UNIVERSITY OF BATH

Citation for published version:
Martinez Hernandez, U, Cedeno-Campos, V & Rubio-Solis, A 2019, Active visual object exploration and recognition with an unmanned aerial vehicle. in *International Joint Conference on Neural Networks (IJCNN).*, 8851738, Proceedings of the International Joint Conference on Neural Networks, vol. 2019-July, IEEE, U. S. A. https://doi.org/10.1109/IJCNN.2019.8851738

DOI:
10.1109/IJCNN.2019.8851738

Publication date:
2019

Document Version
Peer reviewed version

Active visual object exploration and recognition with an unmanned aerial vehicle

Uriel Martinez-Hernandez
EEE Department
University of Bath, Bath, UK
u.martinez@bath.ac.uk

Victor Cedeno-Campos
MEng Department
University of Sheffield, Sheffield, UK
v.cedeno-campos@sheffield.ac.uk

Adrian Rubio-Solis
ACSE Department
University of Sheffield, Sheffield, UK
a.rubiosolis@sheffield.ac.uk

***Abstract*—In this paper, an active control method for visual object exploration and recognition with an unmanned aerial vehicle is presented. This work uses a convolutional neural network for visual object recognition, where input images are obtained with an unmanned aerial vehicle from multiple objects. The object recognition task is an iterative process actively controlled by a saliency map module, which extracts interesting object regions for exploration. The active control allows the unmanned aerial vehicle to autonomously explore better object regions to improve the recognition accuracy. The iterative exploration task stops when the probability from the convolutional neural network exceeds a decision threshold. The active control is validated with offline and real-time experiments for visual exploration and recognition of five objects. Furthermore, passive exploration is also tested for performance comparison. Experiments show that the unmanned aerial vehicle is capable to autonomously explore interesting object regions. Results also show an improvement in recognition accuracy from 88.14% to 95.66% for passive and active exploration, respectively. Overall, this work offers a framework to allow robots to autonomously decide where to move and look next, to improve the performance during a visual object exploration and recognition task.**

I. Introduction

Aerial robotics has shown a rapid progress in the last decade, mainly due to technological advances in unmanned aerial vehicles (UAVs). Maneuverability, lightweight, low cost, high efficiency and robust control are some characteristic that have made UAVs attractive for research in many areas of robotics [1], [2]. Applications for human-robot interaction, surveillance, aerial recording, search and rescue, exploration and manipulation in hazardous environments have received special attention from researchers [3], [4], [5].

Machine learning has played a key role for the development of intelligent robots. Particularly, probabilistic methods have shown to be robust for exploration and decision-making with multimodal sensors and robot platforms [6], [7], [8]. Deep learning techniques have also shown their potential for high accuracy in data analysis and scalability in robotics compared to the performance from traditional numerical methods [9], [10], [11]. Specially, convolutional neural networks (CNNs) have become popular, in recent years, for applications in image classification, speech and object recognition in real-time environments [12], [13], [14], [15], [16].

In this work, an active control approach for object exploration and recognition with a UAV robot is presented (Figure 1). First, a CNN module is used for object recognition with input images obtained by the robotic platform. Input images are preprocessed to segment the object to be explored. Second, a saliency map module is used to extract interesting object regions for active exploration and improvement of the recognition accuracy. The active exploration of relevant object regions is iteratively performed by the UAV until the recognition accuracy from the CNN exceeds a threshold. This process mimics the capability of the human visual system for continuously analysing images, while changing the points of visual fixation according to relevant information [17]. Thus, the fixation point from the UAV camera is autonomously and actively moved to extract better object information for recognition. This bioinspired exploration capability has also been studied using multimodal sensors and robotic platforms for multiple tasks [18], [19], [20], [21], [22], [23].

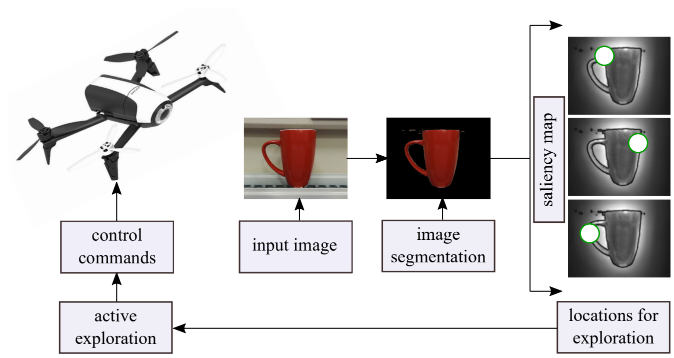


Fig. 1. Active control framework for visual object exploration and recognition with a UAV. Input images are preprocessed and analysed by a CNN and saliency map models. Interesting regions for object exploration and recognition are represented by white colour circles.

