UAV obstacle detection with bio-motivated computer vision

Máté Pethő [*†] and Tamás Zsedrovits[†]

[*] Department of Physiology and Neurobiology, Eötvös Loránd University, Budapest, Hungary,
[†]Pazmany Peter Catholic University, Faculty of Information Technology and Bionics, Budapest, Hungary

***Abstract* — Unmanned aerial vehicles (UAVs) are becoming more and more common. They show excellent potential for multiple types of autonomous work, although they must achieve these tasks safely. For flight safety, it must be assured that the UAV will not endanger its surroundings during autonomous operations; it will avoid collision with any objects in its flight path. Camera-based computer vision and artificial neural networks have shown to be effective in many applications. However, biological vision systems and the brain areas responsible for visual processing may hold solutions capable of acquiring information effectively. Previous work has shown the usability of biologically motivated algorithms using vision systems of insects or even behavioral patterns to solve computer vision problems. We are proposing a novel system, which performs visual cue extraction with algorithms based on the structure and functionality of the retina and the visual cortex of the mammalian visual system. We are also developing a modular artificial neural network with a training dataset, which will perform autonomous obstacle recognition tasks using the data from the image processing algorithm.**

***Index Terms* — vision-aided navigation, bio-motivated algorithms, artificial neural networks**

I. Introduction

Unmanned aerial vehicles (UAVs) are becoming more and more common, and they show excellent potential in many fields, such as aerial imaging, inspection tasks, and so on. [1]. Autonomous UAVs could achieve efficient task completion, but to do so, they would need precise and fast obstacle-avoidance algorithms, so they could operate without endangering themselves or their surroundings. To that end, more and more biologically inspired algorithms become available. For example, algorithms that mimic the behavior of bees [2], UAV structures and algorithms that show resemblance with the vision system of insects [3], and more are the focus of many research projects.

The mammalian vision system is one of the most precise sensor systems, making animals capable of navigating complex tasks in various environments (for example, hunting or fleeing from a predator). The first stage of the visual processing is in the retina, a multi-layered sensor system capable of converting photons into action potential coded information, which is the method of information flow throughout the nervous system. From the photoreceptors, rods are susceptible to luminance, while the three cone types (L-, M-, S-cones) are receptive to red, green, and blue wavelengths of the incoming light, respectively [4]. The retinal ganglion cells operate with concentric receptive fields of varying sizes. The receptive fields can be divided into ON and OFF regions; the interplay of these regions extract visual cues for further processing [4]. Information flows in parallel pathways from the retina towards the brain on the optic nerve, containing contrast [5], movement direction [6], edge information [7], and more. Thus, stimuli reaching the primary visual cortex (V1) already contain multiple types of information. In the primary visual cortex, detailed edge information is acquired from the field of vision by the cortical columns [8].

Further processing (color processing, motion- and depth perception, cognitive mapping) is achieved in the higher-order visual cortices (V2, V3, V4, and so on) [9]. Neural plasticity is the property, which makes neuronal circuits dynamic. The strength of a synapse depends on its previous activity. Long-term potentiation and depression play an essential role in enhancing more critical and meaningful inputs [10].

Artificial neural networks (ANN) try to mimic the behavior of neural networks through nonlinear activation functions and introducing plasticity through machine learning techniques. ANN is perfect for classification problems, such as obstacle recognition in the flight path. Complex ANN can process multiple information modalities, thus increasing the probability of giving the right answer for a problem. To that end, we can increase the number of hidden layers slightly, but the best solution is to build a hierarchical network. These modular artificial neural networks have strongly separated architectures. Each ANN part will compute its domain [11]. Furthermore, applications of deep convolutional networks in computer vision are becoming more and more prevalent [12].

In this paper, we propose a bio-motivated vision system mimicking the processes of the retina and visual cortex of mammals to acquire visual cues (contrast and edge information) — the resulting information then processed by an ANN to determine the obstacle-free pathways. The system could be a base of autonomous flight in the future, or it could work as a pilot helping feature, which allows avoiding a collision in case of human error.

II. Proposed Method

Two main tasks must be performed to achieve obstacle awareness and avoidance: visual cue extraction and obstacle recognition after learning. We implemented algorithms modeling the retina and primary visual cortex functions to model the processes of biological systems in feature extraction. To achieve learning and autonomous obstacle detection, we will create a modular multilayer neural network.

A. *Bio-motivated feature extraction*

To achieve the goal of obstacle avoidance, first, we must extract features from the input images. Existing computer vision algorithms and biological visual systems determine visual cues such as contrast, edge, and movement information. In our model, two main features (contrast and edge information) were obtained using algorithms modeling the retina and a primary visual cortex (example outputs in *Figure 1*).

Contrast can provide information about more significant differences in the input image. The basis of our contrast-detection algorithm is a two-dimensional Gaussian function:

$$\Phi(x,y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{1}{2\pi\sigma^2}(x^2+y^2)\right)}$$

where σ denotes the width of the bell-shaped surface and is the only free variable.

