PROCEEDINGS OF SCIENCE

Visual Localization for Copter based on 3D Model of Environment with CNN Segmentation

Alexander Buyval[a1]**; Ruslan Mustafin**[b]
Innopolis University
Innopolis, Russia
E-mail: [a]a.buyval@innopolis.ru; [b]r.mustafin@innopolis.ru

Mikhail Gavrilenkov
Bryansk State Technical University
Bryansk, Russia
E-mail: gavrilenkov@umlab.ru

Aidar Gabdullin[c]**; Ilya Shimchik**[d]
Innopolis University
Innopolis, Russia
E-mail: [c]a.gabdullin@innopolis.ru; [d]i.shimchik@innopolis.ru

This paper introduces a novel approach to indoor visual localization based on a particle filter, CNN-segmentation and the nearest edge method for particle weight estimation. A main algorithm is used by detecting the edges on the image from camera and then mapping them to a 3D model of the room. The main contribution of the paper is to introduce a novel approach that allows to get rid of such problems as noise generated by textured objects, edges created by dynamic objects and groups of unexpected objects which are considered to be excessive and can be removed from localization algorithm. The algorithm described in this paper was verified on a computer model in ROS/Gazebo environment.

ISCC2017
16-17 December 2017
Guangzhou, China

[1]This research has been supported by Russian Ministry of Education and Science within the Federal Target Program grant (research grant ID RFMEFI60917X0100) and by KAMAZ Company, the industrial partner of the research.

1.Introduction

The problem of localization of a mobile robot has been considered important in recent years. Localization of a mobile robot, UAV in particular, is critical for manipulation of the robot inside a building. In contrast with robots that function in open outdoors environment, robots that work in buildings cannot rely on signals from global positioning systems (GPS/GLONASS). To obtain localization information and make decisions on such conditions, it is common to use inertial sensors, laser rangefinders and data from cameras.

For example, a combination of IMU and LIDAR data can be used to estimate the indoor position of Unmanned Ground Vehicle (UGV) [7]. One of the disadvantages of a LIDAR is its high cost and inability to operate within environments that contain glass or mirror elements. Another approach based on radio signal analysis and RFID beacons have been used in [10], [11]. However, this approach is far less precise as compared to previously discussed. Usage of camera data to detect and classify visual beacons, location of which is known a priori is described in [8]. A vehicle is then able to localize itself based on the known location of beacons. Another approach is to use visual information for localization, but it uses natural indoor objects as beacons: doors, stairs [9].

One of the advanced methods to solve localization problem is to use information from onboard camera and process it by using particle filter. The approach used in this work is based on comparison of images from camera with the images acquired from a 3D model of a building. Such an algorithm outperforms other alternatives from several points of view, but, unfortunately, does not come without shortcomings. One such shortcoming is the redundancy of information obtained from camera as well as noise that comes from variations in texture and dynamic objects.

This paper introduces an approach for camera images pre-processing that allows to get rid of these shortcomings and improve their value for localization tasks.

1.1 Base Algorithm for Localization

The description of an algorithm of visual localization which utilizes edges of the images for the main visual characteristic is presented in [1], [2]. These works also introduce the method of likeness evaluation of two images - gathered from sensor and a model. This measure of likeness is then used as a main criterion for assessing hypotheses of robot location. The similar approach for outdoor localization of ground vehicle is used in works [5,6]. Algorithm of particle filter is used for processing information about frame likeness and information obtained from other sensors. Particles in that filter represent the hypotheses about possible robot location. In case we do not have a priori data about UAV location, the algorithm uniformly distributes particles to all possible robot positions in the environment.

To estimate the probability of each hypothesis (particle), it is necessary to estimate the nearness of borders obtained from camera image to boarders obtained from corresponding modelled image.

(1)The focused algorithm executes the following steps sequentially:
(2)Extraction of borders on image acquired from camera.
(3)transform to find lines from extracted borders. The lines are used as input data to calculate
(4)probabilities of each particle
(5)Rendering the image according to the particle state
(6)Finding contours on the resulting rendered image
(7)Using Hough transform to find lines in the contours from the previous step

For each line a set of points is constructed and a normal of a finite length is built for each point. If the normal intersects with any other line extracted from the camera image, the length of the normal is used to calculate the total weight of the line according to the following:

$$g(d) = exp\left(-\frac{d^2}{2\sigma^2}\right), \tag{1.1}$$

where d – is the normal length, σ – defines the weight of the normal w.r.t the length of the normal (allows to increase or decrease the influence of long normals).