Validation of this work is performed with visual object recognition tasks in offline and real-time modes. Image data for all experiments are collected with the UAV Bebop robot (Figure 1). This robot performs the recognition of five objects while actively moving and visually exploring relevant object regions. For comparison of recognition accuracy, a passive exploration method is also used by the UAV robot. Experiments show the capability of the robot to autonomously move during the object exploration task to improve the accuracy. Results

(A) pen (B) box (C) circle (D) rectangle (E) mug

Fig. 2. Objects used for data collection, feature learning and classification with CNN and saliency map models. These objects are also employed for exploration and recognition with a UAV robot.

also show that the UAV with active exploration achieves an accuracy of 95.66%, which improves over the 88.14% accuracy with passive exploration. Overall, the active control method has the potential to develop robots capable to make autonomous decisions, about where to move and look next, for exploration and recognition tasks.

II. Methods

A. Robotic platform

In this work, the UAV Bebop was used for data collection and all experiments. This UAV robot is a lightweight quadcopter with a 856 $\times$ 480 pixels fish-eye camera and 3-axes image stabilization. The UAV transmits the live video stream of its front facing camera and pose information to a central computer over WiFi. The central computer runs the core software components of the proposed active control method for object exploration and recognition. Similarly, feedback and control commands are sent over the same WiFi link to the UAV. Figure 1 shows an example of the UAV robot while collecting an image for object recognition.

B. Data collection and preprocessing

For training the proposed method for visual object exploration and recognition, multiple datasets were collected with the camera of the UAV robot. For the data collection process, the UAV was placed at random locations closed to the objects used for exploration while flying. In total, 10,000 images with a resolution of 856 $\times$ 480 pixels were collected from a pen, box, circle, rectangle and mug (Figure 2). The collected datasets were split into training (7,000 images) and testing (3,000 images) for training and validation of the proposed method for active exploration.

The images collected with the UAV robot were preprocessed to remove the background. This segmentation process was carried out using OpenCV built-in functions. Figure 3 shows the original image from the mug, captured by the UAV, and the output image from the segmentation process. The object recognition module, implemented in the proposed active exploration method, was trained and validated using the segmented images. The object recognition process is described in the following Section II-C.

C. Active object exploration and recognition

1) Object recognition with CNN: Convolutional Neural Networks (CNN) have shown their potential for speech recognition and image classification [11], [24], [25], [26], [27]. In this work, a CNN is developed and applied to object exploration with UAV robot. The proposed CNN architecture is shown in Figure 4. The first layer uses 32 kernels of 5 $\times$ 5 and 3 $\times$ 3 sizes for convolution and max-pooling. The second layer uses 16 kernels of 3 $\times$ 3 and 3 $\times$ 3 sizes for convolution and max-pooling. The features obtained from the second layer are vectorised by the flatten and fully connected layer. Then, the softmax layer estimates the probability of the current object visually explored by the UAV robot. The CNN receives preprocessed images, of size 856 $\times$ 460, from 5 different objects collected with the UAV camera (see Figure 3). The output map from each convolutional layer in the CNN is obtained as follows:

$$x_{ij}^{l} = b_j + \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} k_{ab} * y_{(i+a)(j+b)}^{l-1} \tag{1}$$

where x_{ij}^{l} is the output of the l convolutional layer of the j-th feature map on the i-th unit. The operator $*$ denotes the convolution between the $m \times m$ kernel k_{ab} and the nonlinear output $y_{(i+a)(j+b)}^{l-1}$ from the convolutional layer $l-1$. The bias is represented by b_j. The nonlinear function σ is applied to the output from Equation (1) as follows:

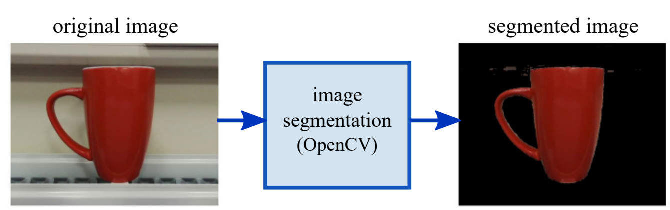


Fig. 3. Object segmentation for exploration and recognition performed with OpenCV built-in functions. The left-hand side shows the original image, captured by the camera of the UAV robot. The right-hand side shows the resulting image from the segmentation process.