The difference of Gaussians (DoG) function models the signal processing of retinal ganglion cells [13]. Retinal ganglion cells have a central and a surrounding receptive field. The interaction of these two regions provides information of contrast, movement, and so on, even before the information reaches higher-level processing brain regions (for example, primary visual cortex) [4]. As in the retina, multiple color channel comparison was performed to gain the contrast information, such as red-green and yellow-blue discrimination, using the following equations:

$$RG_{input} = \frac{(R+G)}{\sqrt{2}}$$

$$YB_{input} = \frac{(R+G-2B)}{\sqrt{6}}$$

where R is red, G is green, and B is the blue color channel of the input picture.

Figure 1. Examples for the output images after processing (the six different contrast images and two cases from the 20 separate images containing edge-information).

Blue color channel and greyscale information were also processed using the DoG function, as in blue midget cells and the rod pathway in the retina [4]. The process of calculating the intensity based on the input of the three cones was performed by averaging all three color channels. The function was convolved on the pre-processed image (channel selection, discrimination, and modification with memory) to imitate the process.

Edge information is also an essential visual cue in the case of obstacle detection. For that purpose, the Gabor function was used as follows:

$$g_c(x,y) := \cos(\omega_x x + \omega_y y) e^{\left\{-\frac{x^2+y^2}{2\sigma^2}\right\}}$$

$$g_s(x,y) := \sin(\omega_x x + \omega_y y) e^{\left\{-\frac{x^2+y^2}{2\sigma^2}\right\}}$$

where σ represents the width of the receptive field and was set to (5,7,9,14,20) to represent the size-dependent sensitivity of the primary visual cortex. ω_x/ω_y gives the preferred orientation of the receptive field, which was chosen to be {-45, 0, 45, 90}, which represents the four main directions. Gabor function was implemented on the output of the three-cone intensity process. The resulting output resembles the information gained in the primary visual cortex, where cortical columns gather information from edges with different directions separately [8].

As the algorithms try to model the functionality of biological visual systems, a memory was constructed to modify the current input, a feature that characterizes neuronal communication. The exponential fit of the current and previous four inputs adjust the current input before the processes modeling retinal ganglion cells; thus, the methods acquire some form of plasticity.

As a result of the previously described processes, the system acquires 26 outputs (6 with contrast information, 20 with edge-information). This information will be the input of the neural network trained to recognize obstacles in the flight path.

B. *Obstacle recognition with ANN*

To achieve autonomous flight in advance, the UAV must learn to interpret the information acquired by the visual processing algorithms presented before. To that end, we segmented the input images into 64*64 grids and want to build a neural network (with a training dataset), which will be able to determine if the grid holds an obstacle or not. To create and train the network, we will use one of the most popular open-source libraries: TensorFlow [14], which Google initially developed.

For the proposed problem, we would like to construct a multi-layered ANN with a subnetwork system. The subnetwork will represent the information incoming from the individual grids. It will contain 26 input neurons, from which 6 depict contrast information, while 20 include edge information. There will be multiple hidden layers, which will be fully interconnected with the input layer. The subnetwork will end in a single neuron as an output layer in the case of each subnetwork. All activation functions are planned to be Rectified Linear Unit (ReLu) activation functions, except the output neuron, which will contain a sigmoid activation function. The number of subnetworks will be equal to the number of grids on the input image. The main network will include multiple hidden layers as well. Again, all the neurons will use the ReLu activation function. The output will have the same number of values as the number of inputs (thus the number of grids). This structure again follows the mammalian visual system, where the processing structure follows the representation of the input.

In a different trial, we would like to examine the usability of the neural network by providing 3D inputs and outputs to it, where the third dimension will be time. In this approach, 10 or 20 consecutive images will be added and converted into 3D input and output after image processing. This approach may provide movement information based on the movement of the UAV. The acquired network will operate in a provided time interval, but it may be sufficient to estimate the motion of the obstacles in the surrounding environment; thus, flight changes may be prompted in time in case of danger of collision.

During training, we would like to examine the effectiveness of different loss (like sigmoid cross-entropy, softmax cross-entropy, and so on) and optimizer functions (like gradient descent optimization, Adam optimizer, and so on) for this problem [15].

We recorded Parrot AR Drone 2.0 flights around display panels in a big hall to create a dataset for training and testing. The recordings were segmented, and the previously described image processing was performed on them. The output was saved as an n*26 matrix in mat files, where n is the number of grids on the image (220 in the 64*64 sized grids) (*Figure 2*).

