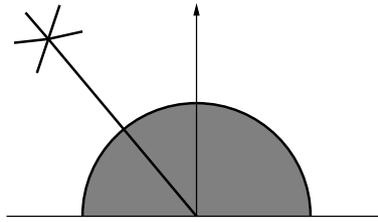
### 1.1 Theoretical Exercises

#### 1.1.1 Light and Matter

- a) Please label on the following illustration the name of the effects that can occur during interaction between light and a solid object.



- b) Complete the following figure such that it illustrates reflection on a Lambertian surface. Draw the direction(s) of the reflected light beam.

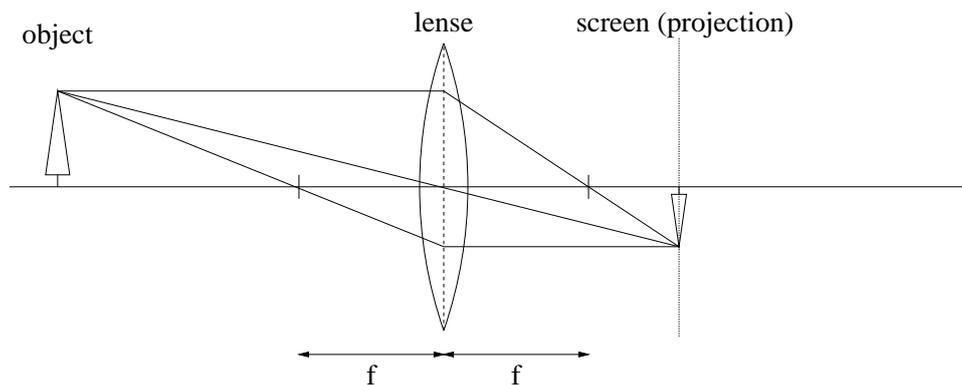


### 1.1.2 Image Acquisition / Camera Projection

#### 1.1.2.1 Acquisition of an image with a convex lens

You want to take a picture of an object with the following simple camera model (a lens of focal length  $f$  and a screen). Where should you place the screen in order to have a sharp image? How large will the object be on the screen?

Draw and label the position of the screen and the size of the object onto the sketch below:



#### 1.1.2.2 Minimal distance between object and screen

Using Gauss's formula (see script), compute the minimum distance between the object and the screen for the above camera setup. Assume that the object is in focus and that the focal length is constant. To simplify the calculation, you may introduce positive variables for both the object distance  $g = Z_0$  and the image distance  $b = -Z_i$ . Hence, you have to minimize  $g + b$ .

Unknown is  $\min(x)$ , with  $x = g + b$ , whereas  $b$  and  $g$  are linked (assuming  $f$  to be constant) by

$$\frac{1}{f} = \frac{1}{b} + \frac{1}{g} \quad \rightarrow \quad b = \frac{gf}{g-f}$$

With  $x = g + b = g + gf/(g-f)$ , the condition for an extreme value is:

$$\frac{\partial x}{\partial g} = 1 + \frac{f \cdot (g-f) - 1 \cdot gf}{(g-f)^2} = \frac{g^2 - 2gf}{(g-f)^2} = 0$$

The two possible solutions are  $g = 0$  or  $g = 2f$ . The first one ( $g = 0$ ) does not make any sense in a physical point of view.

The solution  $g = 2f$  is a real *minimum* of  $x$  as  $x = \infty$  for  $g = f$  or  $g = \infty$  and there are no further extreme values of  $x$  between these two values of  $g$ . Hence, out of  $g = 2f$  we get  $b = 2f$  and finally  $x = b + g = 4f$ .

### 1.1.2.3 Camera Projection

You have a (pinhole-) camera, for which the internal calibration parameters are already known. Now you take a picture with this camera, in which you know the exact position of  $N$  points in the “World Coordinate System”.

How large must  $N$  be to determine all *external* calibration parameters?

Try to answer the question in two ways:

- a) Consider how many unknowns this problem has, and how many equations you obtain from each point. The problem is solved if there are at least as many equations as there are unknowns.

The external calibration parameters are given by the position vector  $\mathbf{C}$  and the  $3 \times 3$ -rotation matrix  $R \in SO(3)$ .  $\mathbf{C}$  has 3 unknown quantities. From the matrix  $\mathbf{R}$  we also get 3 unknown quantities, since the rotation can be expressed with three angles (the Euler angles). An alternative explanation can be given as follows: The matrix has 9 elements in total – but the column vectors  $\mathbf{r}_i$  are of unit length (-3 unknowns) and pairwise orthogonal to each other (-3 again), which leaves us with 3 unknowns.