For each line, the total weight is calculated by summing all weight of all normals according to:

$$l = \frac{\sum_{i=0}^{S} g(d_i)}{S}, \tag{1.2}$$

where S – total number of lines on the image.

The total hypothesis probability is calculated by combining probabilities of each particle according to the following formula:

$$W = \propto \cdot exp\left(k \frac{\sum_{n=0}^{m} l_n}{m}\right), \tag{1.3}$$

where m – number of lines, $\propto, k$ – parameters that regulate strength of individual lines in the final result.

During experiments we obtained the following parameters which are the most suitable for typical indoor environment: α is 1.0, k is 3.0, σ is 0.3. The most major parameter of our algorithm is a number of particles. In this paper we used 100 particles. Details regarding choosing the number of particles can be found in [1].

1.2 Applications of Segmentation to Reduce the Amount of Insignificant Hypotheses on the Image

One of the key disadvantages of the base algorithm is that it uses all edges found on the image. In the real image from UAV, edges may be detected not only from the structure, but also from variations of the texture on the floor or on the walls. These edges will be absent in the 3D model and, as a result, they will reduce weights of corresponding particles in particle filter. The image may also contain dynamical objects (humans, animals), or objects that are missing in the 3D model of the room (furniture, home appliances etc.). The edges extracted from these objects will reduce the accuracy of localization.

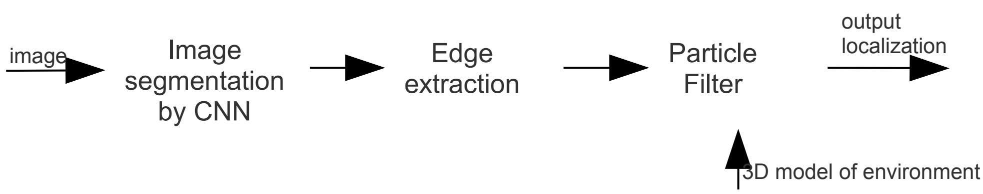

Figure 1: Localization pipeline. An image from camera is fed into segmentation CNN and resulting segmented image is used to detect edges. Detected edges are then processed with particle filter and final location is estimated.

To solve the aforementioned problems, in this work we present an application of segmentation technique to the image from the camera. Then we use the segmented image for further analysis. For the segmentation algorithm, we use convolutional neural networks with the appropriate architecture. Fig. 1 presents a pipeline of image data processing. Object location is a result of the shown pipeline.

As a base neural network for the upgraded algorithm we use SegNet [3], [4]. This choice is motivated by a compromised optimum between the accuracy of segmentation (82-85%) and time of execution (0.1-0.2s). Fig. 2b presents the same frame as Fig. 2a, but segmented using the neural network. As can be observed, the relative borders between ceiling and floor are clearly distinguishable, as well as the door on the left hand side.

Applications of segmentation allow to get rid of the textures on the image and also highlight the objects on the image that might be absent from the 3D model of the environment. Also, this allows to filter out such edges from the analysis.

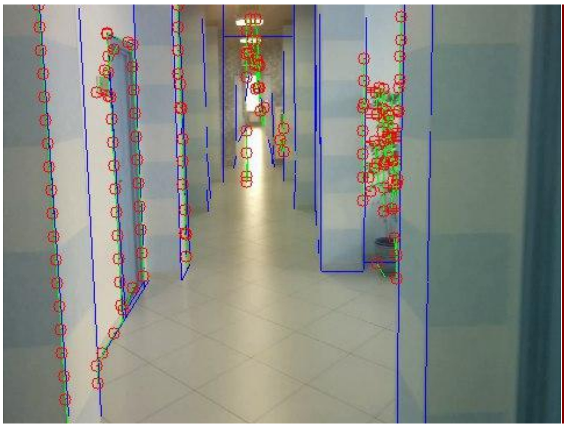

Figure 2a: Edges on the image and corresponding edges from the modelled image (blue lines - edges from the modelled image, green lines and red circles - found on the image from the camera).

