**Bio-motivated vision system for UAV navigation and control**

Master Thesis

Master of Info-bionics Engineering

Bionic interfaces and Integrated Systems Specialization

Author:

**Máté Pethő**

Thesis supervisor:

Tamás Zsedrovits, PhD, senior lecturer

**Hallgatói Nyilatkozat**

Alulírott **Pethő Máté**, a Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Karának hallgatója kijelentem, hogy ezt a szakdolgozatot meg nem engedett segítség nélkül, saját magam készítettem, és a szakdolgozatban csak a megadott forrásokat használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen a forrás megadásával megjelöltem. Ezt a Szakdolgozatot más szakon még nem nyújtottam be.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Pethő Máté

Contents

[Abstract 5](#_Toc497750137)

[Absztrakt 6](#_Toc497750138)

[1. Introduction 7](#_Toc497750139)

[2. Problem statement 8](#_Toc497750140)

[3. Literature review 9](#_Toc497750141)

[3.1 UAV navigation and control 9](#_Toc497750142)

[3.1.1 Dynamics of UAV control 9](#_Toc497750143)

[3.1.2 Previous methods on autonomous flying of UAVs 10](#_Toc497750144)

[3.2 Vision of humans and non-primate mammals 10](#_Toc497750145)

[3.2.1 The retina 10](#_Toc497750146)

[3.2.2 Higher information processing in the visual system 14](#_Toc497750147)

[3.3 Image processing techniques 14](#_Toc497750148)

[4. The model 15](#_Toc497750149)

[5. The software 20](#_Toc497750150)

[6. Results and Discussion 21](#_Toc497750151)

[7. Conclusion 22](#_Toc497750152)

[8. Acknowledgement 23](#_Toc497750153)

[9. References 24](#_Toc497750154)

**Abbreviations**

cGMP – cyclic guanosine-monophosphate

CNS – central nervous system

EKF – extended Kalman filter

GABA – gamma-Aminobutyric acid

IPL – inner plexiform layer

LGN – lateral geniculate nuclei

PID – proportional integrative derivate

RGC – retinal ganglion cell

SLAM – simultaneous localization and mapping

UAV – unmanned aerial vehicle

# Abstract

# Absztrakt

A pilóta nélküli repülőgépek, avagy UAV-k egyre nagyobb ismertségnek örvendenek, valamint számos területen nyújtanak kiváló potenciális alkalmazási lehetőségeket. Újabban az olcsóbb árkategóriás modellekből is egyre nagyobb a választék, ami szélesebb rétegek számára teszi lehetővé, hogy hozzáférjenek ehhez a technológiához. Az egyre változatosabb szerepkörök betöltéséhez egyre könnyebb irányításra és egyre kifinomultabb önálló működésre van szükség.

# 1. Introduction

Unmanned aerial vehicles (UAV) are becoming more and more common and they show excellent potential in many fields. Nowadays there is an increasing range of cheap models available, which makes it possible for SZÉLES RÉTEGNEK, to acquire one. These UAVs have to

In the thesis, I would like to go through the corresponding literature considering UAV navigation and control, the visual system of mammals, then show the model and the software created to …

Basic introduction to UAVs: advantages – disadvantages

Basic purpose of the thesis, what the proposed system wants to achieve.

# 2. Problem statement

The purpose of my thesis was to create a model of the mammalian vision system and to use it to create algorithm, which makes UAVs capable to execute basic autonomous flying tasks.

The main purpose of the proposed system.

The detailed task planned to be solved during the thesis work.

# 3. Literature review

To …. For this we must go through the basics of UAV dynamics and the current state of autonomous control. Because the base of the model originates from biological models and result, it is important to understand the structure and functionality of the visual system. For this purpose, the following chapter walks through the literature considering UAV control, image processing techniques and the visual system.