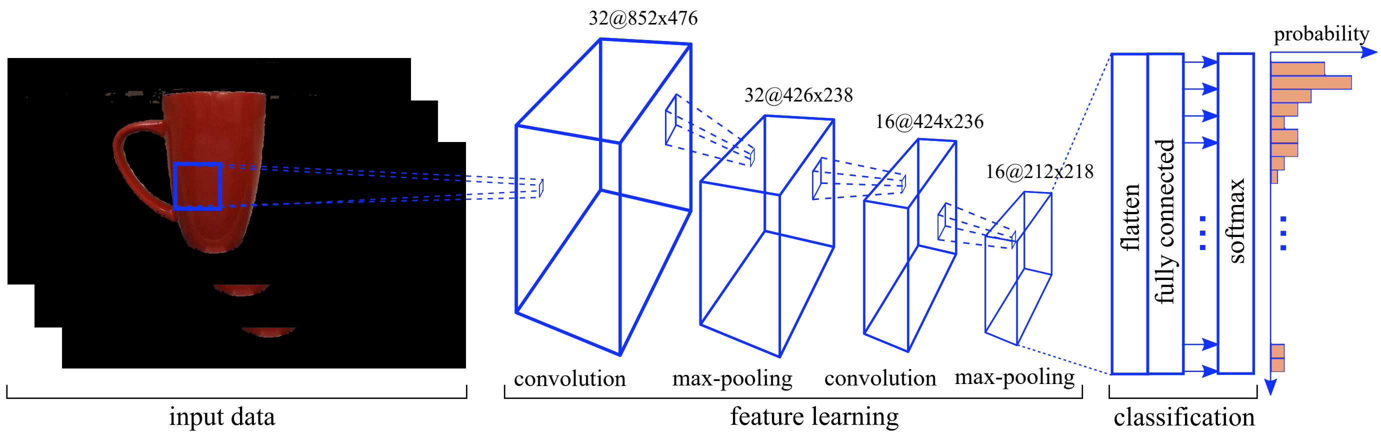


Fig. 4. Convolutional neural network for visual object recognition. This process uses a CNN model composed of 1) an input data module with images from the UAV robot, 2) a feature learning module with two convolutional and max-pooling layers and 3) a classification module with fully connected and softmax layers. The probability, from the classification layer, is used by the active exploration module to decide whether the robot needs to actively explore more interesting object regions, based on the saliency map module, in order to achieve a better recognition accuracy.

$$y_{ij}^l = \sigma(x_{ij}^l) \tag{2}$$

where y_{ij}^l is the nonlinear output from the l convolutional layer and σ is the hyperbolic tangent function $\tanh$. A down-sampling process, for dimensionality reduction, is performed with a max-pooling layer after each convolutional layer. This process takes a $u \times u$ region (3×3 size for our proposed CNN architecture) and outputs the maximum value from that region as follows:

$$y_{ij}^l = \max_{u \times u}(y_{ij}^{l-1}) \tag{3}$$

where y_{ij}^l contains the maximum values from the nonlinear output y_{ij}^{l-1} from the previous layer. The process performed by the convolutional and max-pooling layers is known as feature learning, and its output from the second layer of the CNN are connected to a 1-dimensional feature vector y_c. This vector is used by a softmax function for classification, as follows:

$$P(c|y) = \frac{e^{y^T w_c}}{\sum_{n=1}^{N} e^{y^T w_n}} \tag{4}$$

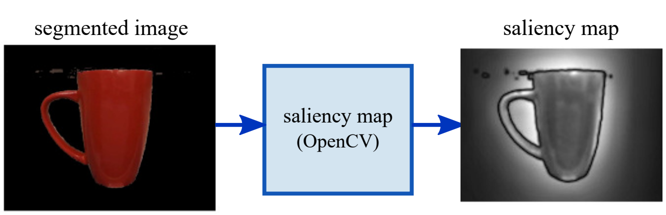


Fig. 5. Saliency map genrated from the segmented image. The saliency map is used to actively control the UAV during the object exploration and recognition task. In this work, the spectral residual saliency map algorithm was employed with OpenCV built-in functions.

$$\hat{c} = \arg\max_c P(c|y) \tag{5}$$

where $P(c|y)$ contains the probabilities for all object classes (pen, box, circle, rectangle and mug), given the sample vector y. The parameters w and N represent the weight vector and total number of classes, respectively. In Equation (5), the recognition of the current object, $\hat{c}$, is obtained with the *maximum a posterior* (MAP) estimate.

The output from the CNN is used to actively control the movements performed by the UAV robot. This process, described in Section II-C2, allows the robot to autonomously explore an object while improving the recognition accuracy.

