Visual Semantic Context of Semantically Segmented Aerial Images

Junwoo Park, Sungjoong Kim, Kyungwoo Hong, and Hyochoong Bang

Abstract This paper proposes a descriptor for the position and heading of aerial vehicles out of down-looking images using the concept of the visual semantic context aided by semantic segmentation and semantic labelled map which can be utilized during aerial navigation. The study presents the derivation of the visual semantic context from the given image while also conducting a feasibility analysis of the visual semantic context by presenting the corresponding error characteristics using both information- and heuristic-based residual metrics over the two contexts. The analysis using semantically segmented aerial images and the semantic labelled map indicates that the proposed concept shows distinct numeric features in a local sense and that it can be utilized as a position and heading fixing tool if associated with exhaustive search and/or data assimilation methods. The local uniqueness of the proposed context also implies that it is possible to use the concept as a validity index for a given aerial image when combined with a filtering paradigm.

Keywords Vision-based navigation · Semantic segmentation · Position descriptor · Visual semantic context

1 Introduction

While inertial navigation system (INS) serves as the primary solution for the position, velocity, and/or attitude of aerial vehicles, it is common that position-fixed information from supporting systems such as a global navigation satellite system (GNSS) should be periodically fed back to the INS for calibration and for improved long-term reliability. A typical drawback of a GNSS, however, is that it is vulnerable to the signal reception condition, particularly critical issue in military applications. Database-referenced navigation (DBRN) can replace the GNSS in the sense

J. Park · S. Kim · K. Hong · H. Bang (✉)
Korea Advanced Institute of Science and Technology, Daejeon, Korea
e-mail: hcbang@kaist.ac.kr

J. Park
e-mail: junwoopark@kaist.ac.kr

S. Lee et al. (eds.), *The Proceedings of the 2021 Asia-Pacific International Symposium on Aerospace Technology (APISAT 2021), Volume 2*, Lecture Notes in Electrical Engineering 913, https://doi.org/10.1007/978-981-19-2635-8_10

that the dependency of the GNSS upon external signals and/or systems, i.e., satellites, is replaced with an onboard database. The idea of matching sensory features with those of such a database is also straightforward [4]. Especially among various DBRN methods, the utilization of aerial images in conjunction with a geo-referenced database provides a powerful localization tool for aerial vehicles, even when the conventional position-fixing systems, a GNSS for instance, are subjected to intended signal interference [16] such as jamming, or spoofing. Visual map navigation, also known as visual navigation, requires only a self-contained vision sensor which is operationally inexpensive and already available from use in most aerial vehicles. Its operation is independent of external systems, robust to deliberate sabotage, and less likely to be detected by opponents, in contrast to alternative DBRN systems such as terrain-referenced navigation (TRN), which makes use of active sensors such as a radar altimeter, and which still has signal problems. Moreover, publicly available map databases or a geographic information system (GIS) that can relate ground features with their corresponding locations, such as Google Maps, OpenStreetMap, and VWorld [15], make visual navigation aided by a geo-referenced feature map more appealing than alternatives, as the application of these systems are not limited nor restricted to military purposes. Several researchers have utilized public map databases in their own studies. Two studies [17, 19], for instance, utilized Google Maps for image registration using histograms of oriented gradients (HOG), and for scene matching using classical feature extractors, respectively.

The performance of visual navigation relies on feature extraction, finding correspondences, and/or the coordination algorithm of the processed information, e.g., filtering and, matching, for data assimilation. In order for visual navigation to be robust over the long term, however, feature extractors should extract time- and climate change-invariant features from aerial images. Classical feature extractor-based navigation is associated with scaling and rotation [9] and is sensitive to flight and/or shooting conditions [5]. Thus, it makes sense for the feature extractor to focus on abstract features such as artifacts, e.g., roads, junctions, buildings, and/or natural objects such as farm fields, rivers, and mountain drainage areas, which rarely vary or vary slowly over centuries such that the consistency with the map database lasts longer. Studies of visual navigation using abstract features of an aerial image, patterns of it, or landmarks within an aerial image include matching road intersections [3], mountain drainage patterns [18], and salient man-made infrastructures [14]. They show promising results regardless of seasonal conditions.

