

---

Robot Localization

MASTER SEMESTER PROJECT LCAV

Frederike Dümbgen

frederike.duembgen@epfl.ch

R. Scheibler, M. Vetterli



December 30, 2015

---

# **1 Introduction**

## **1.1 Acoustic-based SLAM**

Introduction to algorithm to be implemented

## **1.2 Experimental setup**

Overview experimental setup

## 2 Visual Localisation

For visual localisation, a set of 4 cameras are used. the camera's parameters such as their focal width and height and distortion factors (*intrinsic parameters*) are determined using an algorithm provided by OPENCV (2.1). Thanks to color filtering and some basic shape recognition, a set of feature or reference points are localized in the images (2.2). Combining these image positions with the knowledge of the exact positions in the object space, the 3D camera positions (or *extrinsic parameters*) are obtained (2.3). Knowing the intrinsic and extrinsic parameters of the camera, any point can be located given its image. A higher robustness is achieved when the point is triangulated using its image position in more than one cameras (2.4).

### 2.1 Intrinsic Calibration

Every camera is characterised by what is called its *intrinsic parameters*. They consist of the focal width and height,  $f_x$  and  $f_y$ , the principal point  $c_x, c_y$  which is usually at the image center. These parameters are grouped in the camera matrix  $C$ :

$$C = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Furthermore, every camera creates some radial and tangential distortion in the images which needs to be taken into account. If  $x'$  and  $y'$  are the undistorted image points normalized with  $z$  such that  $z' = 1$  and centered around the principal point  $c_x$  and  $c_y$ , and  $r^2 = x'^2 + y'^2$ , then the distorted points  $x''$  and  $y''$  can be modelled using Brown-Conrady's decentering distortion model [2], [3].

$$x'' = x'k(r) + (2p_1x'y' + p_2(r^2 + 2x'^2))(1 + p_3r^2 + p_4r^4 + \dots) \quad (2)$$

$$y'' = y'k(r) + (p_1(r^2 + 2y'^2) + 2p_2x'y')(1 + p_3r^2 + p_4r^4 + \dots) \quad (3)$$

$$\text{with } k(r) = 1 + k_1r^2 + k_2r^4 + \dots \quad (4)$$

The distortion coefficients are indifferent to the camera resolution [7], only the focal width and height and the principal point have to be scaled appropriately. To reduce imprecisions due to scaling, a recalibration was done every time the camera resolution was changed.

An algorithm for finding the intrinsic parameters including the radial distortion coefficients was suggested by Zhang [10]. Its implementation in the Matlab Camera Calibration Toolbox added an estimation algorithm for two tangential distortion coefficients. [1] The OpenCV implementation used in this project is based on the Matlab Toolbox. It is implemented for the first 6 radial distortion coefficients  $r_i$  and the first 2 tangential distortion coefficients  $p_i$  [7].

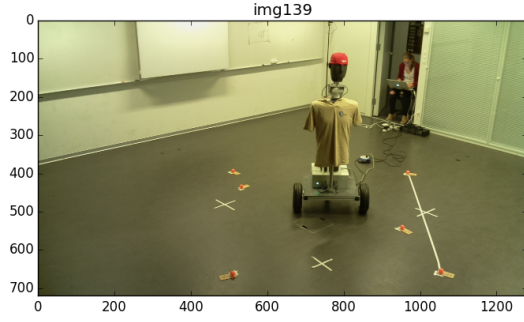
### 2.2 Image Processing

The implemented procedure to detect the image of the robot head and the reference points is based on bright colored, circular reference points and is semi-automatic.

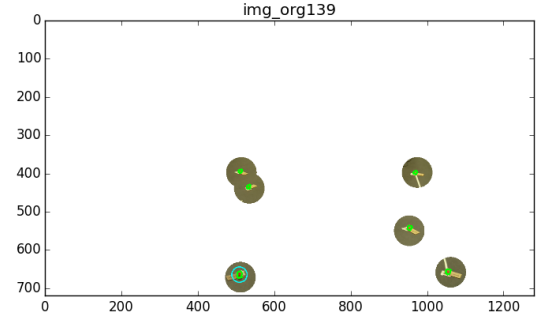
Examining the picture's HSV representation (Figures 1c and 1d), one can see that a combination of the H and S values gives a good criteria for the extraction of the bright colored points. The user defines the regions of interest and the relative position of the reference points by clicking in order on these points in the order of their numbering. The colors in these regions of interest are extracted in two steps: first a rough filter is applied to the image which filters out any pixels that lie outside of a certain color range and sets their values to 0 (Figures. The color distribution of the left over pixels is then characterized by its mean  $\mu$  and standard deviation  $\sigma$  and only colors lying above and below a certain threshold are preserved. Formally the criterion for pixels that are kept is:

$$\frac{I(x, y) - \mu}{\sigma} \leq z \quad (5)$$

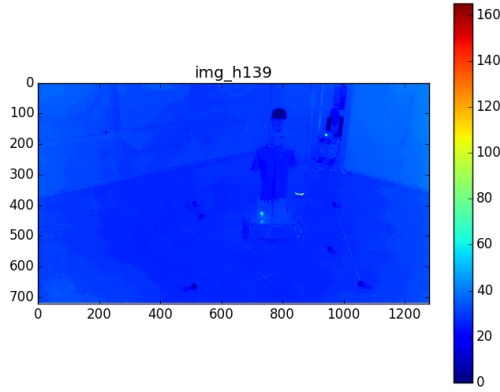
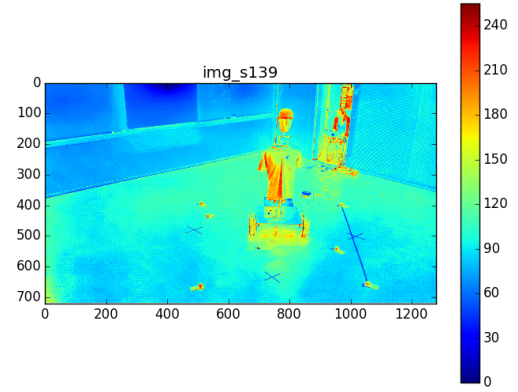
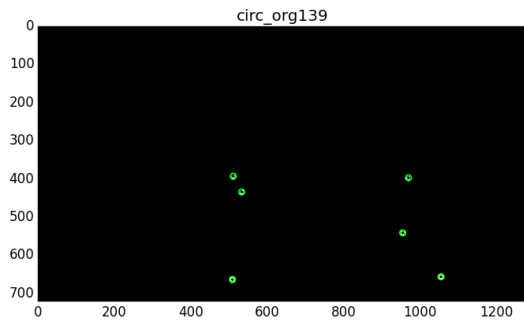
The threshold  $z$  is empirically chosen to be 2, which, if the colorspectrum had the form of a gaussian, would correspond to keeping around 99% of the colors. The contours of the resulting binary image undergo two tests. First, the area within the contour has to lie above a empirically found minimum for the contour to be further treated. The resulting contours are shown and the regions of interest are shown in Figure 1b. Secondly, the circular shape is tested by placing a circular filter of an approximate radius onto the contours and counting the white pixels that lie inside the circle and outside the circle respectively (true positives and false positives). The proportion of true positives has to lie above a certain threshold  $t$  for the contour to be considered valid. The resulting binary image is composed of at least  $N_{pts}$  contours whose centroids can be obtained by calculating the respective moments  $M_{i,j} = \sum_x \sum_y x^i y^j I(x, y)$  ( $x$  and  $y$  correspond to pixel coordinates and  $I(x, y)$  is the pixel intensity). The centroid coordinates are given by  $\bar{x} = \frac{M_{10}}{M_{00}}$  and  $\bar{y} = \frac{M_{01}}{M_{00}}$ . If two centroids are too close to each other, they are considered a duplicate and replaced by their middle. The resulting binary images with the final centroids for each point for reference points and the robot head are shown in Figures 1e and 1f respectively.



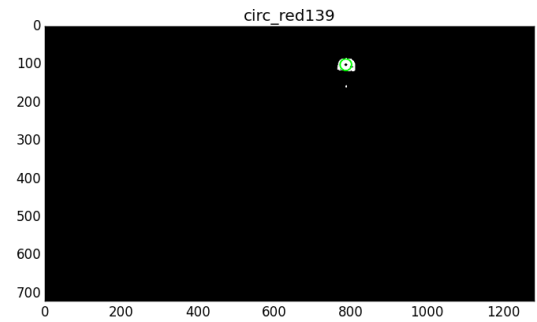
(a) original image



(b) Regions of interest and extracted colors

(c) *Hue* representation of original image(d) *Saturation* representation of original image

(e) Extracted colours, reference points



(f) Extracted colours, robot head

Figure 1: Procedure and tools for feature extraction

## 2.3 Extrinsic Calibration

### 2.3.1 Reference points

**Points** In a first run, 4 to 6 bright orange circular reference points were used for the extrinsic calibration. The points are numbered and the first 2 points serve as the basis for the reference frame. (see Figure 2a). Since the distances between all points can be very accurately measured using a laser

pointer, the euclidean distance matrix is set up and the absolute positions in the reference frame are obtained using the classical Multidimensional Scaling (MDS) method. [8]

The limitations of this method are little number of reference points that can only hardly be extended since all points need to be numbered manually by the user and the number of laser pointer measurements increases with the square of the used number of reference points. Secondly, the radius of the reference points needs to be chosen big enough such that it can be robustly detected from a changing distance. But the bigger the radius, the bigger is the imprecision induced since the points can not be exactly on the ground and are perceived at different heights from different camera angles. For these reasons, a checkerboard approach was considered.

**Checkerboard** A second and more user friendly approach using a checkerboard for reference points was implemented in a second run. Since many robust implementations exist for checkerboard corner detection, its use promises a big number of reference points with minimal effort and high reliability. In addition to that, the checkerboard corners are of much smaller radius than circular reference points and can be placed exactly on the ground.