## 3.1 UAV navigation and control

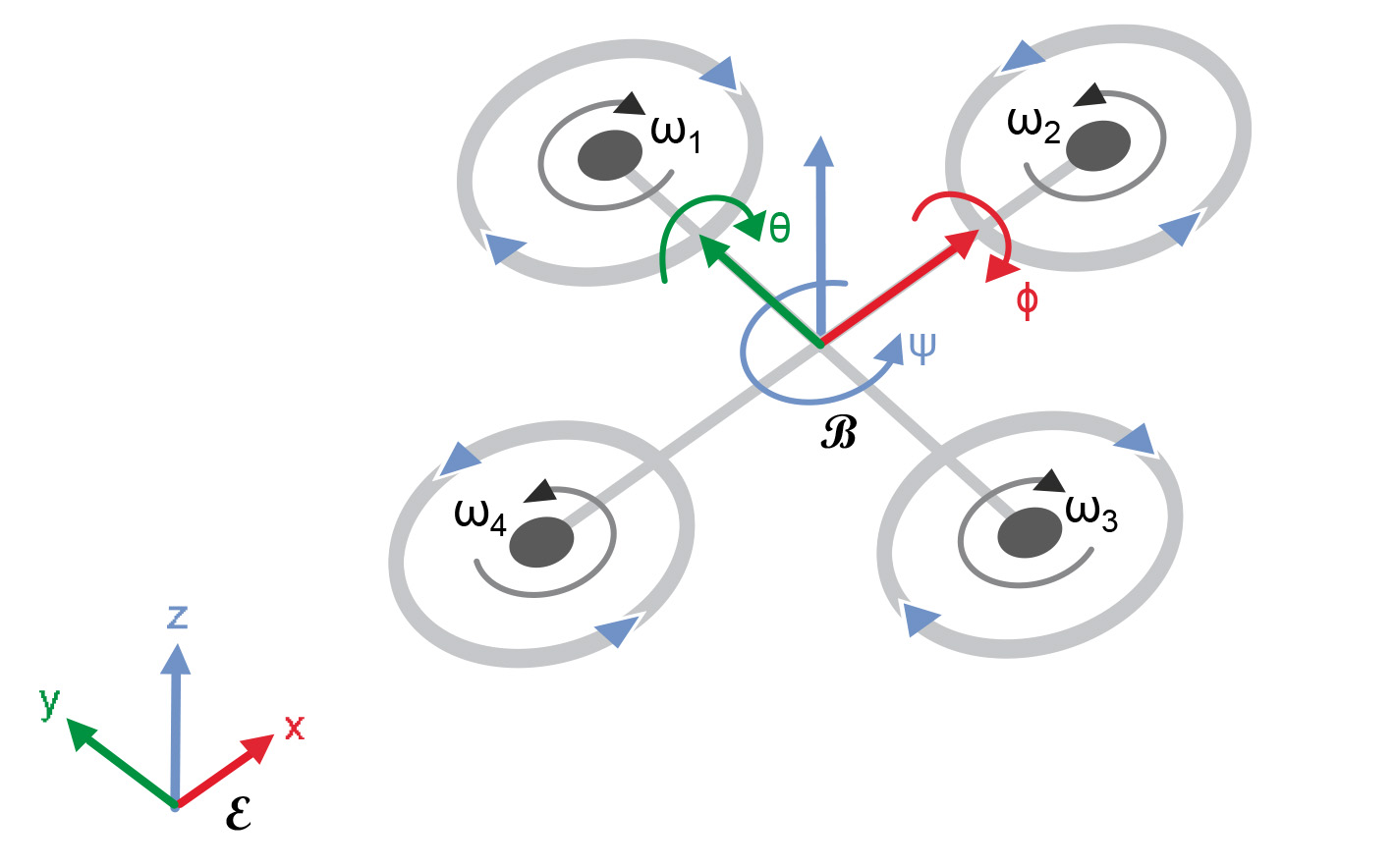
Description of USVs, we are going to use during the thesis work.

### 3.1.1 Dynamics of UAV control

To understand the problems and tasks, which come up during the planning and building an autonomous system for a UAV, we have to go through the basics of the dynamics of a quadcopter.