Recent progress on semantic segmentation using a fully convolutional network [12] has enabled the training of an image segmentation network that outputs pixel-wise semantic of urban aerial images such as roads and buildings [13]. One of these studies utilized a Gaussian mixture model to approximate dim borders across semantics using a clustering method. The noisy characteristics of position estimation were smoothed using a particle filter [5]. An alternative approach utilized patterns between the centers of the segmented semantics [8]. Direct matching of two semantically segmented images [7] is also possible by augmenting pixel-wise semantics as an additional dimension for the image and applying the iterative closest point (ICP) algorithm to find the translation and rotation that yields the best fit between the two

with the least squared error. Nevertheless, such approaches are subject to false correction and/or false matching caused by corrupted measurement or ill-posed conditions of visual matching. Reliable operation of a navigation system aided by a Kalman- or nonparametric Bayesian filter requires an indicator capable of telling whether the given measurement, i.e., the processed image in this case, is valid and/or may cause a failure of the integration filter. Moreover, [5, 8] necessitate additional clustering methods such as mean-shift clustering [2], thus demanding costly work in addition to already expensive semantic segmentation.

As a remedy for these problems, we propose a descriptor for the position and heading of aerial vehicles upon semantic level using semantically segmented aerial images. The descriptor realized by a visual semantic context (VSC) can provide a validity index for semantic-based navigations and can be directly utilized in localization problems when served and matched with a semantically labelled map database. From the observation that the radial distribution of high-level semantics with respect to the given position and heading shows a distinct numeric pattern over the spatial and orientation domain, the VSC is derived to deliver the concept. The detailed process is described in the following section.

The rest of the paper begins with a definition of the VSC, as presented in Sect. 2. In Sect. 3, several information- and heuristic-based comparison methods that measure similarities or differences between the two contexts are introduced. Section 4 shows the numerical error characteristics of the VSC when analyzed using the aforementioned measures. The results delineate the residual distribution in both the spatial and orientational domains. Finally, Sect. 5 summarizes and concludes the paper.

2 Visual Semantic Context

This section describes the concept of the visual semantic context (VSC), which functions as the position and heading descriptor of aerial vehicles given a semantically segmented down-looking image and a semantic labelled map. The semantic labelled map includes public map databases that render 2D locations on aerial/satellite imagery, such as Google Maps, OpenStreetMap, and VWorld. VWorld especially is maintained by the Korean ministry of land, infrastructure and transport (MOLIT) and thus provides more accurate semantic map for Korean domestic areas compared to the others. Therefore, this paper exploits labelled maps from VWorld as reference semantic maps that describe positional distributions of semantics, such as buildings and roads. With the help of a semantic segmentation module [5, 12], semantically labelled image can be acquired from a given aerial image, as described in Fig. 1. Here, the such filter yields pixel-wise semantics differing slightly from the semantics of a semantically labelled map, as presented in Fig. 1-(a) compared to Fig. 1-(c).

Defining the visual semantic context from a semantically segmented image begins with discretizing both the radial and azimuthal directions with respect to the image center. This type of decomposition is illustrated in Fig. 2-(a). The azimuthal direction

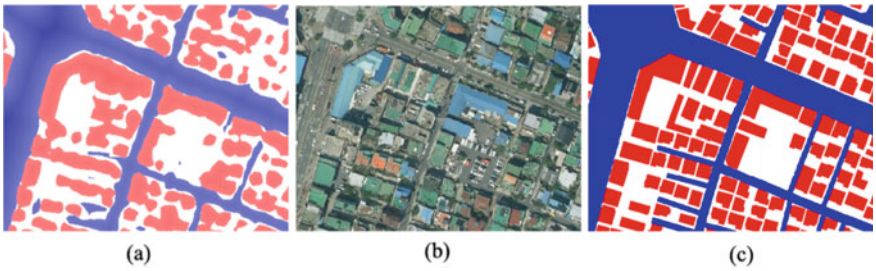

Fig. 1 Semantic segmentation of an aerial image: **a** Semantically segmented version of the original aerial image, **b** original aerial image, and **c** ground-truth semantic label from VWorld

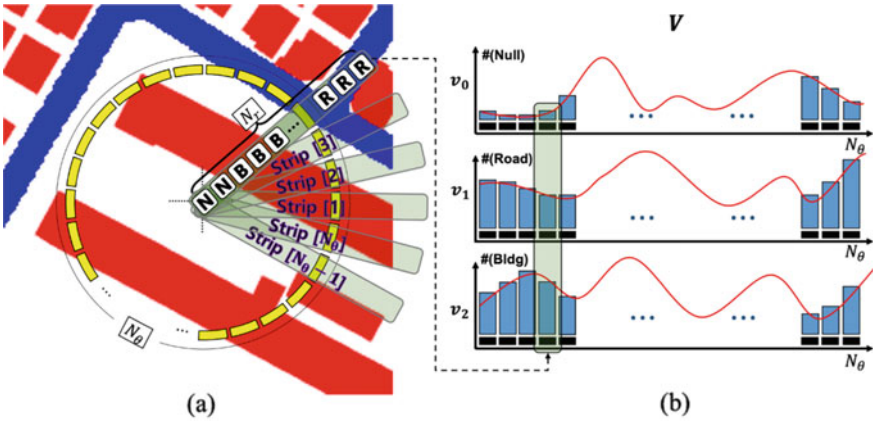

Fig. 2 Schematic description of the visual semantic context: **a** Counting semantics at each strip, **b** semantic-wise azimuthal plots

