Use of semantic description of reference scene for visual navigation solutions

N E Bodunkov and M O Pindiurina
Moscow Aviation Institute, 4, Volokolamskoe Avenue, Moscow, 125993, Russia
E-mail: mapinol@rambler.ru

Abstract. This article is devoted to the description generation of reference scene applying to visual navigation problem. The authors investigate the semantic descriptions that are more resistant to changing conditions of observation, thus being an alternative to existing approaches that use raster, feature, or object descriptions. There are no publications, considering similar approach of utilization for visual navigation problem solution. In the article the original description structure is proposed. The experiments involving descriptions of various scenes under various conditions have been conducted to assess the workability of the proposed approach. It is suggested that semantic descriptions are more resistant to partial occlusion of scene objects or changes in their visual features than other description types and can be used to solve visual navigation problems.

1. Introduction
The problem of visual navigation is currently a topical issue in the light of advances in synthetic vision systems [1]. Visual navigation is a process of comparing descriptions of a current image (CI) received from the onboard observation system with reference descriptions of scenes or objects (landmarks) with known coordinates to identify the current location [2]. The application conditions and the solution quality directly depend on description type and structure.

The modern algorithms, including those that involve neural networks [3], allow identifying the most stable and informative combinations of characteristic features of references to form so-called object descriptions. However, those descriptions are insufficient for solving a number of applied problems.

Semantics is becoming increasingly frequently mentioned in scientific reports related to description generation. However, there are no known issues which provided examples of their applications for visual navigation.

2. Materials and methods
Historically, raster images and correlation comparison algorithms that use them are the basic description type [4]. However, changeable observation conditions (such as occlusions, change of seasons, precipitations and other) result in altered visual characteristics of landmarks, which, in turn, increases numbers of detection errors. At the same time, raster images can be characterized as the most redundant, and, consequently, the most resource-consuming in terms of storage and comparison.

The description types based on identification of characteristic features, e.g., dots, lines, areas, etc [5] are more informative and resistant to altered observation conditions. Using characteristic features considerably reduces memory volumes, compared to pixel-by-pixel image storage, thus speeding up the algorithm operation rate [6]. Most of those algorithms, however, are sensitive to high-frequency noises,
which is why characteristic feature sets and their combinations are used to enhance algorithm stability.

Semiotics is commonly used for the description of scenes received from synthetic vision systems of UAVs. E.g., work [7] proves that the a proposed method for semantic description formulation, involving the PLSA (Probabilistic Latent Semantic Analysis) model is efficient, and can identify, to a high accuracy, the areas of an image, including soil, wooded areas, house roofs, and water surfaces. The work [8] proves that using an algorithm based on deep learning, the average accuracy of recognition of buildings located along a river shore is 90%.

The above mentioned works commonly define semantic descriptions as marking (or classification) of image fragments by using a prepared vocabulary. However, both identification of objects on a scene and creation of a description of a whole scene are essential to solving the research problems. For example, the article [9] discusses a problem of human behavior recognition from an image. The work [10] asserts that image processing involves not only local recognition of objects, but space information also—so-called semantic context that allows linking local features of objects with interactions between them to provide a general idea of a scene.

The overview of modern trends in research suggests that semantic descriptions are commonly used to identify individual objects and to analyze complex scenes in order to provide solutions to various practical problems. It is fair to assert that semantic descriptions that have been proposed so far fail to provide solutions to visual navigation problems. The authors discuss the structure of a semantic description of a scene in order to solve a visual navigation problem.

3. Structure of semantic description of a reference scene

A reference scene (also called a landmark) is an area that contains one or more objects, by unambiguous recognition of which the UAV current coordinates may be identified to an accuracy.

The proposed structure of semantic description is a set of descriptions of individual objects of a scene, which are required for solving a problem, and descriptions of relations between them.

Semantic descriptions are of a hierarchic nature, and consist of descriptions of landmarks (map objects) and relations between them.

The proposed object attributes are as follows: shape class (Square (S), rectangle (R), Parallel Lines (PL)), dimensions (Large (l), Medium (m), Small (s)), and the north turn angle.

Relation description structure contains relation instruction, its value, and a list of bound objects. The used relations are as follows: distance (Very Close (vc), Close (с), Far (f), Very Far (vf)), and orientation (Parallel (pl), Perpendicular (pr), and Diagonal (diag)).