The control of quadcopters can be broken up to two main parts: low level control, which means …, and high-level control, which means the trajectory of the UAV. In case of the UAV, there are two inertial frame of references we must consider. One of them is the frame of the object (in this case it is the UAV) denoted as ℬ, the other one is the frame of Earth, denoted as ℰ (*Figure x*). The movement of UAVs can be characterized by four main components: 1) roll, which is tilt along the longitudinal axis (denoted as ), 2) pitch, which is movement along the lateral axis (denoted as ), 3) yaw, which is movement along the vertical axis, thus it enables sideward movement (denoted as ), 4) the forth component is the vertical displacement (denoted as ), which is created by the synchronous ÉS AZONOS EREDŐJŰ of output increment or decrement of the four rotors REF. If we denote the angular speed of the *ith* rotor as ωi (*Figure x*), then we can formulate the movement components the following way:

where b denotes the thrust coefficient, and d denotes the drag coefficient REF.



**Figure x. The quadrotor model REF.**

Euler Lagrange description of UAV control….

The low-level control of UAV considers the achieving of flight, the maintaining of UAV stability and the following of the designated flight path. Proportional integral derivative (PID) controller is a commonly used method to achieve this kind of flight control.

High level control

### 3.1.2 Previous methods on autonomous flying of UAVs

Previous methods for autonomous control of UAVs. Landing, obstacle avoiding, etc.

## 3.2 Vision of humans and non-primate mammals

As during my thesis, I developed a system based on the visual system of humans and non-primate mammals, it is crucial to go through the literature covering the morphological and functional basis of those systems. The mammalian visual system starts with the eyeball, a complex structure that is capable to focus the light projecting towards the organism and which is able to adapt to the changes of luminous intensity. By a truly complex motor system with three pairs of muscles and the incoming neural integrators from the CNS, our eye is able to perform the fastest movements in our body, the saccadic eye movement REF. The focal point of the incoming light beams will project to the retina, which is the first stop of the sensory system. It converts the incoming photons to actual information from our surroundings REF. As during the thesis work, the image processing was based on this sensory input, only this system will be explained in detail.

### 3.2.1 The retina

The retina is the first part of the mammalian visual system, from there visual information ascends to the neocortex through the optic nerve. After leaving the retina, the information crosses in the optic chiasm, then through the lateral geniculate nucleus, the optic radiation and the superior colliculus, it reaches the visual cortex, where complex image processing is executed and the sense of vision comes into being. The retina has a diameter of 200 µm, furthermore it consists of ten layers and has six main type of cells in it (*Figure x., Panel a*). It is enclosed by the vitreous body from the anterior side and by a layer of pigment cells from the posterior side. REF – Fonyó



**Figure x. Schematics of the mammalian retina. a) Panel A shows a cross-section of the whole retina. The numbers denote the main cell types: 1 – rod, 2 – cone, 3 – horizontal cell, 4 – bipolar cell, 5 – amacrine cell, 6 – ganglion cell. b) The synaptic terminal of cones. c) The synaptic terminal of rods. d) The axon terminal of a cone bipolar cell. e) Bipolar cell ribbon synapse with an amacrine and a ganglion cell REF.**

There are two types of photoreceptors in the mammalian retina: the rods and the cones (although the ratio, localization, importance and even the composition of subtypes may be quite different in case of the various species). Both photoreceptors transduce light into electric signal in the form hyperpolarization. The rods are activated on remarkably low luminance intensity, even one photon is capable to activate them, thus they have a role in the scotopic vision or twilight vision. Colour vision is created by the cones. Cone cells are packed quite densely in the *fovea centralis*, which is responsible for the sharp vison, this is the area where the focal point falls from the eye-lens. There are three different types of cones, these are activated by different wavelengths of visible light. The L or red cones are activated by long-wavelength-, the M or green cones by middle-wavelength-, while the S or blue cones by short-wavelength light, thus they are activated the wavelength of light corresponding to the origin of the names of the individual cone types. Comparison of L and M cones results the red-green (R-G) discrimination, comparison of S and the combination of L and M cones results in the blue-yellow (B-Y) discrimination. The luminance intensity information is acquired by the summation of all three types of cone cells REF – Fonyó + Wassle rew. + webvision. Rods and cones have a quite typical distribution (Fig x.), as stated above cones can be found in the fovea centralis in greatest abundance, while rod density will be higher in the surrounding areas. There will be one area, where none of the photoreceptors can be found, the blind spot, where the axon of the ganglion cells leave the retina REF.