is discretized into N_θ strips, each of which stretches out to a length of N_r pixels from the center.

Let p_{jk} denote a pixel designated by the kth radial element of the jth strip, which exists at a distance of k pixels from the center, and let s_{jk} denote the corresponding semantics of the pixel. Note that every instance of semantics is from the semantic set $\{S_0, S_1, \cdots, S_{N_s}\}$, where S_0 denotes the null semantic and where each $S_i, i \in \{1, 2, \cdots, N_s\}$ denotes the relevant semantic. Specific semantics of interest in this study will include both buildings and roads. Thus, $N_s = 2$, while S_1 indicates a building and S_2 is a road.

Counting the number of ith semantics laid on each strip leaves the azimuthal distribution of the semantic, and collecting them into row vector $v_i \in N^{N_\theta}$ as

$$v_i = \begin{bmatrix} \sum_{k=1}^{N_r} \delta_i(s_{1k}) \\ \sum_{k=1}^{N_r} \delta_i(s_{2k}) \\ \vdots \\ \sum_{k=1}^{N_r} \delta_i\left(s_{N_\theta k}\right) \end{bmatrix}^T + \overrightarrow{1}_{N_\theta} \tag{1}$$

yields the visual semantic context of the ith semantic where $\overrightarrow{1}_n$ represents an all-one row vector of length n and $\delta_i(s)$ is as follows:

$$\delta_i(s) = \begin{cases} 1 & \text{if} s = S_i \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Figure 2-(b) plots graphs in the orthogonal coordinate system that span the azimuthal distributions of each semantic v_i. In (1), $\overrightarrow{1}_{N_\theta}$ is added to ensure that every element of v_i is nonzero so that further analyses are well-posed. This can prevent infeasible calculations from occurring, such as division by zero, or calculating log0. The visual semantic context is then defined as a $(N_s + 1) \times N_\theta$ matrix by augmenting v_is in the column direction, as

$$\mathbf{V} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N_s} \end{bmatrix} \in N^{(N_s+1)\times N_\theta}. \tag{3}$$

Note that shifting columns of $\mathbf{V}$ by $m \in \mathbb{N}$ columns to the right is equivalent to recalculating $\mathbf{V}$ from an image that is rotated $2\pi m/N_\theta$rad counterclockwise. Thus, column circulation of $\mathbf{V}$ provides a heading description tool when compared to a label map. A couple of semantic-wise plots of the visual semantic contexts and the points from which specific contexts are sampled are shown in Fig. 3.

The contexts sampled from different positions show distinguishable patterns. In this study, $N_\theta = 240$ and $N_r = 40$, obtained by means of trial and error for the best empirical performance, were used. The author found that an enlargement of N_r beyond that value rarely improves the uniqueness of the concept, as a larger N_r ambiguates local characteristics. A larger value of N_θ will increase the azimuthal resolution, though there is a trade-off with regard to the computation cost. These parameters can be adjusted further in accordance with imagery conditions such as scale, resolution, and/or field of view information.

The proposed concept of the visual semantic context is inspired by earlier work [6] and is analogous to this work to some extent. The novelty of the study, however, is in how it is applied to semantically segmented aerial images, analyzing the numerical error characteristics of the proposed concept and providing the potential to be fused with a vision-based integrated navigation system.

