1 Theoretical Exercises

Please answer the theoretical part of the exercise on this answer sheet and label it with your group number. Submit only one solution per group by 26.10.2017. In case of insufficient space, please use your own paper and label it accordingly with your group number and task number.

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.

1.2 Image Acquisition / Camera Projection

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:

1.2.2 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 $x = g + b = g + \frac{gf}{g-f}$, the condition for an extreme value is:

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

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 $x = b + g = 4f$. 

acquisition-solution.tex
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 \( \mathbf{R} \in SO(3) \). \( \mathbf{C} \) 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.
2 Practical Exercise

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. Part 1 treats spatial sampling techniques, while part 2 considers pixel intensity values. To complete this exercise, you may use either the Python script templates `sample.py` and `requant.py` or the Jupyter notebook templates `sample.ipynb` and `requant.ipynb`, all located in the folder `~/cvcourse/materials/`. Just copy the templates into your local folder by typing:

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

2.1 Sampling

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 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).

![Output Image](image)

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 task** is to implement a function that filters the input image with a Gaussian filter, which is a good approximation of a 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.
2.2 Modification of the pixel intensities

In this part of the exercise, you will manipulate the actual pixel values. For example, if you take a look at the provided stereo images `cvcourse/pics/cube_left.pgm` and `cvcourse/pics/cube_right.pgm`, you will see that image acquisition was not optimized and both images are underexposed. Furthermore, the acquisition hardware used, i.e. the camera, performs a non-linear mapping from the captured irradiance values (light coming from the scene) to resulting pixel intensities. In the following step you will learn how to adjust the range of pixel values.

Requantization of the pixel intensities: Digital images are discrete not only in space but also in intensity. The process of mapping a continuous intensity range to a finite number of intensity levels is called quantization. Your task is to complete the provided code skeleton in order to uniformly requantize an input image down to a given number $n_g$ of gray levels. After requantization, the $n_g$ gray levels should span the entire intensity range $[0, 255]$ of the output image that corresponds to 8 bits per pixel; note that this is not the same as having 256 gray levels.

Experiment with different quantization levels from 8 bits ($2^8 = 256$ gray levels) down to 1 bit (2 gray levels) and observe the output images in comparison to the corresponding input images.

1. As the number of quantization levels decreases, the quality of the image degrades. Why would one want to use fewer quantization levels then?
   
   For compression purposes, one may need to use fewer quantization levels.

2. Is it possible to decrease the error of quantization—measured as the mean of some norm of the difference between input and output—while keeping the same number of quantization levels?
   
   In principle, one can further decrease the error for a given formula of it by choosing non-uniform quantization levels based on images statistics (see Lloyd-Max quantization in script).