2) Active exploration with saliency maps: Visual saliency has been used for different applications in robotics [28], [29]. Most robot applications rely on the extraction of local information of salient features. In this work, visual saliency map is employed as an active method for extraction of interesting object regions for active exploration with the

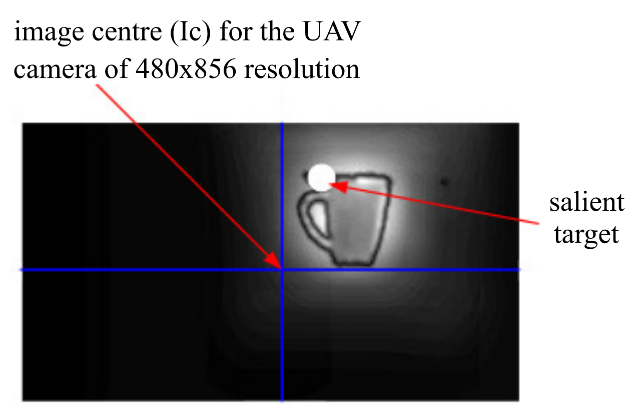
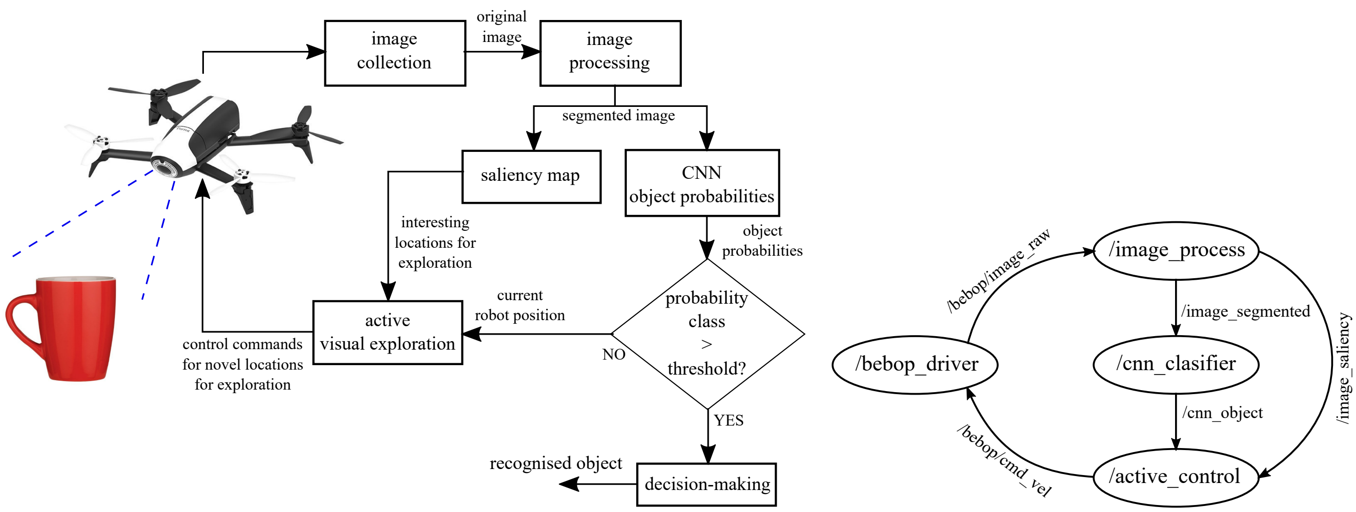


Fig. 6. Estimation of the trajectory to be actively followed by the UAV based on interesting object regions obtained with the saliency map model.



(A) Active UAV control architecture

(B) ROS nodes and topics for communication and control

Fig. 7. Architecture for active UAV control. A) Processes implemented in the UAB for active object exploration and recognition using vision sensing. This approach allows the UAV robot to autonomously identify and move towards interesting object regions to improve the recognition accuracy. B) ROS graph composed by the nodes and topics used for all the processes during the active object exploration and recognition task.

UAV robot. Here, interesting object regions are represented by groups of pixels with high intensity levels, which are employed for active control and repositioning of the UAV during the exploration process. Here, images from the UAV Bebop camera are captured at every 30 ms. Then, the processes for object segmentation and extraction of object regions for active exploration, with saliency map method, are continuously performed. Figure 5 shows an example of the saliency map obtained from the mug object. In this work, the saliency map was implemented using the spectral residual algorithm with C++ language and OpenCV functions.

Once the saliency map is obtained from an input image, it is used to actively control the exploration movements of the UAV robot. This is an iterative process where the UAV position is continuously updated, in x- and y-axes, to gradually reach the object region. This process is based on the estimation of the distance, $D(x_d, y_d)$, between the centre of the camera of the UAV robot, $I_c(x_c, y_c)$ where $x_c = \frac{856}{2}$ and $y_c = \frac{480}{2}$, and the exploration region, $S_r(x_r, y_r)$, obtained from the saliency map (Figure 6). Then, the distance, D, and number of movements, $nSteps$, performed by the UAV to reach the salient region S_r is calculated as follows:

$$D(x_d, y_d) = S_r(x_r, y_r) - I_c(x_c, y_c) \tag{6}$$

$$nSteps(x_s, y_s) = floor\left(\frac{x_d}{\Delta step_x}, \frac{y_d}{\Delta step_y}\right) \tag{7}$$

where D is the distance between the centre of the UAV camera and the object region for exploration, $nSteps$ is the number of movements to reach the salient region, and $\Delta step_x$ and $\Delta step_y$ are the step size in x- and y-axes performed by the UAV robot. It was found that the smallest step that the UAV robot can performed, in $\Delta step_x$ and $\Delta step_y$, is 10 pixels. This allows the robot to perform fast and smooth active exploration movements.