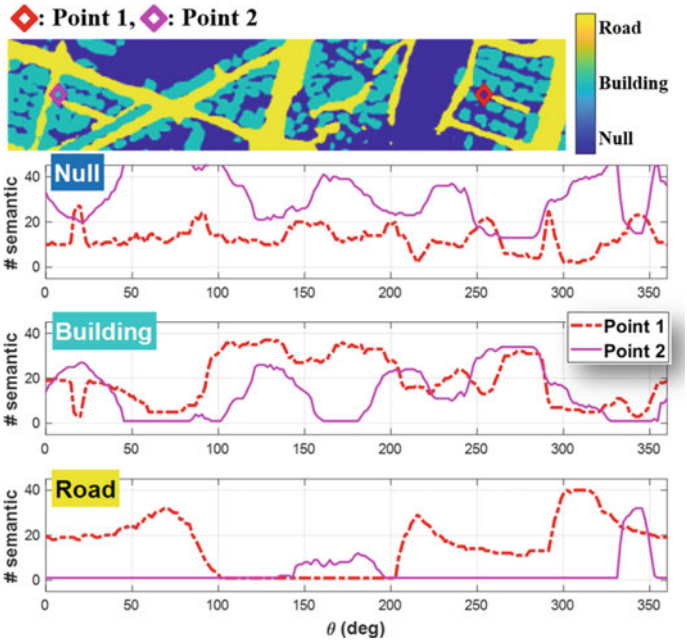

Fig. 3 Two locations from which visual semantic contexts are sampled (up), and respective row-wise plots of the visual semantic contexts (down)

3 Measures of Differences

In this section, we briefly introduce numerous information- and heuristic-based measures of difference between the two visual semantic contexts $\mathbf{V}^{(1)}, \mathbf{V}^{(2)}$ while omitting detailed derivations or backgrounds for the sake of conciseness. The error characteristic analysis of the visual semantic context in the following section makes use of the measures introduced herein.

3.1 Measure of Difference Between Two Probability Mass Function

Dividing the elements of v_i by corresponding summation $\sum_{j=1}^{N_\theta} v_{i,j}$, where $v_{i,j}$ denotes the jth element of v_i, yields the normalized ith semantic context, as

$$\tilde{v}_i = \left(\sum\nolimits_{j=1}^{N_\theta} v_{i,j} \right)^{-1} v_i, \tag{4}$$

which can be interpreted as a point mass function given that all of the elements are positive and $\sum_{j=1}^{N_\theta} \tilde{v}_{i,j} = 1$. Then, the measure of the difference between the two visual semantic contexts in the form of (5) is tractable.

$$D(\mathbf{V}^{(1)}, \mathbf{V}^{(2)}) = \left\| \begin{bmatrix} d(v_0^{(1)}, v_0^{(2)}) \\ \vdots \\ d(v_{N_s}^{(1)}, v_{N_s}^{(2)}) \end{bmatrix} \right\|_2, \tag{5}$$

Where $d(\cdot, \cdot)$ denotes one of the followings:

- Kullback–Leibler divergence (Relative entropy) [10]

$$d_{\mathrm{KL}}\left(v_i^{(1)}, v_i^{(2)}\right) = \sum\nolimits_{j=1}^{N_\theta} \tilde{v}_{i,j}^{(1)} \log\left(\frac{\tilde{v}_{i,j}^{(1)}}{\tilde{v}_{i,j}^{(2)}}\right). \tag{6}$$

- Bhattacharyya distance [1]

$$d_{\mathrm{Bh}}\left(v_i^{(1)}, v_i^{(2)}\right) = -\log\left(\sum\nolimits_{j=1}^{N_\theta} \sqrt{\tilde{v}_{i,j}^{(1)} \tilde{v}_{i,j}^{(2)}}\right) \tag{7}$$

- Jensen-Shannon distance [11]

$$d_{\mathrm{JS}}\left(v_i^{(1)}, v_i^{(2)}\right) = \sqrt{\frac{1}{2}\left(d_{\mathrm{KL}}\left(v_i^{(1)}, v_i^{'}\right) + d_{\mathrm{KL}}\left(v_i^{(2)}, v_i^{'}\right)\right)} \tag{8}$$

- χ^2 distance

$$d_{\chi^2}\left(v_i^{(1)}, v_i^{(2)}\right) = \sum\nolimits_{j=1}^{N_\theta} \frac{\left(v_{i,j}^{(1)} - v_{i,j}^{(2)}\right)^2}{v_{i,j}^{(1)} + v_{i,j}^{(2)}} \tag{9}$$

$v_i^{'}$ denotes $\frac{1}{2}\left(v_i^{(1)} + v_i^{(2)}\right)$ in (8). Note that the semantic-wise difference is augmented in (5) and the L_2-norm is then applied in order to ensure that two $\mathbf{V}$s having a small difference are similarly distributed for all semantics. L_1-norm is also feasible via the same rationale. There will be inexhaustible numbers of measures other than those mentioned above; however, only those that exhibit locally distinct patterns are introduced.