In total we get **6** unknowns from  $\mathbf{C}$  and  $\mathbf{R}$ , which can be determined with **3** points, because each point contributes 2 equations.

- b) Consider which degrees of freedom are lost for each additionally considered point. If no more degrees of freedom exist, i.e. the location and orientation of the camera are fixed, then the camera is externally calibrated.

The camera has 6 degrees of freedom. Point by point they are fixed as follows:

**1. Point:** Without loss of generality, we suppose the first point is lying on the optical axis. The camera can now be translated along the optical axis and rotated freely around the first point. This means we fixed two degrees of freedom, namely the translations perpendicular to the optical axis.

**2. Point:** After fixing a second point, the camera can still move around the connecting axis between the two points. It can also move in the plane which is spanned by the axis between the two points and the optical axis. These movements can not be expressed by elementary translation or rotation. Rather they are a combination of those: a translation along the optical axis, and a rotation around an axis, which is perpendicular to the optical axis and the axis between the two fixed points and is going through the point of intersection of those axes. This means, that the second point eliminated one degree of freedom for the rotation and fixes a dependence between the remaining rotation and translation.

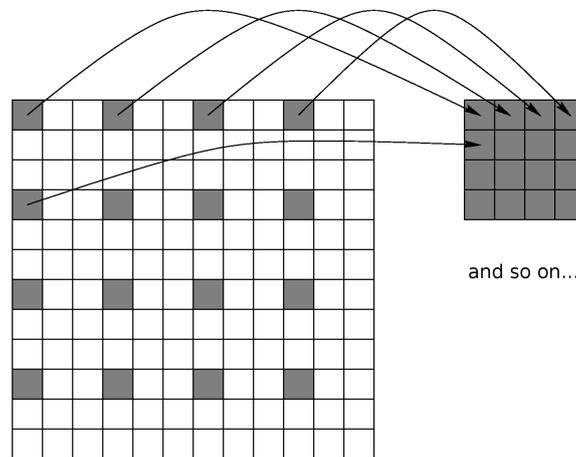
**3. Point:** No movements of the camera possible any more. The camera is fully externally calibrated.

## 1.2 Practical Exercise: Sampling

This first programming exercise addresses the acquisition of digital, discretely sampled images and helps you to understand how to access individual pixels in a given input image by treating spatial sampling techniques. To complete this exercise, you may use either the Python script template `1_sample.py` or the Jupyter notebook template `1_sample.ipynb`, all located in the folder `~cvcourse/material/`. Just copy the templates into your local folder by typing:

```
cp ~cvcourse/material/1.* . (do not forget the point at the end)
```

Digital images are sampled representations of continuous space. Each sample in the image is a discrete entity known as a pixel. This part of the exercise explores the properties of spatial sampling by considering an already sampled input image and resampling it with a lower sampling rate, which is known as *downsampling* or *decimation*. **Your first programming task** is to implement a function that extracts every  $n$ -th pixel from a given input image (where  $n$  is an integer) both in the horizontal and vertical direction to generate an output downsampled image for which the total number of pixels is reduced by a factor of approximately  $n^2$ . Parameter  $n$  is known as the *decimation factor*. For  $n = 3$ , the following output image (right part) will result from the input image (left part).



Using your function, experiment with various decimation factors  $n$  on different images in the directory `~cvcourse/pics`.

1. What happens if you resample an input image containing fine regular structures (e.g. `~cvcourse/pics/carpet.png`) with a much lower sampling rate, i.e., using very large  $n$ ?

If the sampling rate is less than twice the spatial frequency of the fine regular structures, then aliasing will occur.

2. What is aliasing and why does it occur?

Aliasing is the distortion of sampled images as a result of insufficiently high sampling rates.

3. The higher the decimation factor  $n$ , the lower the sampling rate. Without modifying the input image, what is the highest decimation factor  $n_{max}$  that one can apply to a given image without any information loss incurred?

To avoid information loss, the frequency corresponding to  $n_{max}$  must be at least twice as much as the highest spatial frequency that is present in the input image. This is also known as the Nyquist frequency.

A standard *anti-aliasing* step is to *pre-filter* the input image with a low-pass filter, in order to suppress its high-frequency content, before downsampling it. In this way, it is possible to use quite large decimation factors  $n$  and still obtain a meaningful output. **Your second programming task** is to implement a function that filters the input image with a Gaussian filter, which is a good approximation of an ideal low-pass filter. Selecting a proper built-in Python function to achieve this

is highly recommended. After that, experiment with large decimation factors  $n$  for downsampling the filtered image and compare to the case of no pre-filtering.