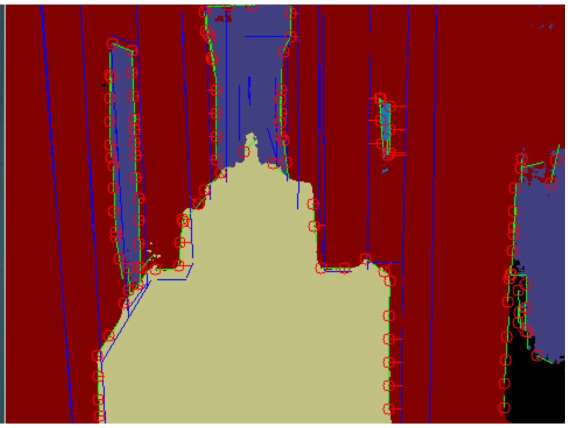

Figure 2b: Segmented image with subsequent searching of the edges and comparison of the found edges

1.3 Experimental Results

For experimental validation we have implemented the algorithm based on ROS as a set of nodes: segmentation node, localization node and edge correspondence node.

All experiments were carried out on a desktop PC with the following configuration: OS Ubuntu 16.04, ROS Kinetic, RAM: 8GB; CPU: Intel Core i7 7700HQ 2.8ГГц; graphics card: NVIDIA GeForce GTX 1050(m) 2GB.

In Gazebo simulator we created a model of an office corridor with the length of 10 meters, the longest side corresponding to the Y axis. In the experiment, the quadcopter was controlled manually and moved alongside the corridor.

Firstly, we evaluated the execution time for each component of the system. Table 1 presents the results of our measurements. As can be observed, segmentation was made full use, however, the localization processing time for a single frame does not exceed 200ms. Assuming low movement velocities of an UAV, the processing time is satisfactory. More powerful computational resources and more advanced neural network architecture will help to reduce computational times.

	Image acquisition, ms	**Image preprocessing , ms**	**SegNet processing, ms**	**Result presentation, ms**	**Total, ms**
min	4,43	1,35	165,74	4,51	176,05
avg	4,47	1,54	167,10	4,23	177,36
max	4,50	1,74	168,47	3,94	178,67

Table 1: Calculation Time in Segmentation Module

Fig. 3 and Fig. 4 present the comparison of measured and real position of the UAV along X and Y axes respectively. Information from Gazebo simulation was used as the ground truth.

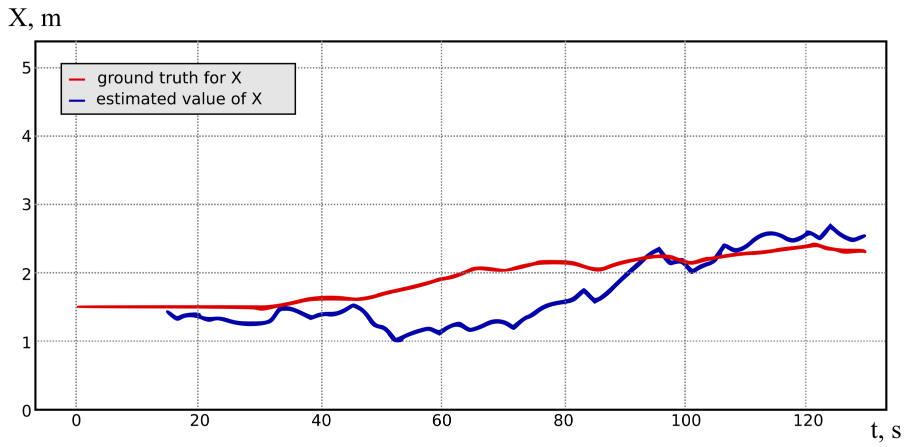

Figure 3: Comparison of Computed and Ground Truth Position of the Copter along Axe X

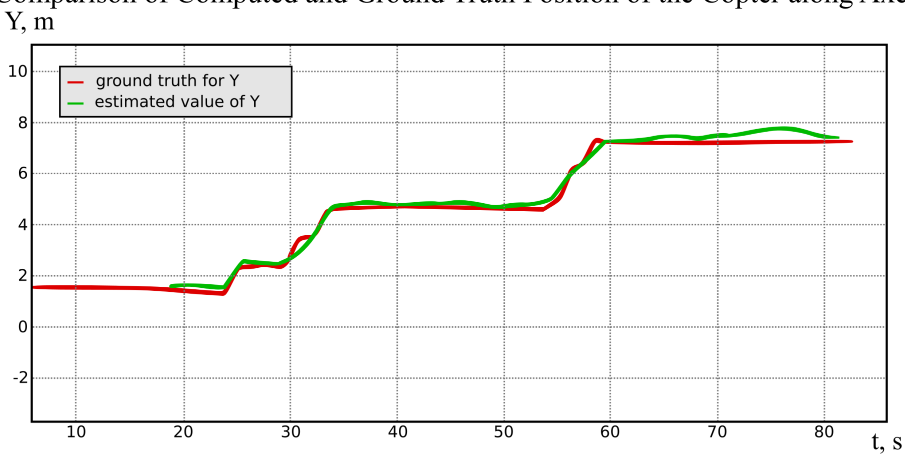

Figure 4: Comparison of Computed and Ground Truth Position of the Copter along Axe Y

From Fig. 4, it is possible to observe that ground truth position data and estimated position data are very close to each other, implying that estimated robot position matches the real one for the whole duration of the experiment along Y axis.