Particular attention is required for the numbering of the reference points. Since it shall be avoided that the significant number of reference points needs to be numbered manually, an automatic numbering based on three reference points placed in the corners of the checkerboard is defined. The three reference points are detected manually and the numbering starts at the chessboard corner closest to point 1, goes on in direction of point 2 and then up in direction of point 3 (see Figure 2b).

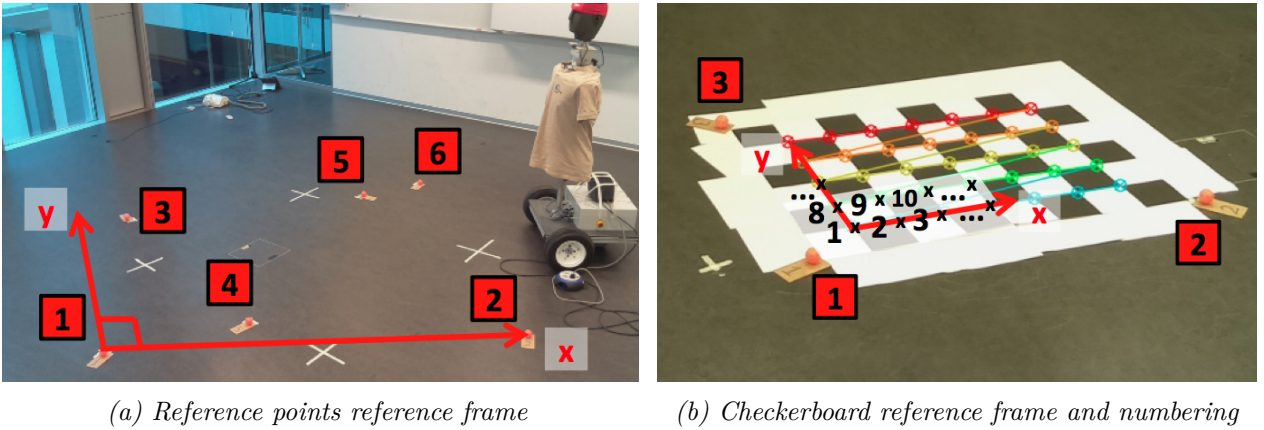


Figure 2: Reference point and checkerboard layout conventions

**Algorithm** The reference points are detected as explained in 2.2. The so-found image-object correspondances can be used to determine the camera position and orientation by solving the following system of equations for  $R$  and  $t$ .

$$\mathbf{x}_i = \text{proj}(\mathbf{X}_i, \mathbf{P}) = \mathbf{C} \quad \mathbf{P} \quad \mathbf{X}_i \quad (6)$$

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{R} & | & \mathbf{t} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad \text{for } i = 1 \dots N_{pts} \quad (7)$$

Where  $N_{pts}$  is the number of points (4 to 6 or number of checkerboard corners),  $\begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T$  are the homogenous image coordinates with scaling factor  $s$  and  $\begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T$  are the homogenous object point coordinates.  $C$  is the intrinsic camera matrix, determined as explained in 2.1.

The extrinsic camera matrix  $\mathbf{P}$  that one needs to solve for can be decomposed in a rotation and a

translation component:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad \text{Camera rotation matrix} \quad (8)$$

$$\mathbf{t} = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^T \quad \text{Camera translation vector} \quad (9)$$

There are 12 unknowns and each new point provides 3 independent equations. Therefore at least 4 points are required for solving this set of equations.

Various methods have been proposed for solving this kind of problem, also known as Perspective- $n$ -Point or PnP problems. The methods can be embedded in a Ransac scheme, which makes them more resistant to outliers. However, outliers will not occur in the present experimental setup because of its deterministic nature: all reference points are precisely defined and need to be detected, as opposed to setups where a undefined amount of feature points are extracted from images.

Three different methods for solving this PnP problem are implemented in OpenCV. *P3P* is based on a technique that is limited to 4 points only, so it was immediately rejected. The *EPNP* method [6] provides a non-iterative solution to the problem which is more stable and computationally inexpensive. Since in the present case, the number of points is limited to 40 and the calculation time turned out to be acceptable, the method chosen is the *ITERATIVE* method, based on reprojection error minimization.