4. Assessment results for semantic description workability

In order to assess the possibility of using descriptions proposed for visual navigation, experiments were conducted to calculate recognition accuracy of a scene being observed (CI).

A base map image (the reference image, RI) was chosen as a test scene (Figure 1). Characteristic scene objects were identified, and a semantic description was built, based on them (Table 1). Further on, CIs of the same scene were considered under various outer conditions (Figure 2). The recognition accuracy coefficient was calculated for each CI.

The normalized coefficient of recognition accuracy of a CI was found from the formula below, similar to the Hamming distance

\[ K_P = \frac{1}{n} \sum_{i=1}^{n} K_{pi} \]

where \( n \) was the number of cells in the semantic description table (Table 1); and \( K_{pi} \) was the coefficient of coincidence of a RI cell with its respective CI cell, its value being 0 when cells did not coincide, or 1 when they coincided.

The similar calculations were done for two other scenes (Figure 3 and Figure 4) for which Figure 2.a and Figure 3.a were chosen as their respective reference images.

Figure 5 shows diagrams of the normalized accuracy coefficients found for altered outer conditions.
Figure 1. Reference image of Scene 1 with characteristic objects highlighted (1 – River; 2, 3, and 4 – Roads; 5 – Waterbody)

Table 1. Semantic description of Scene

| Objects | 1 River | 2 Road 1 | 3 Road 2 | 4 Road 3 | 5 Waterbody |
|---------|---------|----------|----------|----------|-------------|
| 1 River | “River” Large $\alpha=\pi 3/4$ | C diag | F diag | F pr | C diag |
| 2 Road 1 | C diag | “Road” Small $\alpha=\pi$ | Vc pl | Vc pr | Vc pr |
| 3 Road 2 | F diag | Vc pl | “Road” Med $\alpha=\pi$ | Vc pr | C pr |
| 4 Road 3 | F pr | Vc pr | Vc pr | “Road” Med $\alpha=\pi/2$ | Vc pl |
| 5 Waterbody | C diag | Vc pr | C pr | Vc pl | “Waterbody” Large $\alpha=\pi/2$ |
Figure 2. Current images of Scene 1, obtained under various observation conditions: a – season “Summer”, b – season “Winter”, c – season “Fall”, d – season “Spring”.

The authors consider calculation of $K_p$ for CI4, by way of illustration.

Table 2. Semantic description of CI4.

| Object         | 1 River | 2 Road 1 | 3 Road 2 | 4 Road 3 | 5 Waterbody |
|----------------|---------|----------|----------|----------|-------------|
| 1 River        | “River” | C        | F        | F        | Vc diag     |
|                | Large   | diag     | diag     | pr       |             |
|                | $\alpha=\pi$ 3/4 |          |          |          |             |
| 2 Road 1       | C       | Vc       | Vc       | Vc       | Vc diag     |
|                | diag    | pl       | pr       | pr       |             |
| 3 Road 2       | F       | Vc       | “Road”   | Vc       | Vc diag     |
|                | diag    | pl       | Med      | pr       |             |
|                | $\alpha=\pi$ |          |          |          |             |
| 4 Road 3       | F       | Vc       | Vc       | “Road”   | Vc diag     |
|                | pr      | pr       | Med      | pr       |             |
|                | $\alpha=\pi$ 2/3 |          |          |          |             |
| 5 Waterbody    | Vc      | Vc       | Vc       | Vc       | “Field”     |
|                | diag    | pr       | pr       | pl       | Large       |
|                |          |          |          |          | $\alpha=\pi$ 2/3 |

$K_p (CI4) = 20/25 = 0.8$

$K_p (CI1) = 1; K_p (CI2) = 0.9; K_p (CI3) = 0.96$. 

4
Scene 2

Figure 3. Current images of Scene 2, obtained under various observation conditions: a – season “Spring”, b – season “Summer”, c – season “Summer” cloudy, d – season “Fall”. $Kp(CI1)=1$; $Kp(CI2)=0.9$; $Kp(CI3)=0.36$; $Kp=0.64$.

Scene 3

Figure 4. Current images of Scene 3, obtained under various observation conditions: a – season “Spring”, b – season “Winter”, c – season “Fall”, d – season “Summer” cloudy. $Kp(CI1)=1$; $Kp(CI2)=0.64$; $Kp(CI3)=0.92$; $Kp=0.96$.