3.2 Measure of Difference Between Two Gaussian Distributions

Alternatively, one can calculate the sample mean μ and unbiased sample covariance $\mathbf{\Sigma}$ from $\mathbf{V}$, as

$$\mu = \begin{bmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_{N_s} \end{bmatrix}, \mu_i = \frac{1}{N_\theta} \sum\nolimits_{j=1}^{N_\theta} v_{i,j}, \tag{10}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{v_0 v_0} & \cdots & \sigma_{v_0 v_{N_s}} \\ \vdots & \ddots & \vdots \\ \sigma_{v_{N_s} v_0} & \cdots & \sigma_{v_{N_s} v_{N_s}} \end{bmatrix}, \sigma_{v_i v_j} = \frac{1}{N_\theta - 1} \sum\nolimits_{k=1}^{N_\theta} \left(v_{i,k} - \mu_i\right)\left(v_{j,k} - \mu_j\right). \tag{11}$$

Given two Gaussian distributions $p(x) \sim \mathcal{N}\left(\mu_p, \mathbf{\Sigma}_p\right), q(x) \sim \mathcal{N}\left(\mu_q, \mathbf{\Sigma}_q\right)$, the L_2 distance of the two are analytically calculated as

$$\begin{aligned} \int (p(x) - \mathrm{q}(x))^2 dx &= \int (p(x))^2 dx + \int (q(x))^2 dx - 2 \int p(x) q(x) dx \\ &= \begin{matrix} \mathcal{N}\left(\mu_p; \mu_p, 2\mathbf{\Sigma}_p\right) + \mathcal{N}\left(\mu_q; \mu_q, 2\mathbf{\Sigma}_q\right) \\ -2\mathcal{N}\left(\mu_p; \mu_q, \mathbf{\Sigma}_p + \mathbf{\Sigma}_q\right) \end{matrix}. \end{aligned} \tag{12}$$

If we treat the columns of $\mathbf{V}^{(i)}$ as a set of samples drawn from the multivariate Gaussian distribution $\mathcal{N}\left(\mu^{(i)}, \mathbf{\Sigma}^{(i)}\right), i \in \{1, 2, \cdots, N_s\}$, following measures are tractable:

- L_2 distance

$$\mathrm{D}_{\mathcal{N}-\mathrm{L}_2}\left(\mathbf{V}^{(1)}, \mathbf{V}^{(2)}\right) = \begin{matrix} \mathcal{N}\left(\mu^{(1)}; \mu^{(1)}, 2\mathbf{\Sigma}^{(1)}\right) + \mathcal{N}\left(\mu^{(2)}; \mu^{(2)}, 2\mathbf{\Sigma}^{(2)}\right) \\ -2\mathcal{N}\left(\mu^{(1)}; \mu^{(2)}, \mathbf{\Sigma}^{(1)} + \mathbf{\Sigma}^{(2)}\right) \end{matrix}. \tag{13}$$

- Bhattacharyya distance

$$\mathrm{D}_{\mathcal{N}-\mathrm{Bh}}\left(\mathbf{V}^{(1)}, \mathbf{V}^{(2)}\right) = \begin{matrix} \frac{1}{8}\left(\mu^{(1)} - \mu^{(2)}\right)^T \mathbf{\Sigma}'^{-1} \left(\mu^{(1)} - \mu^{(2)}\right) \\ + \frac{1}{2} \log\left(\frac{\left|\mathbf{\Sigma}'\right|}{\sqrt{\left|\mathbf{\Sigma}^{(1)}\right|\left|\mathbf{\Sigma}^{(2)}\right|}}\right) \end{matrix}. \tag{14}$$

In (14), $\mathbf{\Sigma}' = \left(\mathbf{\Sigma}^{(1)} + \mathbf{\Sigma}^{(2)}\right)/2$, $|\mathbf{\Sigma}|$ is a determinant of $\mathbf{\Sigma}$. Note that (14) is an analytic realization of (7). The existence of such analytic solutions enables an inexpensive calculation of the difference between the two visual semantic contexts. Only nonzero semantics, or any N_s semantics selected from $\mathbf{V}$, should be utilized when calculating (11), as $\mathbf{V}$, and thus $\mathbf{\Sigma}$, lack one rank assuming that $N_s \ll N_\theta$. Moreover, (13) and (14) are invariant to column permutations

of **V**, while the others are sensitive to permutations. Thus, (13) and (14) are not applicable to describe the heading of an image, though they can provide a global position comparison tool regardless of the absence of prior heading information.