<PICTURE OF ROD-ONE DISTRIBUTION>

**Figure x. The distribution of cone and rod cells in the retina. x-axis show the angle of eccentricity, y-axis shows the receptor density.**

Cones react to light with graded hyperpolarization. Their neurotransmitter is glutamate, which is an excitatory neurotransmitter. Thus, during dark there is a continuous release of glutamate, as the cone cells are in a depolarized state. This is caused by the so called “dark current”, which is created by cyclic guanosine-monophosphate (cGMP) activated ion-channels. cGMP level is modulated by the visual phototransduction – a signal transduction cascade based on the opsin G-protein coupled receptor. As light decreases the level of cGMP, it will be followed up by the hyperpolarization of the photoreceptor, thus decreased amount of glutamate will be released to the synaptic cleft REF - Fonyó. The synapse of the cones, which is called cone pedicle, is probably the single most complex synapse in the whole nervous system (*Figure x., Panel b*). Each pedicle contains 20-25 presynaptic ribbons and has more than 500 contacts [[1](#_ENREF_1)]. Three types of horizontal cells and at least nine types of bipolar cells might connect with the cone cell, thus after the first synapse, the light information proceeds forward on multiply parallel pathways. L-M CSAPOK KAPCSOLATA HORIZONTRÁLIS SEJTEKEN KERESZTÜL. This connection in the network ameliorates the response to light stimuli by filtering the uncorrelated noise by averaging [[1](#_ENREF_1)]. + egyéb ref

Glutamate released by the cone cells depolarize the OFF bipolar and horizontal cells during the dark period through ionotropic AMPA glutamate receptors, thus these cells are hyperpolarised in case the illumination of their receptive field, because of the decreasing glutamate level in the synaptic cleft. In contrary, ON cells are excited during illumination, while it remains hyperpolarized during dark periods because of the G-protein coupled glutamate receptors. OFF bipolar cells give excitatory output to OFF retinal ganglion cells (RGC) in the inner plexiform layer, which is the seventh layer of the retina. ON bipolar cells excite ON bipolar cells in case of activation during illumination. Consequently, OFF cells are excited by stimuli darker than the background, while ON cells are excited by stimuli brighter than the background. So far, every examined mammalian retina has had at least four types of ON- and four types of OFF cells, respectively REF.

Horizontal cells approach the axon terminal of cone cells from lateral position. They sum up the signal of several cone cells and they create lateral inhibition by its GABAergic output. Thus, the response of a given cone cell is modified depending on the illumination of adjacent cells. Consequently, edges will generate greater response, than adjacent areas with the same level of illumination REF. Horizontal cells sum up the input of light and subtract it from the local signal REF. Horizontal cells have greater receptive field than their dendritic tree, as they are interconnected by gap junctions REF.

There are at least 10-15 retinal ganglion cells, which are distinguishable by the size and shape of their dendritic tree. However, one thing is common in all cases, that is the cumulative receptive fields of the RGCs cover the whole retina, thus the incoming light or dark stimuli will excite at least one RGC of each type. As all these types tend to other “image processing” procedures, it can be stated, that the signal separated to at least 10-15 parallel channels [[1](#_ENREF_1)]. In the inner plexiform layer, seven types of RGCs (PV1-PV7) can be differentiated based on their dendritic arborisation. All these RGCs have circular receptive fields and each of them is excited by different kind of stimuli. For example, PV3 ganglion cells fire in close proximity to the edge of the stimuli in case of a flash of light, then they will fire again at the edge of the stimuli after it disappears. PV7 [[2](#_ENREF_2)].