: The reprojection error is the sum of the squared distances between observed projections  $\mathbf{x}_i = [u_i, v_i, 1]^T$  and the projected object points ( $proj(\mathbf{X}_i, \mathbf{P})$ ), defined as in (7) and calculated with the current estimation of the extrinsic camera matrix  $\mathbf{P}$ . Formally, this means:

$$S(\beta) = \sum_{i=1}^{N_{pts}} \|\mathbf{x}_i - proj(\mathbf{X}_i, \mathbf{P}(\beta))\|^2 \quad (10)$$

And the optimization problem can be written as

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} S(\beta) \quad (11)$$

For an adequate convergence in all directions, this minimization problem is solved with the Levenberg-Marquardt algorithm, also called damped least-squares. The optimization parameters are the entries of the matrix  $\mathbf{P}$ , which are grouped in the vector  $\beta$ . At each minimization step a damped update of the parameter vector  $\beta = \beta_{old} + \delta$  is used depending on the decent of the optimization function, as defined in (12) [9]

$$(\mathbf{J}^T \mathbf{J} + \lambda \operatorname{diag}(\mathbf{J}^T \mathbf{J})) \delta = \mathbf{J}^T [\mathbf{x} - proj(\mathbf{X}, \mathbf{P})] \quad (12)$$

where  $\mathbf{J}$  is the Jacobian matrix of the reprojection function and  $\lambda$  is the damping factor. It is tuned such that the convergence is moderated in the case of very fast descending functions, preventing from instability, and enhanced for slowly converging problems.

## 2.4 Triangulation

**Introduction** The basics for the triangulation technique are the camera projection equations (7). The difference is that in this case, one wants to find the real position of one point given its image points in images from multiple cameras. The governing equation is thus given by:

$$\mathbf{x}_j = \mathbf{C} \mathbf{P} \mathbf{X} \quad (13)$$

$$s_j \begin{bmatrix} u_j \\ v_j \\ 1 \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{R} & | & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \text{for } j = 1 \dots N_{cameras} \quad (14)$$

Both cases with fixed or free height  $Z$  are considered. The resulting system of equations has 2 or 3 unknowns and needs to be solved accordingly. The method applied is the one proposed by Hartley Zisserman ([4], Chapter 12.2). First of all, for each camera, the following matrix needs to be set up.

$$\mathbf{A}_j = \begin{bmatrix} u_j f_3 - f_1 \\ v_j f_3 - f_2 \end{bmatrix} \quad (15)$$

where  $f_k = Proj(k, :)$  denotes the  $k$ th row of the projection matrix and  $u_j, v_j$  are the image points of the required point in Camera  $j$ . The matrix  $\mathbf{A}$  is then composed of all rows  $\mathbf{A}_j$ , vertically stacked. It is of size  $4 \times (2N_{cameras})$ .

**DLT fixed height** If the height is specified, then finding the solution to (14) comes back to solving

$$\mathbf{A}'\mathbf{x} = \mathbf{b} \quad (16)$$

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}^T \mathbf{x} = -\mathbf{a}_3 Z - \mathbf{a}_4 \quad (17)$$

Where  $f_k$  denotes the  $k$ th row of the matrix  $\mathbf{A}$ . This system of equations can be solved in the least-squares sense, which leads to the solution  $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ . Adding the third component  $Z$  to  $\hat{\mathbf{x}}$ , the solution is obtained.

**DLT free height** If the height is not specified, one can simply use a single value decomposition of  $\mathbf{A}$ . The eigenvector with biggest eigenvalue corresponds to the solution of (14) in the least squares sense.



### 3 Echo-SLAM

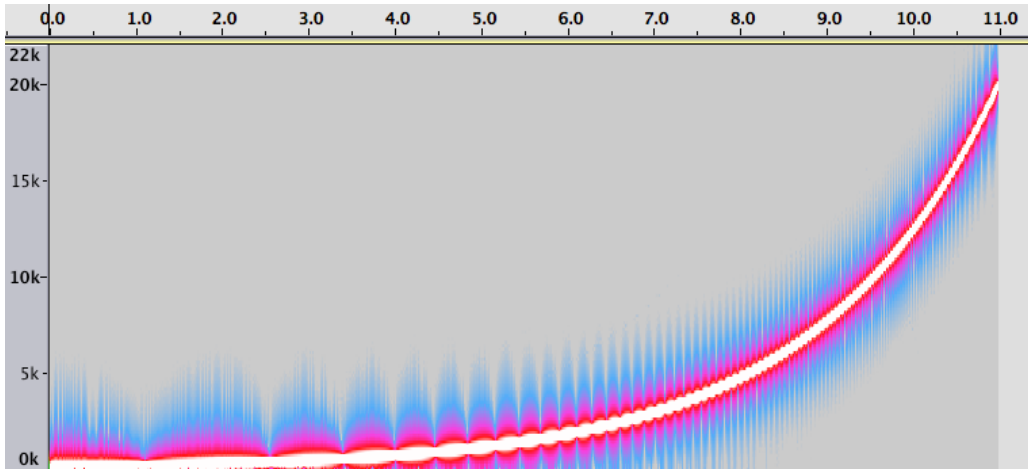
#### 3.1 Sine Sweep Generation

A common signal for recording room impulse responses is the sine sweep. Since for the scale of the present setup, mostly low frequencies are of interest, an exponential increase of frequencies is chosen, going from  $f_1 = 100\text{Hz}$  to  $f_2 = 20000\text{Hz}$ .