Every time the UAV position is updated during the exploration process, an object image is captured and sent to the CNN model for recognition of the object being explored. This exploration process is repeated until a decision threshold is exceeded by the CNN model. The robustness of this threshold crossing approach has been demonstrated with a variety robot exploration tasks [30].

The control architecture, with all modules for active visual object exploration and recognition with a UAV, is shown in Figure 7A. This architecture also shows the interaction between all modules and iteration of processes for control of the UAV in real-time. In this work, C++ and Python programming language were used for implementation of the data collection process, image segmentation, saliency map, object regions for active exploration, CNN model and control commands. All these processes were communicated and synchronised using the Robot Operating System (ROS), which has become the standard open-source middleware for robotics [31], [32]. The nodes and topics employed for the control architecture are shown in Figure 7B.

III. Results

A. *Offline visual object recognition*

Validation of the object recognition process was performed in offline mode. For this experiment, images collected with the UAV Bebop robot were used for recognition of multiple objects (see Section II-B). Input images from the UAV robot were preprocessed removing the background to keep the object to be explored only (see Figure 3).

The CNN model, proposed for object recognition is shown in Figure 4. The performance of this CNN model was validated

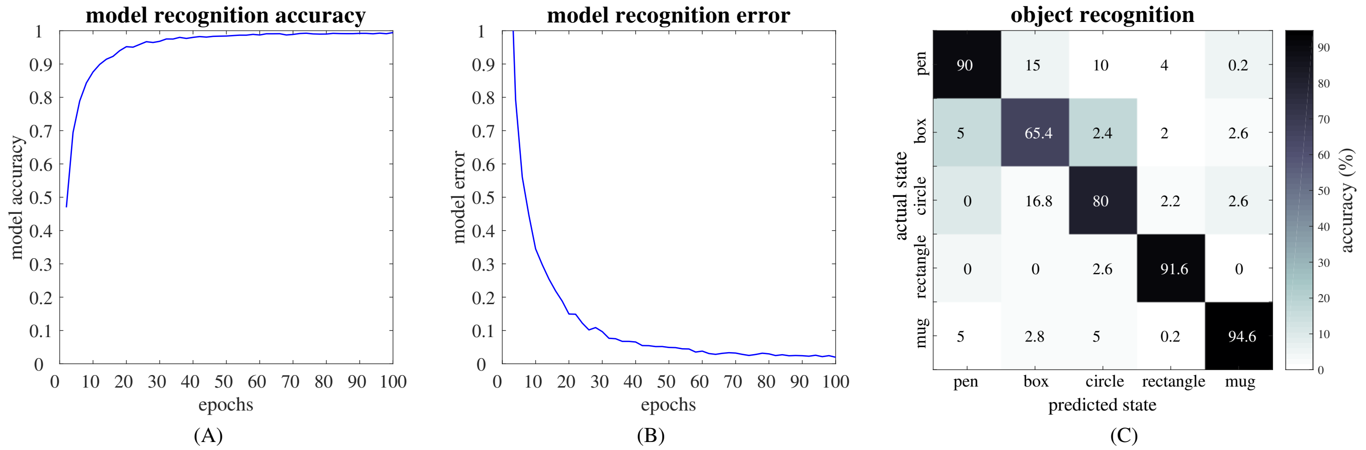


Fig. 8. Results from object recognition in offline mode. (A), (B) Accuracy and error of the CNN model, during the recognition of multiple objects, using real data collected with the UAV robot. (C) Recognition accuracy results for each object using passive exploration in offline mode.

using 10,000 images of 5 different objects. In this experiment, images were recognised in a passive mode, which means that the robot did not move to different object regions, during the exploration process, to improve the recognition accuracy. The results from the accuracy recognition and loss function using the CNN model are shown in Figures 8A and 8B, respectively. These results were achieved with 100 epochs. The recognition model achieved an accuracy of 98.95%, while the loss function achieved an error of 2.2% using the training datasets. An accuracy of 84.32% was achieved for recognition of all objects using the testing datasets. Figure 8C shows the confusion matrix with the accuracy results from the recognition of each explored object: pen, box, circle, rectangle and mug. The results show that recognition of the mug and box achieved the highest and lowest accuracies with the UAV robot in offline mode. The overall recognition accuracy could be affected by the amount of data employed for training. However, nowadays, it is possible to collect and have access to large datasets given the advances in sensor technology.