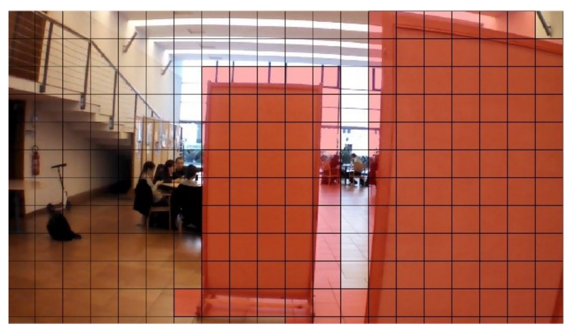

Figure 2. An example for the manual obstacle labeling using predefined script sizes (64*64 pixels in the current case)

Using a custom-written MATLAB script, grids containing obstacles (the display panel) were annotated on all images manually. Our dataset contains roughly 4000 images, but the preliminary tests showed that it is not enough to learn the desired features with the desired accuracy in the current ANN.

We are currently increasing the number of test images. The final training data set planned to contain around 8000 images, from which roughly half of them will include the avoidable obstacle (display panel).

III. CONCLUSION

After the implementation and testing of the proposed neural network, we would like to examine the real-time usability of the system. For that purpose, the image processing algorithms are reimplemented in C++, to achieve the best possible computational speed as well as make the software more compatible with the core system of the UAV.

We want to widen the obstacle types. Currently, we use display panels, but we would want to extend the recognizable obstacles to any kind of object in the room. In the long run, we would also like to implement the avoidance of the borders of the operation area (walls, floor, ceiling in case of an indoor environment).

We also would like to extend the training set with more environments and obstacle types, as we could not find publicly available data sets for this problem. With this work, we would like to lay down the foundation for such a dataset.

Later on, we would like to implement a “fovea”-like region, which is the location of sharp vision in the retina. It would be a small segment of the sight of the forward camera of the UAV, depending on its movement direction. We would try to use a neural network trained for such a smaller image size to decrease computation cost. The whole sight will be inspected if this smaller area holds obstacle(s) and flight path modification is needed.

ACKNOWLEDGMENT

This work has received funding from Pázmány Péter Catholic University KAP R&D and equipment grants (KAP20-51005-1.1-ITK, KAP20-53019-3.6-ITK, and KAP20-53019-3.6-ITK). The research was partially funded by the Hungarian National Research, Development and Innovation Office through the research grant TUDFO/51757-1/2019-ITM. This research has been partially funded by the Hungarian National Research, Development, and Innovation Office Postdoctoral Researcher funding PD 128699.

REFERENCES

[1] K. Mathe and L. Busoniu, "Vision and Control for UAVs: A Survey of General Methods and of Inexpensive Platforms for Infrastructure Inspection," *Sensors (Basel),* vol. 15, pp. 14887-916, Jun 25 2015.

[2] N. J. Sanket, C. D. Singh, K. Ganguly, C. Fermuller, and Y. Aloimonos, "GapFlyt: Active Vision Based Minimalist Structure-Less Gap Detection For Quadrotor Flight," *Ieee Robotics and Automation Letters,* vol. 3, pp. 2799-2806, Oct 2018.

[3] J. R. Stowers, "Biologically Inspired Visual Control of Flying Robots," PhD, University of Canterbury, Christchurch, New Zealand, 2012.

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

[5] C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," *J Physiol,* vol. 187, pp. 517-52, Dec 1966.

[6] H. B. Barlow, R. M. Hill, and W. R. Levick, "Retinal Ganglion Cells Responding Selectively to Direction and Speed of Image Motion in the Rabbit," *J Physiol,* vol. 173, pp. 377-407, Oct 1964.

[7] T. A. Munch, R. A. da Silveira, S. Siegert, T. J. Viney, G. B. Awatramani, and B. Roska, "Approach sensitivity in the retina processed by a multifunctional neural circuit," *Nature Neuroscience,* vol. 12, pp. 1308-U135, Oct 2009.

[8] T. S. Lee, D. Mumford, R. Romero, and V. A. Lamme, "The role of the primary visual cortex in higher level vision," *Vision Res,* vol. 38, pp. 2429-54, Aug 1998.

[9] K. Grill-Spector and R. Malach, "The human visual cortex," *Annu Rev Neurosci,* vol. 27, pp. 649-77, 2004.

[10] T. V. Bliss and G. L. Collingridge, "A synaptic model of memory: long-term potentiation in the hippocampus," *Nature,* vol. 361, pp. 31-9, Jan 7 1993.

[11] E. J. W. Boers and H. Kuiper, "Biological metaphors and the design of modular artificial neural networks," MSc, 1992.

[12] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience,* 2018.

[13] R. A. Young, "The Gaussian derivative model for spatial vision: I. Retinal mechanisms," *Spat Vis,* vol. 2, pp. 273-93, 1987.

[14] M. Abadi, "TensorFlow: Learning Functions at Scale," *Acm Sigplan Notices,* vol. 51, pp. 1-1, Sep 2016.

[15] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*: MIT Press Cambridge, 2016.