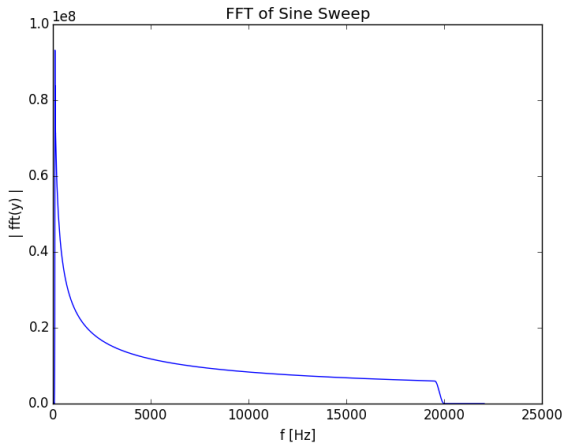
$$x_{exp}(i) = fac \times amp \times \sin\left(\frac{2\pi f_1 N}{F_s \log(\frac{f_2}{f_1})} (e^{\frac{i}{N} \log(\frac{f_2}{f_1})} - 1)\right) \quad \text{for } i = 0 \dots N \quad (18)$$

Where  $N = T_{in} F_s$  and  $F_s$  is the sampling frequency. The Fourier transform visualizes the richness of the signal in low frequencies (Figure 3b) and the spectrogram gives a more intuitive visualization (Figure 3a). The type chosen for the wavfile is a signed Integer of 16 bits, which can be read by *PyAudio*, so the signal is amplified by  $amp = 2^{16-1}$  and damped with a factor of  $fac = 0.8$ .

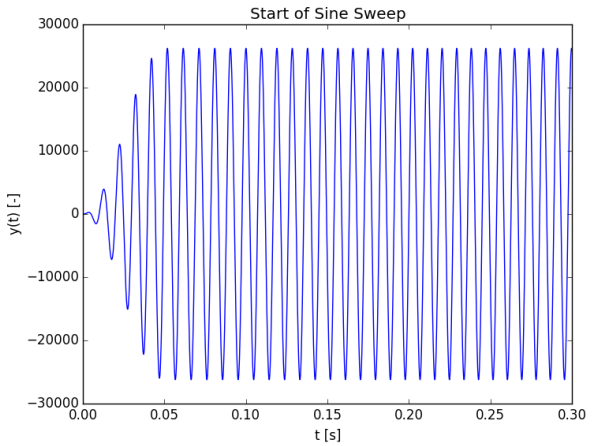
To avoid a too abrupt start of which the speakers would not be capable, a hanning window is applied to the start of the signal (see Figure 3c)



(a) Spectrogram of sine sweep



(b) FFT of sine sweep



(c) Zoom of smoothed start of sine sweep

Figure 3: Characteristics of sine sweep used for room impulse response recording

#### 3.2 Recording

The recording of the impulse response is implemented in a python script using the library *PyAudio*. For a duration of  $T_{out} = T_{in} + \Delta T (= 3\text{s})$ , frames read from the input wav-file are sent out through the output stream and simultaneously, the data is read on the input stream. The incoming data can be

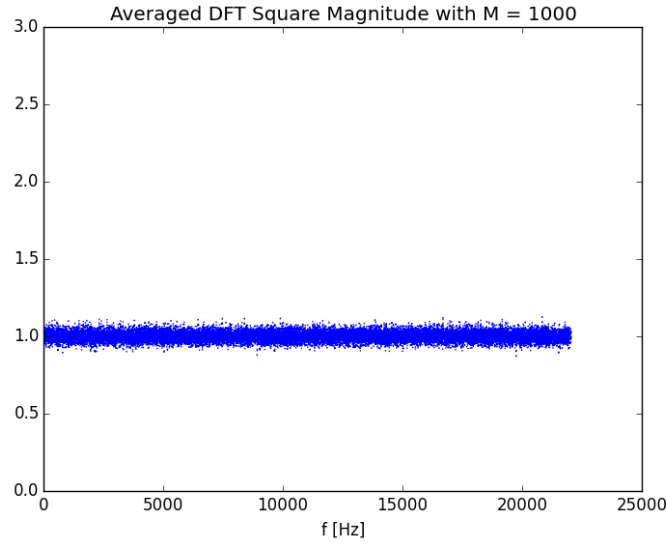
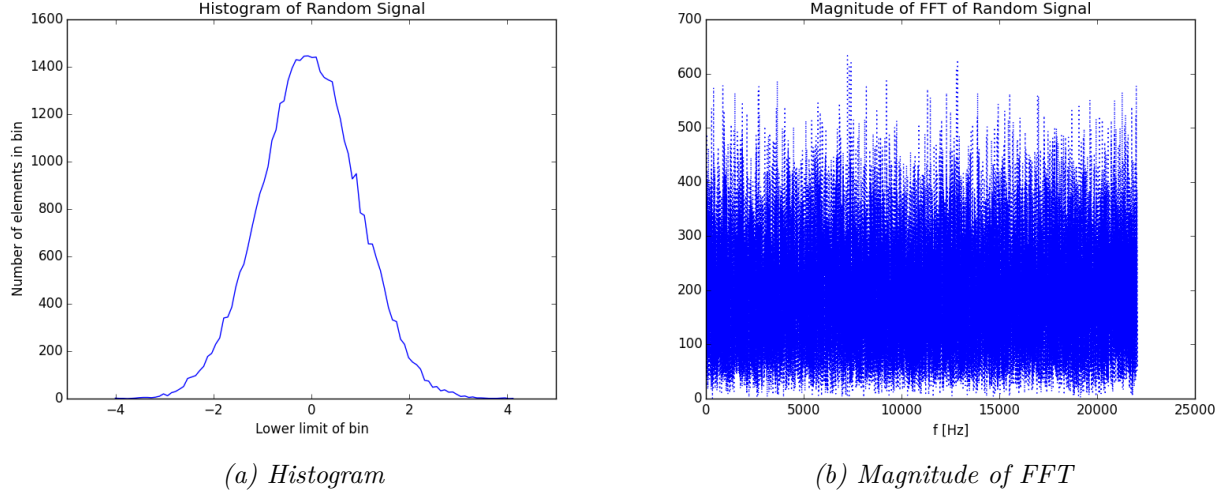
read from multiple channels ( $N_{channels}$ ). The data is stored in a matrix of size  $N_{Buffers} \times (N_{Channels} \times S_{Chunks})$  where  $N_{Buffers}$  denotes the number of buffers, calculated from  $N_{Buffers} = \frac{T_{out} \times F_s}{S_{Chunks}}$ . The frames from the different channels are stored in alternating order and need to be unwrapped in order to store them in separate files. The resulting data matrices are single-channelled and of size  $N_{Buffers} \times S_{Chunks}$ .