The performance of this experiment, in passive mode, can be improved by the used of the proposed active exploration module. The experiments and results with the active module, which allows the robot platform to autonomously explore interesting object regions to achieve a better recognition accuracy, are described in the following Section III-B.

B. *Real-time visual object recognition*

The CNN model together with the active exploration model were validated with object recognition tasks in real-time. For this experiment, the UAV robot performed the object recognition task with the following procedure. The robot was placed at random initial locations closed to the object to be visually explored and recognised (Figure 9A). The images collected in real-time, with the UAV camera, were used as input to the CNN model for recognition. The saliency map module was employed to identify interesting object regions for exploration, in order to improve the recognition accuracy. The saliency map used the same input images as the CNN architecture. Figures 9B-J shows examples of sequences of object regions, from the mug, circle and box objects, obtained by the saliency map model for active object exploration and recognition with the UAV robot.

Outputs from the CNN and saliency map models were the responsible to control the active exploration and decision-making of the object employed for recognition. First, the object probability from the CNN model was compared to a predefined decision threshold. Second, if the object probability did not exceed the predefined decision threshold, then, the robot was actively moved to another object region in order to improve the recognition probability from the CNN model. Active movements or object regions for exploration were obtained with the saliency map module for each image capture by the UAV robot. Third, a decision about the object being explored was made when the probability from the CNN exceeded the decision threshold. Once the object being explored was recognised, the robot landed and waited for the command to start the active exploration of the next object.

Recognition accuracy results, from passive and active object exploration, are presented by the confusion matrices in Figure 10. In passive exploration mode, the robot was placed at random initial locations closed to the object. However, the robot performed the decision-making process without moving to other object regions to improve the accuracy. Results from the passive recognition approach achieved a mean accuracy of 88.14% (Figure 10A). In this passive exploration experiment, the box and pen objects achieved the lowest and highest accuracies with 80% and 95%, respectively. These results contrast with the performance obtained by the UAV robot using the active object exploring approach. In this case, the robot was also placed at a random initial location closed to the object. However, this time, the robot was capable to identify interesting object regions for exploration, based on the saliency map module, which were employed by the UAV to actively move and improve the recognition accuracy. Results from the

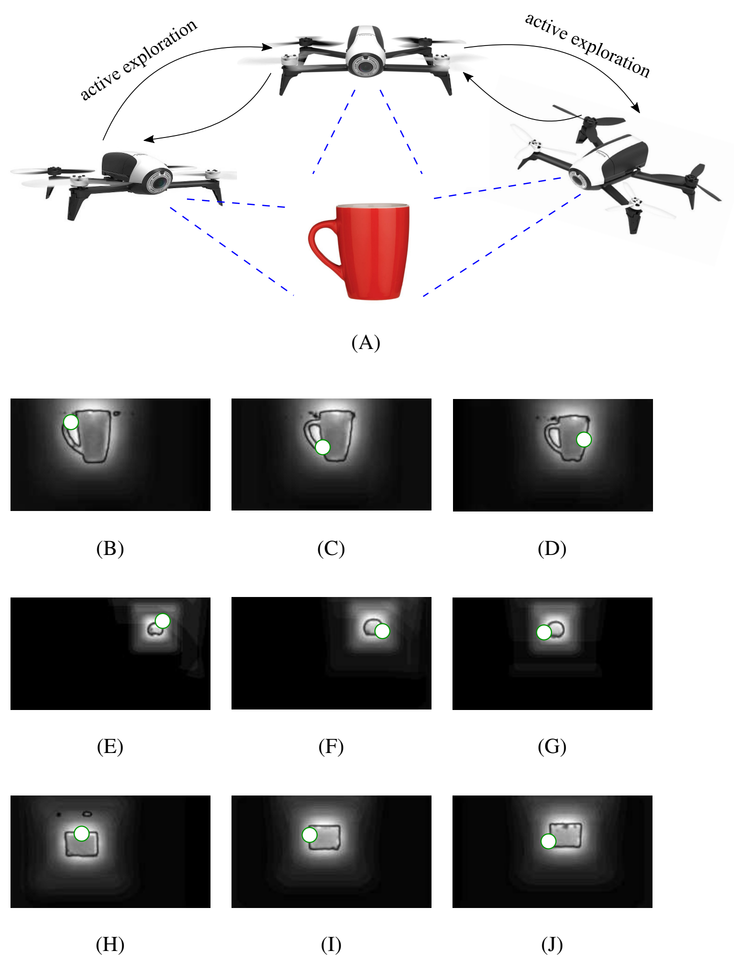


Fig. 9. (A) Real-time sequence of object regions used by the UAV robot for active exploration and recognition. Object regions are represented by white colour circles. (B)-(D) Object regions actively explored by the robot for recognition of the mug. (E)-(G) Object regions actively explored by the robot for recognition of the circle. (H)-(J) Object regions identified on the box for active exploration and recognition.

active exploration and recognition process achieved a mean accuracy of 95.66% (Figure 10B). In this experiment, the circle and pen objects achieved the lowest and highest accuracies with 92% and 100%, respectively. These results show that active exploration allowed the UAV robot to improve the recognition accuracy for all objects, over the results achieved with passive exploration.