Figure 5. Normalized coefficients of recognition accuracy against numbers of current images

Reasoning from theoretical premises, the main consequence of nanopowder introduction into the melt terminating a spericity coefficient and an average size for them.

5. Discussion

It can be seen from Figure 5, Scene 1, that the “waterbody” object changes its visual features greatly, including texture, geometrical shape, size, etc, when seasons change. This, however, practically does not affect general scene recognition, as visual features of a number of other characteristic objects (e.g., the river and the roads) and relations between them remain unchanged, and therefore recognizable. As a result, the normalized coefficient of recognition accuracy decreases almost 1.5 of total coincidence, which can be clearly seen on diagram above.

The clouds blocking the view are the main destabilizing factor of Scene 2. For example it is possible to see that on CI3 the clouds are blocking approx 60% of characteristic objects. The RI and EI3 are
obviously hard to compare; however, the above mentioned factor decreases the semantic recognition accuracy three times, while pixel-by-pixel comparison between similar scenes would have resulted in greater losses: e.g., the normalized correlation coefficient would have decreased 77 times.

The precipitations (snow) play an important role in Scene 3, as they heavily affect visual features of all the objects visible. The above diagram suggests that the criterion decreased by 0.36, which was approx 1.5 times.

All in all, the conducted experiments have proved the possibility of using the description structure proposed here for navigation problems within a wide range of application conditions. However, research and comparison with other description types, e.g., raster descriptions, are required to assess quality of the results obtained.

6. Conclusion
The semantic descriptions of objects, attributes, and relations between them have been proposed in this article as an alternative to raster imaging of reference descriptions.

A new version of semantic description structure and a method for its formulation have been proposed. The experiments have proved workability of the approach proposed and its resistance to various destabilizing factors, including blocked view, and altered observation conditions.

7. Acknowledgments
This research was conducted with financial support from the Russian Foundation for Basic Research (RFBR) as part of Scientific Project No. 19-08-00613 A.

References
[1] Lupanchuk V Y 2018 The Development of Navigation Cartography Techniques for the Control of Spatial Positioning of UAVs Moscow Aviation Institute Herald 25(1) 132–42
[2] Veremeenko K K, Zheltov S Y, Kim N V, Kozorez D A, Krasilshikov M N, Sebryakov G G, Sypalo K I and Chernomorsky A I 2009 Modern Information Technologies Used in Problems of Navigation and Targeting of Maneuverable Unmanned Aircraft (Moscow: Fizmatlit)
[3] Bodunkov N E 2015 The Extension of Operation Conditions of Visual Navigation Systems of Standalone Unmanned Aircraft PhD Dissertation (Moscow)
[4] Volkov V, Germer R, Oneshko A and Oralov D 2011 Object description and extraction by the use of straight line segments in digital images Int. Conf. on Image Processing, Computer Vision and Pattern Recognition (IPCV’2011) pp 588–94
[5] Vizilter Yu, Rubis A, Vygolov O and Zheltov S 2018 Vision-Based Change Detection Using Comparative Morphology Computer Vision in Control Systems-3 Aerial and Satellite Image Processing Intelligent Systems Reference Library 135 65–9
[6] Kim N V, Bodunkov N E and Cherkezov R I 2015 Ensuring the accuracy of traffic monitoring using unmanned aerial vehicles vision systems Modern Applied Sci. 9(5) 323–36
[7] Wenbin Y and Hong T 2009 Experimental analysis on classification of unmanned aerial vehicles images using the Probabilistic Latent Semantic Analysis Proc. of SPIE The Int. Society for Optical Engineer.
[8] Boonpook W, Tan Yu, Ye Y, Torteeka P, Torsri K and Shengxian D A 2018 Deep Learning Approach on Building Detection from UAV-Based Images in Riverbank Monitoring Sensors 18 3921
[9] Vinicius K B P, Conci N and Cavallaro A 2013 Video-based human behavior understanding a survey Transactions on Circuits and Systems for Video Technology 23(11)
[10] Messing R 2011 Human Activity Recognition in Video: Extending Statistical Features Across Time Space and Semantic Context (New York: University of Rochester)