### 3.3 Analysis

#### 3.3.1 Calibration

It is required to differentiate the delay induced by the physical distance between microphone, walls and the speakers from the delay induced by the audio system itself. Therefore an analysis of the latency of the audio system is performed.

For this purpose, the speaker is placed at a well known position with respect to the microphone, so that the physical delay  $\Delta t_{iph}$  can be precisely calculated. It is then sufficient to get the total delay of the signal,  $\Delta t_{tot}$  which is composed of the physical delay and the latency ( $\Delta t_{tot} = \Delta t_{ph} + \Delta t_l$ ). The total delay is found by sending a reference signal  $u_{N1}$  (in this case, a approximation of white noise) and recording the response of the microphone,  $y_{N2}$ . The white noise is generated with a random signal of length  $T = 1$ s. Its histogram and fourier transform are shown in Figures 4a and 4b respectively. The power spectral density of this signal should be equal to one if it is indeed a perfect random signal. Indeed, when averaging over 1000 iterations of random signal generation, the obtained averaged squared magnitude is indeed close to 1 for all frequencies (see Figure 4c).



(c) Averaged Squared Magnitude of FFT with  $M=1000$

Figure 4: Characteristics of random signal used for calibration

The response is a scaled and delayed noisy version of the input. Finding the delay comes back to laying the output signal over the input signal with different phase lags until one gets a maximum similarity. The lag corresponding to this maximum is the total delay between the input and the output. The tool that performs these steps is the cross-correlation, which, for discrete signals can be written as:

$$r_{uy}[k] = \sum_{n=-\infty}^{\infty} u[n]y^*[n-k] \quad k = 0, \pm 1, \pm 2, \dots \quad (19)$$

As the cross-correlation is quite computationally expensive, only the first  $N_{max}$  samples of both input and output signals are correlated, which is sufficient  $N_{max}$  is chosen significantly bigger than the sample index of the expected delay (for example,  $N_{max} = 2 \text{ s} \times F_s = 88200$ )