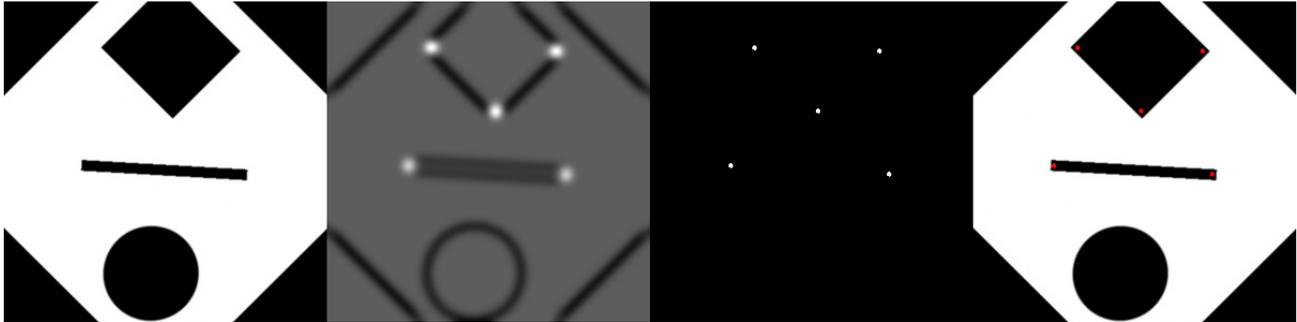
4. How is the scale parameter  $\sigma$  of the Gaussian filter related to the bandwidth of an ideal low-pass filter?

Gaussian blurring with a larger  $\sigma$  is similar to low-pass filtering with a smaller bandwidth (i.e., the two are inversely proportional). Note, however, that high-frequency information can be recovered after Gaussian blurring, but not after ideal low-pass filtering.

## 2 Harris Corner Detection

Visual features are attributes or characteristics of an image. Feature extraction is a common step used in many computer vision routines such as object detection, stereo correspondence, motion extraction, etc. In this exercise, you will explore two types of low-level feature points: **corners** and **edges**.

The Harris corner detector finds corners in an image. Conceptually, corner detection can be thought of as an auto-correlation of an image patch. Consider a window which slides over an image patch. If the image patch is constant or “flat”, then there will be little to no intensity changes in the window. If the image patch has an edge, then there will be no intensity changes in the window along the direction of the edge. If a corner is present, however, then there will be a strong intensity change in the window regardless of the direction.



### 2.1 Theoretical Exercises

- Is the Harris corner detector robust with respect to intensity changes in the image? Why or why not?

The Harris corner detector uses image gradients, and as such, is robust with respect to intensity *shifts* (i.e.  $I + b$ ) but not to a intensity *scalings* (i.e.  $aI$ ).

- Is the Harris corner detector robust with respect to rotation? Why or why not?

The Harris corner detector is rotation invariant. The eigenvalue decomposition of the second-order moment matrix will change in direction when the image is rotated (i.e. the eigenvectors indicating the direction of the largest intensity changes) but the eigenvalues will remain the same.

### 2.2 Practical Exercise

Implement a Harris corner detector. Code templates are provided in Python (`1_harris.py`) and Jupyter (`1_harris.ipynb`). Test your implementation on `~cvcourse/pics/CircleLineRect.png` and `~cvcourse/pics/zurlim.png`.

- **Derivatives:** Approximate the image gradients  $f_x$  and  $f_y$  by using either a Sobel operator or the first derivative of a Gaussian combined with smoothing.
- **Second order moments:** Use the gradients computed above to determine the second-order moments  $f_x^2$ ,  $f_y^2$  and  $f_x * f_y$ .
- **Harris response:** Smooth your second order moments (i.e. with Gaussian smoothing) and compute the determinant, trace and finally the Harris response. Write a function to extract the local maxima of the Harris response (see code template) and to visualize your detected corners.
- **Rotation and scaling invariance:** Try different values for the `rotate` parameter in the `main` function provided in the code template to rotate the input image. Apply the Harris corner detector again and observe. Is the Harris detector invariant for scale changes?

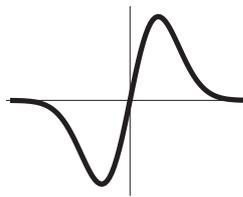
### 3 Canny Edge Detection

The Canny filter determines edges based on searching for local minima or maxima in the approximated first-order derivative of a signal.