4 Error Characteristics of the Visual Semantic Context

In this section, we present the error characteristics of the proposed concept by plotting residuals in both the spatial and orientational domains. Here, residual indicates the difference between one context computed from the semantically segmented aerial image and another context computed using a semantically labelled map from various locations and with various heading values. Figure 4 and Fig. 5 delineate the spatial error characteristics using the various error metrics presented earlier. Note that the heading is assumed to be given when using (6)–(9), as they are sensitive to permutations. It is clear that permutation-sensitive measures such as (7) show a sharp convex shape that becomes unimodal around the true position in a local sense if served with prior heading information, as shown in Fig. 4. Rotation-invariant measures that make use of the Gaussian statistics in Fig. 5 yield rather smoothed distributions that spread widely; however, they still leave their minimum close to the true position and are applicable without any prior knowledge. This is quite encouraging, as one context is calculated from a noisy segmented image and another is calculated from a fine database. The presented graphs cover a 50×50 m area in the real world. This implies that a unique position fixed solution can be gained simply by comparing the visual semantic contexts when the position uncertainty is less than 100 m, which is highly likely in most cases.

If we assume certain noise characteristics, i.e., additive Gaussian, readers may consider the graphs as likelihoods plotted in the spatial domain before passing through a normal distribution with known variance. Thus, the recursive Bayesian filtering paradigm is also applicable by continuing to multiply this likelihood, similar to the squared exponential of Figs. 4 or 5, by the prior position. The actual location of the segmented aerial image is shown at top of Fig. 4.

Figure 6 illustrates the orientational error characteristics. Residuals are calculated using the aforementioned measures by shifting a single column of the visual semantic context calculated from the labelled map at a time. Graphs other than that designated as the true position were based on an erroneous position on the map. In particular, the distance between the graphs in the real world is 10 m. When we know precisely the position at which the image was taken, the heading estimation shows promising results, leaving the global optimum at the true heading regardless of measures (6)–(9), as the graph in the center presents. The heading estimation, however, is degraded more drastically than the position when the position information is uncertain. Position error close to 10 m may induce significant heading error if estimated solely by matching the visual semantic context and is not supported by additional tools or prior heading knowledge. This phenomenon depends on the map, as position deviation towards the upper/right direction, where distinct road patterns