This active exploration approach is inspired by the way in that humans explore an scene with their eyes, looking for key locations to extract better information. Here, the UAV moves towards salient object locations to extract information that can improve the recognition accuracy. Overall, experiments in offline and real-time showed that the object recognition accuracy with the CNN model can be improved, actively moving the UAV robot towards better object regions.

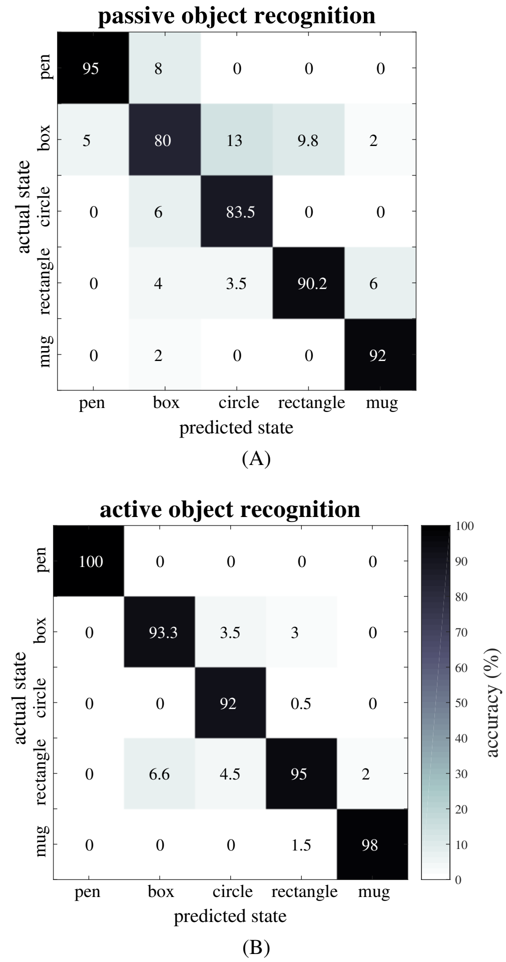


Fig. 10. Confusion matrices with object recognition results from the real-time experiment. (A) Passive object recognition accuracy with the UAV. (B) Active object recognition improved with the active recognition approach.

IV. Conclusion

In this work, an approach for active object exploration and recognition with vision sensing from an UAV robot was presented. This approach, composed of multiple modules, allowed the UAV robot to make autonomous decisions about where to move and look next, during an object exploration task, to obtain better information and improve the recognition accuracy. First, input images from the robot were preprocessed, segmenting the object to be explored and identifying interesting object regions for exploration using a saliency map module. Second, the object exploration task was implemented in passive and active modes for comparison of performance. For the experiment in passive mode, where the UAV robot was not able to move to better object regions, recognition accuracies of 84.32% and 88.14% were achieved in offline and real-time modes, respectively. These low accuracies are related to the inability of the robot to autonomously explore object regions that contain better information. In contrast, in the active exploration experiment, the UAV was capable to make autonomous decisions about where to move and look next to collect better information, and thus, to improve the object recognition task. This capability offered by the active process was reflected with the improved accuracy of 95.66%. Overall, the proposed active object exploration method, offers a framework suitable for the development of intelligent UAVs, capable to make autonomous decisions and actions while interacting with the environment.

REFERENCES

[1] A. L. Salih, M. Moghavvemi, H. A. Mohamed, and K. S. Gaeid, "Flight pid controller design for a uav quadrotor," *Scientific research and essays*, vol. 5, no. 23, pp. 3660–3667, 2010.

[2] S. Gupte, P. I. T. Mohandas, and J. M. Conrad, "A survey of quadrotor unmanned aerial vehicles," in *Southeastcon, 2012 proceedings of ieee*. IEEE, 2012, pp. 1–6.

[3] N. Smolyanskiy, A. Kamenev, J. Smith, and S. Birchfield, "Toward low-flying autonomous mav trail navigation using deep neural networks for environmental awareness," *arXiv preprint arXiv:1705.02550*, 2017.

[4] J. Apaza, D. Scipión, D. Lume, and C. Saito, "Development of two uavs for volcano studies in southern peru," in *Electronics, Electrical Engineering and Computing (INTERCON), 2017 IEEE XXIV International Conference on*. IEEE, 2017, pp. 1–4.

[5] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the sky: Leveraging uavs for disaster management," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 24–32, 2017.