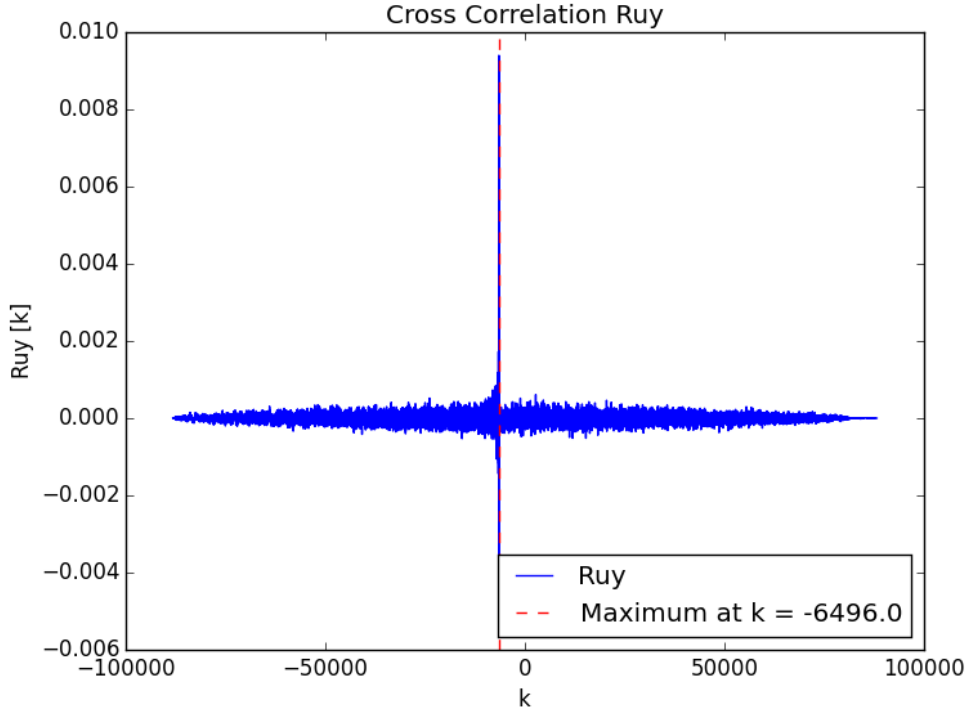


Figure 5: Cross correlation of white noise for latency estimation

From Figure 5, one can see that the maximum occurs at sample  $k_{max} = -6496$ , which corresponds to a delay of  $\Delta t_{tot} = k_{max}/F_s = 147.3$  ms. The distance between microphone and speaker being  $d = 660$  mm, one finds  $\Delta t_{ph} = d/c = 1.9$  ms, so the estimated latency of the sound system is about  $\Delta t_l = 145.4$  ms.

### 3.3.2 Echo SLAM

The Room Impulse Responses can be calculated from the frequency response of the input  $u(t)$  and of the recorded output  $y(t)$  as follows:

$$H(\omega) = \frac{Y(\omega)}{U(\omega)} \quad (20)$$

The analysis of the impulse response is done for one specific position of the robot. For the first considerations, the microphones are assumed to be omnidirectional and placed in the exact center of the robot. The estimated times of arrivals can then be found from the robot position and stored in a matrix  $U$  following the notation proposed in [5]:

$$U[n, k] = \tau_{n,k} = \frac{\|\tilde{\mathbf{s}}_{n,k} - \mathbf{r}_n\|}{C} = \frac{2d_{n,k}}{C} \quad (21)$$

where  $d_{n,k}$  denotes the distance between wall  $k$  and the robot at position  $n$ ,  $\tau_{n,k}$  denotes the corresponding time of arrival and  $\tilde{\mathbf{s}}_{n,k}$  and  $\mathbf{r}_n$  denote the position of the virtual source and the robot respectively.

## **4 Robot movement**

### **4.1 Control**

How robot is controlled.

### **4.2 Odometry**

Odometry algorithm used

## 5 Experimental Setup

### 5.1 Router

The Router used is a TP\_LINK TL\_MR3020 up to 150 Mbps

computer to router: (when connected to internet upload: 45, download: 50) sending 11MB, 11536384 (test10.zip) raspberry to router: 1.57 MB/s, (in 7.0s) computer to router: 1.73 MB/s (in 6.4s)

### 5.2 Camera Setup

**Hardware** The camera is made of a Raspberry Pi 2 Model B with the corresponding Raspi-Camera module. The OS loaded to the camera is the common Raspian WHEEZY system with a few scripts running on startup in the background. The camera connected is the Raspi-Camera module.

**Webcam setup** In order to connect to the LAN on startup, a few lines need to be added in */etc/network/interfaces*:

*Code 1: /etc/network/interfaces*

```
1 auto lo
2 iface lo inet loopback
3
4 iface eth0 inet static
5 address 172.16.156.138
6 netmask 255.255.255.0
7 gateway 172.16.156.1
8
9 auto wlan0
10 allow-hotplug wlan0
11 iface wlan0 inet static
12 address 172.16.156.139
13 netmask 255.255.255.0
14 gateway 172.16.156.1
15 wireless-essid korebot
```

The camera module comes with an own library with modules for taking single images (*raspistill*) or videos (*raspivid*). The following lines of code are added in a file *~/start\_camera.sh* in order to take pictures at regular intervals and save them to */tmp/stream*. This is where they will be found by the *mjpg\_streamer* module which sends the images to the corresponding server.

*Code 2: ~/start\_camera.sh*

```
1 #!/bin/bash
2 echo "Start camera stream"
3
4 sudo mkdir -p /tmp/stream
5 sudo chmod 777 /tmp/stream
6 raspistill -w 1280 -h 720 -q 50 -ex backlight -mm backlit -o /tmp/stream/pic.jpg
   ↳ -tl 100 -t 2147483647 &
7 LD_LIBRARY_PATH=/usr/local/lib mjpg_streamer -i "input_file.so -f /tmp/stream -n
   ↳ pic.jpg" -o "output_http.so -w /usr/local/www"
```

Line 6 tells the camera to take a picture of resolution  $w \times h = 1280 \times 720$  of quality 50%, 100% meaning no compression at all, at a time interval of  $tl = 100ms$  until the maximum time of  $t = 2147483647ms$  is reached, which is the maximum for a 32-bit signed integer. Other parameters like

the exposure (set to backlit, so that background is enhanced: TRY FIXEDFPS for constant lighting?), and the metering mode (what is this???) can be set. With these settings, the resulting picture size is not more than  $N_{pic} = 803kB$  (779412). (Apparent size: 786kB) As this size is much too high, the quality is set down to 10, which is equivalent to a quality of around 85 for most applications. This leads to a size of around 100kB. In addition to that, the thumbnail, which is a bitmap that also takes up unnecessary storage space, is deleted and a final image size of 80kB is obtained.