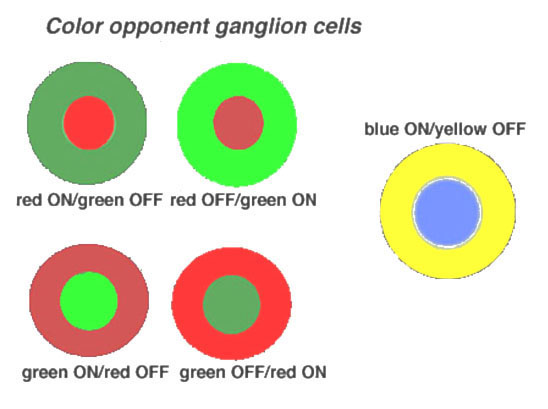
In case of bipolar- and ganglion cells – as seen before -, we can define a so called receptive field. From this area, the given bipolar- or ganglion cell receives excitation. In most cases, we can divide these receptive fields to two subparts: centre and periphery. In the level of bipolar cells, illumination of cone cells found in the periphery decreases the depolarization of the ON bipolar cell of the centrum. The previously described horizontal cells relay this lateral effect REF. In case of the receptive fields of RGCs, the field has a shape of a circle. The diameter of this circle depends on the type of the RGC. In the fovea centralis, RGCs has a smaller receptive field to the cells around the area of sharp view REF.

WEBVISION COLOUR DISCRIMINATION

MOZGÁS ÉRZÉKENY NEURONOK - Reichardt-Hassenstein model!

In case of the rod cell, there are two main pathways for the light information depending on the degree of darkness adaptation. In case of moderate dark adaptation, the signal from the rods will flow through the adjacent cones as they are connected by gap junctions. In case of complete dark adaptation, the information will flow through a different pathway. At that point, the gap junctions are closed, thus the information will propagate toward the rod bipolar cells, from where it will reach the RGC-s through the AII amacrine cells. The RGCs become sensitized during dark adaptation REF - Fonyó. Rods release glutamate during darkness as well as the cones, thus light stimuli hyperpolarises rod cells. The rod bipolar cell (from which there is only one type) is a ON cell, so it is excited by light. The AII amacrine cells are also excited by light and they connect to ON ganglion cells through gap junctions and they also connect to OFF ganglion cells, through inhibitory synapses REF.

State of information leaving the retina.



**Figure x. Colour opponency of ganglion cells.**

### 3.2.2 Higher information processing in the visual system

Purpose of higher visual processing: what information we will get after the processing + main parts of the visual system after the retina, with their respective roles.

Main locations with their morphology and functionality in details (most details must come from the components, which has been modelled in MATLAB).

chiasma opticum, tractus opticus, CGL, radiation ptica colliculus superiror

## 3.3 Image processing techniques

# 4. The model

Before building a C++ code for the image processing, I created a MATLAB model of the visual system, which was meant to perform image processing in a similar manner as it is executed in the visual system of humans and other mammals. In the following chapter, I would like to present the model and the algorithms behind it. The model manages to acquire image processing parameters (contrast information, edge information, main movement direction information) using a similar approach as the different retinal cell types and as the V1 region of neocortex. In the former case the information is created and evaluated by the functions imitating the retinal ganglion cells mainly, while the cortical columns of the V1 region was modelled by Gabor filter. The final part of the model imitated the role of higher order visual processing, it had to create a map….

During all processes in the model, there were some common aspects. In all cases - before the given procedure was executed – the input was saved in a memory, which contained the current and the previous four inputs. This memory played a cardinal role in the simulation of synaptic plasticity, as it gave possibility to produce a synaptic strength weighting matrix, for all inputs using the events of the past, thus it could modify the “synapse” of the given input. Synaptic strength was created the following way:

where is the m x n element of the synaptic strength vector, is the sum of one and the values received by the exponential distributions of the consecutive inputs stored in the memory in the nth row and the mth column. Sigmoid function is a common choice to model the synaptic transmission between neurons REF. This basic plasticity described above are present in all of the procedures of the model.