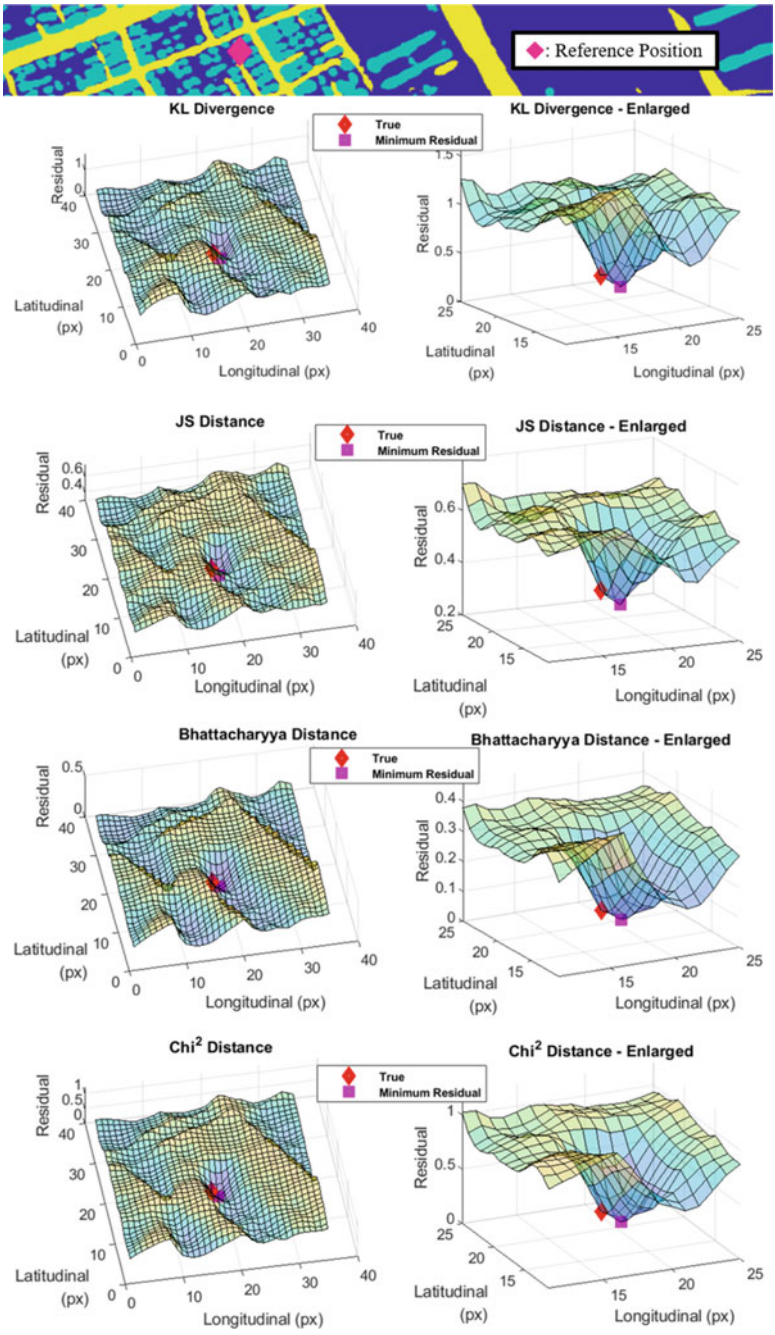

Fig. 4 Spatial error characteristics of the visual semantic context using (6)–(9). The true position where the reference visual semantic context (measurement) is calculated is illustrated at the top

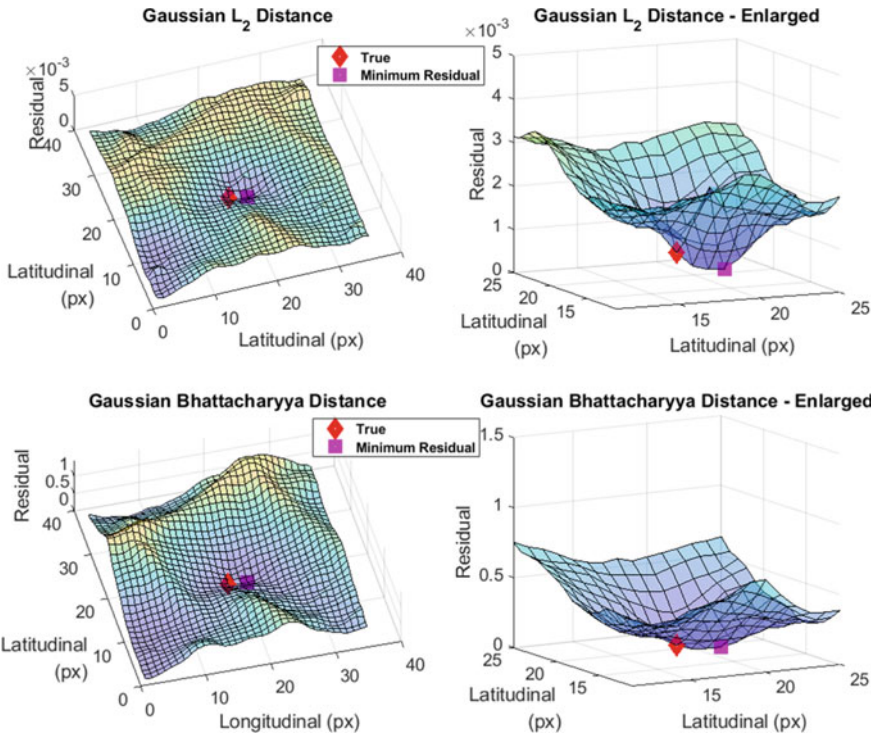

Fig. 5 Spatial error characteristics of the visual semantic context using (13) and (14)

and/or junctions are present, as evident in the map of Fig. 4, leaves little heading estimation error while position deviation towards the other directions results in a completely incorrect estimation. Thus, prior position information would be necessary for the reliable heading estimations so that we can filter out false estimations, which are less likely to occur.