$N_{pic} = 803kB$  qual size 100 533 kB 50 455 kB 10 74 5 37 png 1549kB bmp 2773kB The network speed that needs to be provided to avoid idle times or image distortion is therefore:

Line 7 starts the mjpg streamer from where the picture has been stored and sends it to the local network.

$$S_{network} \geq \frac{N_{pic}}{tl} = \frac{1.1MB}{0.1s} = \frac{1.05Mbits}{0.1s} = 10.5Mbps \quad (22)$$

The router capacity is ??? which would be

**Button** A button is added to the camera such that it can be restarted or turned off without logging into the camera. This allows the Raspberry to be turned off correctly even when the network has not be set up properly or the ethernet is not working. The button is a pull-down resistor, thus the signal at the output goes from 0 to 1 when the button is pressed. An interrupt is triggered when the button is pressed in which it counts during 6 seconds weather the button is pressed or not at a rate of 1s. For robustness, a total of 3 signals is sufficient for the system to shut down. If less than 3 signals are registered during the 6 seconds, the system reboots.

Code 3: `~/switchoff.sh`

```

1  -*- coding: utf8 -*-
2  import RPi.GPIO as GPIO
3  import os
4  import time
5  #set up GPIO using BCM numbering
6  GPIO.setmode(GPIO.BCM)
7  GPIO.setup(10,GPIO.IN, pull_up_down=GPIO.PUD_DOWN)
8
9  #function called on pin interrupt
10 def button_triggered(channel):
11     counter = 0
12     counter_on = 0
13     while (counter <= 6):
14         time.sleep(1)
15         counter+=1
16         if (GPIO.input(10)):
17             counter_on+=1
18             if (counter_on >= 3):
19                 break
20
21     if (counter_on >= 3):
22         print("switchoff.py: Raspberry shutting down now")
23         os.system("sudo halt")
24     elif (counter_on < 3):
25         print("switchoff.py: Rapsberry is going to reboot now")
26         os.system("sudo reboot")
27 #setup pin interrupt
28 GPIO.add_event_detect(10,GPIO.RISING,callback=button_triggered,bouncetime=300)
29
30 #wait forever
31 while True:
32     time.sleep(0.001)
33
34 GPIO.cleanup()

```

**Setup** Both scripts above need to be started on startup of the camera. Therefore, the following lines are put in the file `/etc/rc.local`:

*Code 4: /etc/rc.local*

```
1 #Auto start camera
2 sudo /home/pi/start_camera.sh &
3 #Auto start shutdown
4 sudo python /home/pi/switchoff.py &
```

picture quality and compression: [www.raspberrypi.org/forum/viewtopic.php?f=43&t=73174&p=527300#p527300](http://www.raspberrypi.org/forum/viewtopic.php?f=43&t=73174&p=527300#p527300) wget: <https://www.maketecheasier.com/test-internet-connection-speed-from-terminal/>  
speedtest: <http://www.speedtest.net/>



## References

- [1] Jean-Yves Bouguet. Matlab calibration tool. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/).
- [2] DC Brown. Decentering Distortion of Lenses. *Photometric Engineering*, 32(3):444–462, 1966.
- [3] AE Conrady. Decentred Lens-Systems. *Monthly Notices of the Royal Astronomical Soceity*, 79:384–390, 1919.
- [4] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003.
- [5] Miranda Krekovic, Ivan Dokmanic, and Martin Vetterli. EchoSLAM : SIMULTANEOUS LOCALIZATION AND MAPPING WITH ACOUSTIC ECHOES. *LCAV*, 2015.
- [6] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. EPnP: An Accurate  $O(n)$  Solution to the PnP Problem. *International Journal of Computer Vision*, 81(2):155–166, 2009.
- [7] OpenCV. Camera calibration and 3d reconstruction. [http://docs.opencv.org/2.4/modules/calib3d/doc/camera\\_calibration\\_and\\_3d\\_reconstruction.html](http://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html). Accessed December 30, 2015.
- [8] Florian Wickelmaier. An introduction to MDS. *Reports from the Sound Quality Research Unit*, ..., page 26, 2003.
- [9] Wikipedia. Levenberg-marquardt algorithm. [https://en.wikipedia.org/wiki/Levenberg-Marquardt\\_algorithm](https://en.wikipedia.org/wiki/Levenberg-Marquardt_algorithm). Accessed December 31, 2015.
- [10] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000.