As stated in the Literature review, the retina is characterized by the convergence of information REF. Accordingly, most cell types of the retina has a receptive field, which includes all the inputs of a given neuron. Using this feature, to …. the following way:

|  |  |  |
| --- | --- | --- |
|  | i = 1:number of rows | |
|  |  | j = number of columns |
|  |  | receptiveField = inputMatrix(j:j + sizeOfReceptiveField, i:i + sizeOfReceptiveField) |

The chosen parameters of the receptive field of the various cell types imitated in the model can be observed in Table x.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| name of cell type or processing region | size of the receptive field | step | width of centrum | width of periphery |
| rod | 5\*5 | 1 | - | - |
| AII amacrine cell | 5\*5 | 1 | - | - |
| red-green discriminator cell | 4\*4 | 1 | 2 | 2 |
| yellow-blue discriminator cell | 4\*4 | 1 | 2 | 2 |
| direction selective cell | 8\*4 v. 4\*8 | 2 | - | - |
| contrast classification | 2\*2 | 2 | - | - |

**Table x. Size of receptive fields of the cell types in the model.**

The input values of each receptive field was summed and averaged modelling neuronal summation in the following way:

where width is width of the receptive field or its centrum or periphery, inpi is the ith input converging from the receptive field, type is the constant of the cell type (it is 0, if we are on the ON pathway, while -1 if we are on the OFF pathway), out is the output of the neuronal summation. If the current cell type does not support ON/OFF pathways, this constant is 0. As stated before, both ON- and OFF-cells have a centrum and a periphery. As we saw, the centrum and periphery is excited by the opposite stimuli REF. In the model this is achieved by the following conditionality:

IF periphery > centrum+0.5

out = 0;

ELSE IF periphery + 0.5 < centrum

out = 1;

ELSE IF periphery > centrum

out = out \* (1 – (periphery – centrum));

ELSE IF periphery < centrum

out = out \* (1 + (centrum – periphery));

The model also tried to imitate the variancies in the distribution of photoreceptors. To this end, the model contains a fovae, which is the centre of sharp view. In the model it is the central 10% of the view, and here both width of the receptive field and the step size are lower than the base values of the given cell. In these areas the system creates smaller kernels, mostly with a 2x2 pixel area and a 1 pixel centrum. It was created, so in the C++ program, this centrum would be adjustable to the quadcopter’s direction of travel.

In the following, we examine the properties of the cells imitated by the model. Rods get their scotopic input from the rgb picture using the rgb2grey() MATLAB function. As these cells are extremely sensitive to light, there was no threshold set up in this case. Other than this, a threshold was always used, so that the answer from the …. The output of the rods flow through the AII amacrine cell function, which soothed the contrast information originating from the rod pathway. The output was a value between 1.1-1.0 in case of the ON intensity ganglion function (see later), or 0.9-1.0 in case of the OFF intensity ganglion function.

In case of the cone pathway, all three main cone types were modelled by separating along the third dimension of the RGB input. Three types of comparison was performed: 1. red-green discrimination, 2. yellow-blue discrimination, 3. the cumulative discrimination of all colour components. The interaction of periphery and centrum is created by the collaboration of the photoreceptors, amacrine-, and bipolar cells in case of ON- and OFF-cells, respectively. The three type of contrast information processing was performed by separate functions. In case of the green-red and the yellow-blue discriminations, the functions performed the following steps: 1. initialization of ON/OFF function, 2. modification of synaptic strength matrix based on the memory, the multiplication of the input with the acquired synaptic weight, 3. creation of receptive field, iterating it through on the input as a kernel, 4. calculating the output matrix by the comparison of the normalised averages of the periphery and the centrum using the previously introduced conditionality.