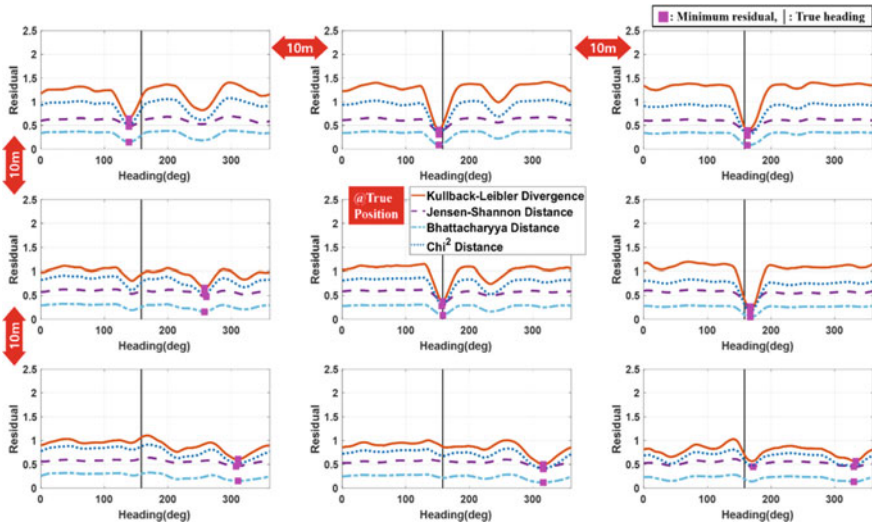

Fig. 6 Orientational (heading) error characteristics of the visual semantic context using (6)–(9). Locations where the reference visual semantic contexts, subject to column circulation, for each graph are calculated are mutually distant by approximately 10 m

5 Conclusion

We proposed the concept of the visual semantic context (VSC), which describes the position of an aerial vehicle using a semantically segmented aerial image and a semantic labelled map applicable to aerial navigation. Because the semantic segmentation of an aerial image aided by a deep learning technique is already an expensive task, introducing complicated patterns as well makes little sense in terms of practical navigation. Thus, the proposed concept focuses on the simplest feature that is both easy to implement and readily explainable. The authors believe that the concept should be robust to imprecise altitude information of an aerial vehicle, i.e., the scale, up to certain point due to its scalability, attained by rearranging the semantics in polar coordinates. The presented analysis, however, showed that the optimum heading estimation is sensitive to position prior. If associated with a filtering paradigm or with a two-phased approach that estimates the position initially such that prior (position) knowledge can be exploited, recursively estimating both the position and heading of a vehicle is expected to be feasible because the two alternately correct each other. We are planning to undertake VSC-based navigation of an aerial vehicle, i.e., position fixing, which incorporates either the prior knowledge of the vehicle state using recursive filtering or with a rejection scheme that prevents false estimations.

References

1. Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. Bull Calcutta Math Soc 35(1):99–109
2. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24(5):603–619
3. Dumble SJ, Gibbens PW (2015) Airborne vision-aided navigation using road intersection features. J Intell Rob Syst 78(2):185–204
4. Groves P (2013) Principles of GNSS, inertial, and multisensor integrated navigation systems, 2nd edn. Artech, Boston
5. Hong K, Kim S, Bang H (2020) Vision-based navigation using Gaussian mixture model of terrain features. AIAA Scitech Forum, Orlando
6. Kim G, Kim A (2018) Scan context: egocentric spatial descriptor for place recognition within 3d point cloud map. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), Madrid, Spain
7. Kim S (2019) Vision-based map referenced UAV navigation using terrain classification. Master's thesis, Korea Advanced Institute of Science and Technology
8. Kim Y (2021) Aerial map-based navigation using semantic segmentation and pattern matching. s.l., arXiv preprint arXiv:2107.00689
9. Koch T, Zhuo X, Reinartz P, Fraundorfer F (2016) A new paradigm for matching UAV-and aerial images. ISPRS Ann Photogram Rem Sens Spatial Inf Sci III(3):83–90
10. Kullback SRAL (1951) On information and sufficiency. Ann Math Stat 22(1):79–86
11. Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory 37(1):145–151
12. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA
13. Masselli A, Hanten R, Zell A (2016) Localization of unmanned aerial vehicles using terrain classification from aerial images. Proc Intell Autonom Syst 13:831–842
14. Michaelsen E, Jaeger K (2009) A google-earth based test bed for structural image-based UAV navigation. In: IEEE international conference on information fusion, Seattle, WA
15. MOLIT (2012) VWorld Data Center, operated by Ministry of Land, Infrastructure and Transport (MOLIT) of South Korea. https://data.vworld.kr/data/v4dc_usrmain.do. Accessed 7 Oct 2021
16. Park J, Kim Y, Bang H (2017) A new measurement model of interferometric radar altimeter for terrain referenced navigation using particle filter. In: IEEE European Navigation Conference (ENC), Lausanne, Switzerland
17. Shan M et al. (2015) Google map aided visual navigation for UAVs in GPS-denied environment. In: IEEE international conference on robotics and biomimetics (ROBIO), Zhuhai, China
18. Wang T, Celik K, Somani AK (2016) Characterization of mountain drainage patterns for GPS-denied UAS navigation augmentation. Mach Vis Appl 27(1):87–101
19. Zhuo X et al (2017) Automatic UAV image geo-registration by matching UAV images to georeferenced image data. Remote Sensing 9(4):376