#### 3.1 Theoretical Exercises

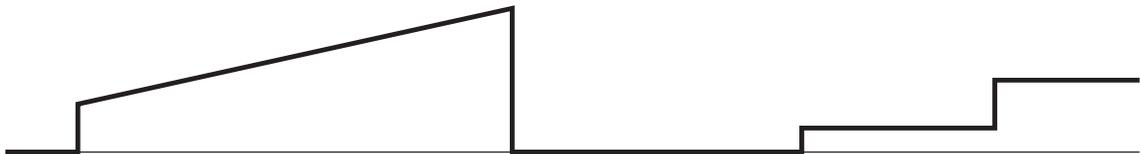
The final version of the Canny filter is as follows:

$$\frac{d}{dx}G(x) = \frac{d}{dx} \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \right)$$

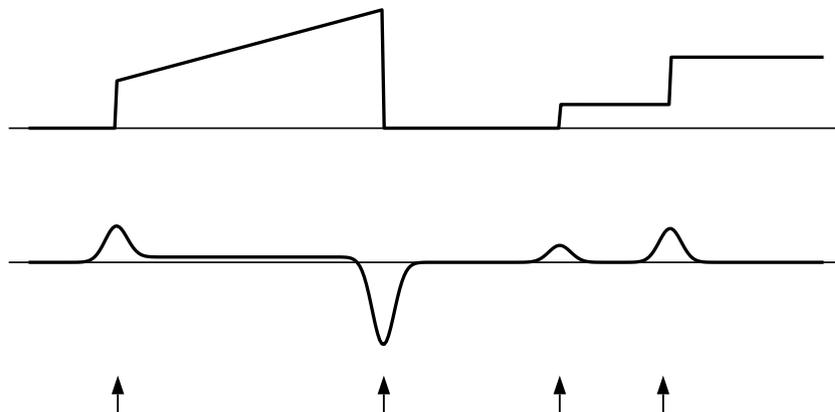


mirrored  $\frac{d}{dx}G$ -Filter

- a) Apply the Canny-Filter to the following profile in order to detect the edges. What is the magnitude of the filter response and the polarity at the edges? (draw a sketch).



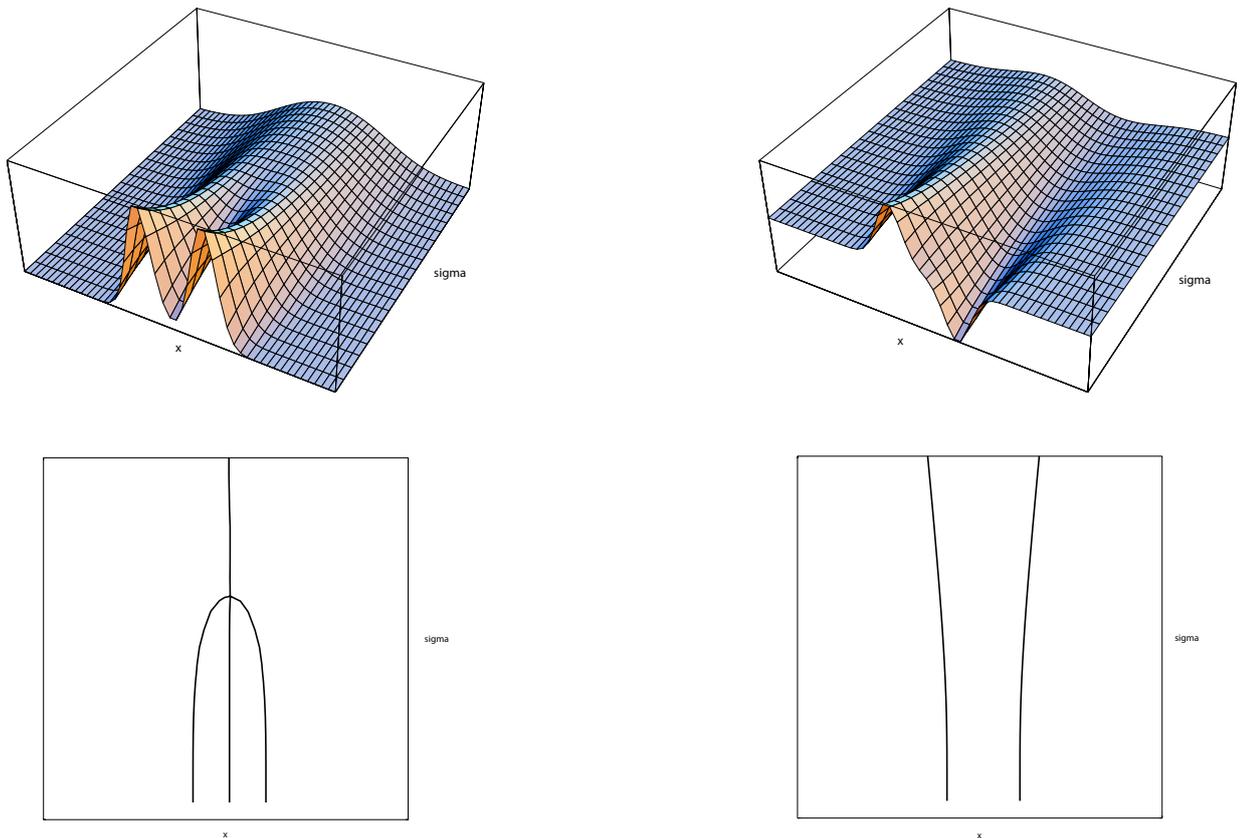
The strength of the filter response is proportional to the height of the edge. The polarity of the edge determines the sign. The filter produces a non-zero value for a slope, which can lead to false edges in the presence of noise. The edges can be identified as maxima in the absolute value of the filter response (marked with arrows).