**Figure x. Outputs of the red-green discrimination of a landscape. ON pathway on the left, OFF pathway on the right.**

****

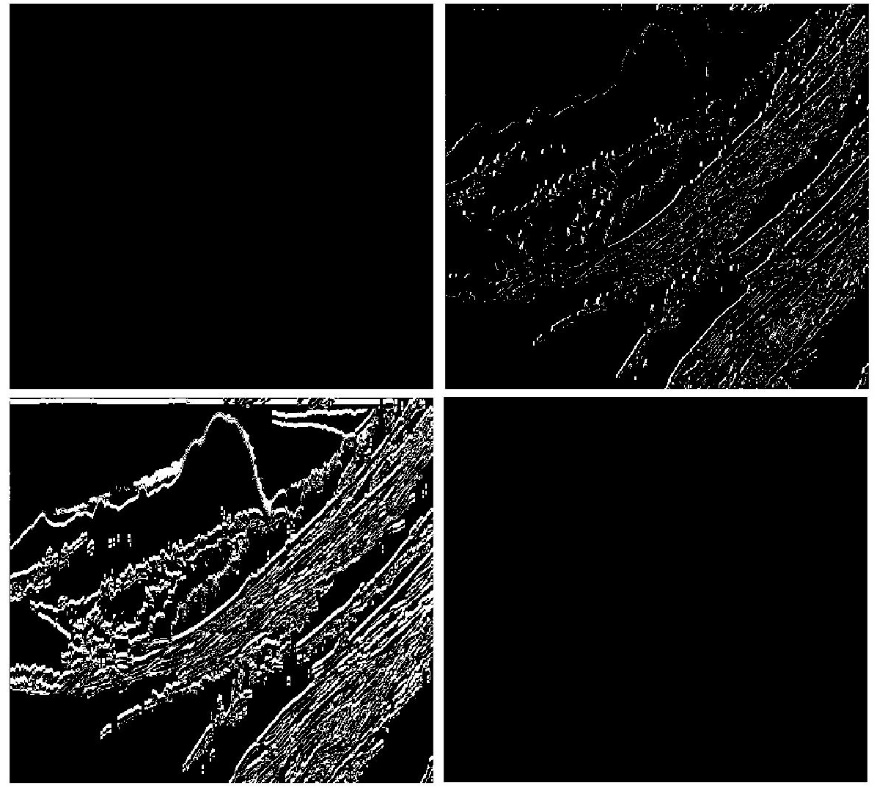
**Figure x. Outputs of the yellow-blue discrimination of a landscape. ON pathway on the left, OFF pathway on the right.**

In case of the three channel discrimination (Figure x.) the output data originated from the averaged values from the three separate channels from the S-, L-, M-cones. This averaged value was further modified by the rod pathway, as it was multiplied by the output of the AII amacrine cell. This procedure was executed in case of ON- and OFF-pathway, respectively.



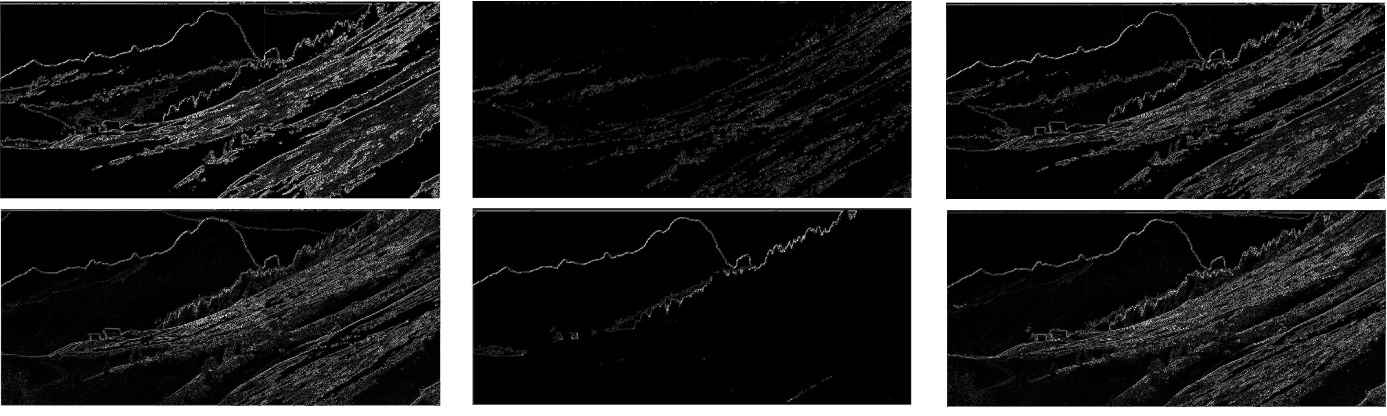
**Figure x. Outputs based on the three cone discrimination modified by the AII amacrine cell input from the rod pathway. ON pathway on the left, OFF pathway on the right.**