Similar situation can be observed in Fig. 4, however, there is a substantial deviation (around 0.4m) from the ground truth in the middle of the experiment, despite of the fact that deviations in all other parts were within reasonable limits.

After a series of simulation experiments, it was calculated that the average localization error of proposed algorithm is 0.32 meters. Higher accuracy of 0.1 meters can be achieved with classical indoor localization methods which adopts LIDAR sensor. Presence of LIDAR sensor significantly increases the cost of a UAV, and also LIDAR has low accuracy with glass windows and mirrors. In comparison with the methods of visual SLAM, this method is more robust for a typical office room because it uses typical visual features for this type of room (the edges of walls, windows, doors, etc.). The detailed comparison of different visual SLAM algorithms is presented in [12].

1.4 Conclusion

This work presents a CNN segmentation-based approach for an image-based visual localization. This approach allows to reduce the influence of insignificant edges on visual localization algorithms that use a known 3D model of the environment.

The efficiency and effectiveness of the approach was demonstrated in the Gazebo/ROS simulation.

References

[1] Buyval A. and Gavrilenkov M., *Vision-based pose estimation for indoor navigation of unmanned micro aerial vehicle based on the 3D model of environmen*t, in IEEE Int. Conf. on Mechanical Engineering, Automation and Control Systems (Tomsk, Russia, 2015).

[2] Buyval A., Gavrilenkov M., Magid E. *A multithreaded algorithm of UAV visual localization based on a 3D model of environment: implementation with CUDA technology and CNN filtering of minor importance objects.* // In Proceedings of the 2017 International Conference on Artificial Life and Robotics (ICAROB 2017). January 19-22, 2017, Miyazaki, Japan. P.356-359.

[3] Alex Kendall, Vijay Badrinarayanan and Roberto Cipolla "*Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding*." arXiv preprint arXiv:1511.02680, 2015.

[4] Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla "S*egNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.*" PAMI, 2017. arXiv:1511.00561v3 [cs.CV] 10 Oct 2016

[5] Nuske, S., Roberts, J., & Wyeth, G.. *Outdoor visual localization in industrial building environments.* // IEEE International Conference on Robotics and Automation, Pasadena, CA., 2008, 544-550

[6] Nuske, S., Roberts, J., & Wyeth, G. *Robust outdoor visual localization using a three‐dimensional edge map* // Journal of Field Robotics. 2009 №26(9), 728-756.

[7] M. Burri, , et al., "*Real-time visual-inertial mapping, re-localization and planning onboard MAVs in unknown environments*," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, 2015, pp. 1872-1878.

[8] K. Okuyama, T. Kawasaki and V. Kroumov, "*Localization and position correction for mobile robot using artificial visual landmarks*," *The 2011 International Conference on Advanced Mechatronic Systems*, Zhengzhou, 2011, pp. 414-418.

[9] Souto, Leonardo AV, et al. "*Stairs and Doors Recognition as Natural Landmarks Based on Clouds of 3D Edge-Points from RGB-D Sensors for Mobile Robot Localization.*" Sensors 17.8 (2017): 1824.

[10] J. Biswas and M. Veloso, "WiFi localization and navigation for autonomous indoor mobile robots," *2010 IEEE International Conference on Robotics and Automation*, Anchorage, AK, 2010, pp. 4379-4384.

[11] Sunhong Park and Shuji Hashimoto, "Indoor localization for autonomous mobile robot based on passive RFID," *2008 IEEE International Conference on Robotics and Biomimetics*, Bangkok, 2009, pp. 1856-1861.

[12] Buyval A., Afanasyev I,and Magid E., *Comparative analysis of ROS-based monocular SLAM methods for indoor navigation,* Proc. SPIE 10341, Ninth International Conference on Machine Vision (ICMV 2016), (March 17, 2017)

PoS(ISCC 2017)035