- b) Apply the Canny-Filter with non-maximum suppression for different values for  $\sigma$  (width of the filter). For the following profiles, draw the positions of the detected edges as a function of  $\sigma$ .



The following figures show the filter response as a function of position and filter width (sigma). The detected edges can be localized as local maxima of the absolute value of the filter response:



The left image shows that for a small sigma two independent edges are detected. For a larger sigma they merge into a single edge.

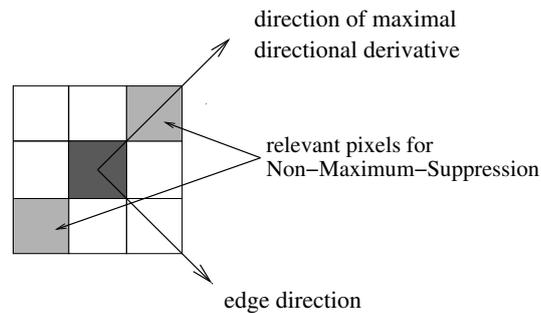
The right image shows how the positions of the edges are detected correctly for small values of sigma. For increasing sigma, the two responses start overlapping and the maxima of the response are not at the accurate edge positions anymore.

- c) Is it possible to detect lines (  ) and edges (  ) of different width and shape with a linear filter? Justify your answer.

It is not possible to recognize both lines and edges with a linear filter. Applying an asymmetric filter (e.g. Canny filter) will result in a zero value for the center of symmetric structures. A symmetric filter on the other hand is well suited to detect lines but can not detect the correct position of step edges.

### 3.2 Practical Exercise

Having seen how the Canny filter works in the 1-D case in the above theoretical exercises, we now move onto the 2-D case. In 2-D, the gradient magnitude is a measure for the strength of an edge, while direction of the gradient vector gives edge orientation. Edge detection can be done by thresholding the gradient magnitude, though this generally results in thick contours which are poorly localized. To thin the contours into single-pixel-wide lines, one should use non-maximum suppression. A pixel in the contour is kept as edge pixel only if its gradient magnitude is greater than that of its two neighboring pixels in the direction orthogonal to the edge, i.e. where the gradient attains a maximum. Code templates are provided in Python (`1_canny.py`) and Jupyter (`1_canny.ipynb`).



Implement this edge detection approach in 2-D. You should produce the following images:

- **Gaussian Smooth Image:** Use the code provided in the templates.
- **Gradient Image:** Normalize the gradient values in order to cover the full range of grey values.
- **Orientation Map:** At each pixel, the orientation of an edge can be obtained from the gradient. Note that edge directions  $\alpha$  and  $\alpha + \pi$  are equivalent. The resulting directions can be coded as grey values.
- **Edge image according to the threshold criteria:** Each point whose absolute value of the gradient reaches a given threshold will be displayed.
- **Edge image after the Non-Maximum-Suppression:** Each point, which is fulfilling the threshold criteria and has a larger edge strength than its direct neighboring pixels in the direction orthogonal to the edge, will be displayed. Note that this is not the same as the non-maximum suppression in Harris corner detection!

Extract edges from the image `~cvcourse/pics/zurlim.png` with your implementation. Appropriate parameters are suggested in the code template. In addition, apply your implementation to the image of the two cuboids.