As the direction of moving can contain a great deal of information from a two dimensional picture, the model also tried to catch the motion sensing ganglion cells, the previously introduced direction cells of the retina. The goal was to get information from motion in the four main direction: upward, downward, rightward and leftward. Do to so, the model uses rectangular shaped receptive fields (Table 1.). These receptive fields was divided into two equal squares, these were meant to model the two main dendritic branch of the direction selective cells REF. The function behind the motion sensing evaluate two consecutive time … and two different conditions must be fulfilled simultaneously: 1. The difference in the intensity of the square during the two consecutive time … in the opposite side to the preferred direction of the movement (for which the given direction cell is sensitive) must reach a given threshold (VALUE), 2. the output of the square in the direction of preferred movement in time t must be higher than the output of the square in the opposite side of the direction of the preferred movement in time t-1 by a given threshold.



**Figure x. Output of movement sensitive ganglion cells. A. upward motion, B. downward motion, C. right-hand motion, D. left-hand motion.**

To gain the edge information from the input image, the model had to utilize Gabor filter, which is well-known to show similarities with the edge processing in the primary visual cortex REF. Thus it gave a perfect opportunity to gain the edge information using the previously produced contrast pictures in. For the execution in MATLAB a function was used to create the filter REF, than it was utilized by another function to perform the filtering. …. The created filters filter in four direction (-45̊, 90̊, 45̊, 0̊), furthermore with 10 different receptive field size (from 5 to 23).



**Figure X. The output from Gabor filter. From left to right: Filtering performed on the 1. red-green discrimination output, 2. yellow-blue discrimination output, 3. summarized discrimination output in case of ON- (upper) and OFF-pathway (lower), respectively.**

The last part of the model is the classificatory. It was created, so it would make an obstruction map, which could yield information for the quadcopter from safe passage during the autonomous flight. For this the classificatory should use three source of information: 1. contrast information, 2. edge information, 3. motion information. Using these modalities we may classify regions of the input image, an estimation if it a “safe passage” for the UAV. To identify obstructions, we have to consider multiply conditions. ….

As the classificatory function was mainly developed in C++, thus the underlying algorithm and its connection will be described in the next chapter.

The created model was able to perform image processing, it KINYERT information from the 2D image which can be used to plan the course of flight.

# 5. The software

Main problems and task for the conversion of the model in to C++ code + some words maybe from the chosen language, OpenCV and from the IDE of choice.

# 6. Results and Discussion

Details on the task desired to be achieved during the autonomous flight.

Detailed analysis on the success rate of the UAV, the analysis of performance and detailed description of it.

Discussion considering the result of the analysis and the worthiness of the created algorithm.

Discussion of future possibilities for enhancing and developing the system (higher navigational process, memory, etc.).

# 7. Conclusion

Conclusion of the thesis.

# 8. Acknowledgement

# 9. References

[1] H. Wassle, "Parallel processing in the mammalian retina," *Nat Rev Neurosci,* vol. 5, pp. 747-57, Oct 2004.

[2] B. Roska and F. Werblin, "Vertical interactions across ten parallel, stacked representations in the mammalian retina," *Nature,* vol. 410, pp. 583-7, Mar 29 2001.