[6] U. Martinez-Hernandez and A. A. Dehghani-Sanij, "Adaptive bayesian inference system for recognition of walking activities and prediction of gait events using wearable sensors," *Neural Networks*, vol. 102, pp. 107–119, 2018.

[7] T. J. Prescott, D. Camilleri, U. Martinez-Hernandez, A. Damianou, and N. D. Lawrence, "Memory and mental time travel in humans and social robots," *Philosophical Transactions of the Royal Society B*, vol. 374, no. 1771, p. 20180025, 2019.

[8] J. Ferreira, J. Lobo, P. Bessiere, M. Castelo-Branco, and J. Dias, "A bayesian framework for active artificial perception," *IEEE transactions on cybernetics*, vol. 43, no. 2, pp. 699–711, 2013.

[9] B. K. Kim, H.-S. Kang, and S.-O. Park, "Drone classification using convolutional neural networks with merged doppler images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 38–42, 2017.

[10] M. A. Olivares Mendez and P. Campoy, "Vision based fuzzy control approaches for unmanned aerial vehicles," in *9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*, 2015.

[11] A. Rubio-Solis and G. Panoutsos, "Interval type-2 radial basis function neural network: a modeling framework," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 2, pp. 457–473, 2015.

[12] I. Ševo and A. Avramović, "Convolutional neural network based automatic object detection on aerial images," *IEEE geoscience and remote sensing letters*, vol. 13, no. 5, pp. 740–744, 2016.

[13] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy, "Visual odometry and mapping for autonomous flight using an rgb-d camera," in *Robotics Research*. Springer, 2017, pp. 235–252.

[14] U. Martinez-Hernandez, A. Rubio-Solis, and A. A. Dehghani-Sanij, "Recognition of walking activity and prediction of gait periods with a cnn and first-order mc strategy," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 2018, pp. 897–902.

[15] N. Attari, F. Ofli, M. Awad, J. Lucas, and S. Chawla, "Nazr-cnn: Fine-grained classification of uav imagery for damage assessment," in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2017, pp. 50–59.

[16] N. Chumerin, A. Gibaldi, S. P. Sabatini, and M. M. Van Hulle, "Learning eye vergence control from a distributed disparity representation," *International journal of neural systems*, vol. 20, no. 04, pp. 267–278, 2010.

[17] J. J. Gibson, "The perception of the visual world." 1950.

[18] U. Martinez-Hernandez, T. J. Dodd, and T. J. Prescott, "Feeling the shape: Active exploration behaviors for object recognition with a robotic hand," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017.

[19] U. Martinez-Hernandez, T. J. Dodd, M. H. Evans, T. J. Prescott, and N. F. Lepora, "Active sensorimotor control for tactile exploration," *Robotics and Autonomous Systems*, vol. 87, pp. 15–27, 2017.

[20] A. J. Davison and N. Kita, "3d simultaneous localisation and map-building using active vision for a robot moving on undulating terrain," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–I.

[21] F. Rea, G. Metta, and C. Bartolozzi, "Event-driven visual attention for the humanoid robot icub," *Frontiers in neuroscience*, vol. 7, p. 234, 2013.

[22] M. Antonelli, A. P. Del Pobil, and M. Rucci, "Bayesian multimodal integration in a robot replicating human head and eye movements," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 2868–2873.

[23] M. Antonelli, A. Gibaldi, F. Beuth, A. J. Duran, A. Canessa, M. Chessa, F. Solari, A. P. Del Pobil, F. Hamker, E. Chinellato, *et al.*, "A hierarchical system for a distributed representation of the peripersonal space of a humanoid robot," *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 4, pp. 259–273, 2014.

[24] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Transactions on audio, speech, and language processing*, vol. 22, no. 10, pp. 1533–1545, 2014.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[26] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "Car detection from low-altitude uav imagery with the faster r-cnn," *Journal of Advanced Transportation*, vol. 2017, 2017.

[27] L. Carnimeo, "A cnn-based vision system for pattern recognition in mobile robots," in *Proc. of the 15th IEEE European Conf. on Circuit Theory & Design, Espoo, Finland*, 2001.

[28] L. Itti, "Models of bottom-up attention and saliency," in *Neurobiology of attention*. Elsevier, 2005, pp. 576–582.

[29] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[30] U. Martinez-Hernandez and T. J. Prescott, "Adaptive perception: Learning from sensory predictions to extract object shape with a biomimetic fingertip," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 6735–6740.

[31] S. Mokaram, J. M. Aitken, U. Martinez-Hernandez, I. Eimontaite, D. Cameron, J. Rolph, I. Gwilt, O. McAree, and J. Law, "A ros-integrated api for the kuka lbr iiwa collaborative robot," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 15 859–15 864, 2017.

[32] A. Koubâa, *Robot Operating System (ROS)*. Springer, 2017.