

# An Entropy-Based Model for Indoor Self-Localization Through Dialogue

Kimia Amoozandeh<sup>1</sup> ✉ 

The University of Melbourne, Parkville, Australia

Ehsan Hamzei ✉ 

The University of Melbourne, Parkville, Australia

Martin Tomko ✉ 

The University of Melbourne, Parkville, Australia

---

## Abstract

People can be localized at a particular location in an indoor environment using verbal descriptions referring to distinct visible objects (e.g., landmarks). When a user provides an incomplete initial location description their location may remain ambiguous. Here, we consider a dialogue initiated to update the initial description, which continues until the updated description can be related to a location in the environment. In each interaction, the wayfinder is incrementally asked about the visibility of a particular object to update the initial description. This paper presents an entropy-based model to minimize the number of interactions. We show how this entropy-based model leads to a significant reduction of interactions (i.e., reduction of conversation length, measured by the number of additional referents) compared to baseline models. Moreover, the effect of the initial description, i.e., the first set of visible objects with different combinations, is investigated.

**2012 ACM Subject Classification** Human-centered computing → Interaction paradigms; Human-centered computing → Interactive systems and tools; Information systems → Location based services

**Keywords and phrases** Indoor self-localization, Dialogue, Entropy

**Digital Object Identifier** 10.4230/LIPIcs.COSIT.2022.24

**Category** Short Paper

**Supplementary Material** *Software (Source Code and Data)*: <https://github.com/hamzeiehsan/dialogue-based-self-localization>

**Funding** The support by the Australian Research Council grant DP210101156 is acknowledged.

## 1 Introduction

Self-localization is a fundamental prerequisite of navigation systems for indoor environments [8, 7]. Existing indoor localization systems often rely on infrastructures (e.g., WiFi, beacons) [6]. Yet, such systems are not universally available and are associated with significant costs of installation, affected by disruptions, and require maintenance by operators [11]. Infrastructure-independent approaches for localization based on verbal communication may help overcome these challenges.

People can be localized within a space through verbal descriptions containing references to a set of visible objects (e.g., landmarks) [1]. In this approach, the space is decomposed into a set of discrete regions that are each characterized by a set of visible objects in the environment known as *visibility signature*.

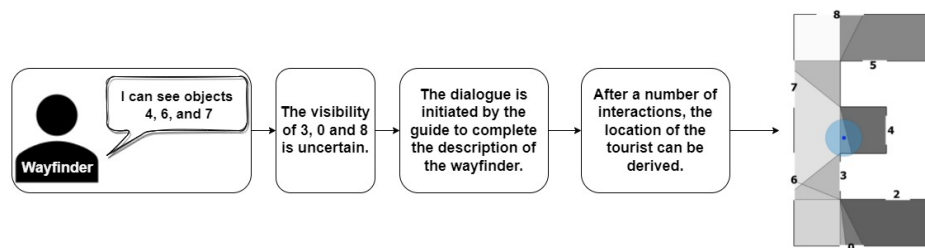
Yet, when self-localizing with verbal descriptions, people may provide incomplete descriptions, resulting in ambiguous specifications of their position in the environment. As a solution, we envisage the initiation of dialogue to resolve description ambiguity (see Figure 1).

---

<sup>1</sup> Corresponding author



This localization dialogue thus involves two agents (i.e., computer and human): a *guide* and a *wayfinder*, who interact to achieve a common goal – i.e., identifying the location of the wayfinder [2]. For this purpose, Gintner et al. [3] introduced a method to improve the localization process for visually impaired people by interacting with the user to derive the direction of the user and the particular sidewalk of the street on which they are. From another point of view, Sernani et al.[9] introduced a chatbot system based on voice and written interaction in museums combining indoor localization and chatbots to offer customized visits. In their approach, wearable sensors and a mobile map are required for localization with Ultra-WideBand (UWB) radio technology. This paper proposes a dialogue-based localization approach independent of any particular infrastructure.



■ **Figure 1** The human description of their location, through reference to an incomplete set of surrounding objects, i.e., facades. A dialogue is initiated to update the description and identify the location of the wayfinder in the environment.

Hua et al. [5] introduced a qualitative place map to enable landmark-based interactive localization. In their approach, the agent describes the nearest landmarks based on different qualitative relations, such as direction and distance. The agent’s location is then derived based on the description matched to an existing qualitative graph map. The challenge of their [5] approach is that since the vertices of the place graph are landmarks and turning points, the location of the agent can only be assigned to these vertices. Hence, in some complex environments (e.g., airports), the resulting position determination may not be accurate enough.

In order to enable indoor localization through dialogue in our approach, we assume that the *guide* has access to a map of decomposed regions based on location signatures, as recently proposed by Amoozandeh et al. [1]. We further assume that the *wayfinder* has a 360-degree view of the environment and the ability to identify visible objects that are captured in the signatures. The agents then interact and collaborate to locate the wayfinder through dialogue [2]. The wayfinder initiates the dialogue with reference to the first set of visible objects. If incomplete, this initial description needs to be updated through dialogue to add references until an unambiguous match to a signature of one of the decomposed regions is eventually found. This dialogue may consist of multiple iterative rounds of interaction.

In this paper, we hypothesize that the number of interactions in the dialogue (i.e., dialogue length) can be effectively minimized using a measure of *entropy* evaluating references selected for the update of the initial description. This paper addresses the following research question:

- How can the number of interactions for localizing a person in an environment be minimized?
- How does the amount of references in the initial description affect the number of interactions in the entropy-based reference search compared to the baselines?

Since *entropy* measures the quantity of information held by a quantum of data [10], in order to minimize the number of interactions, objects with a high value of entropy are hypothesized to lead to less ambiguous descriptions faster (Section 2). In order to answer the

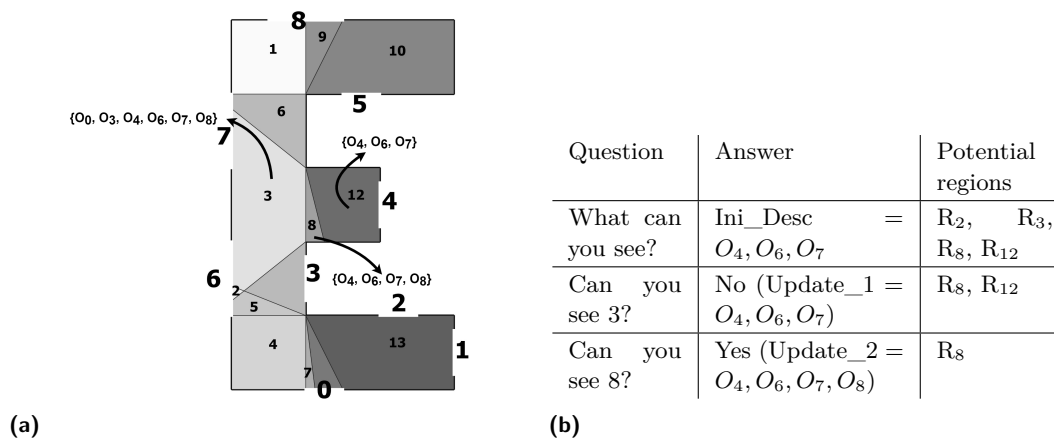
research questions, we compare the *entropy*-based approach to two baseline models, which update the initial description querying about the visibility of objects that are not included in the initial description either randomly or in a supervised (guided) manner.

The three approaches are tested with different combinations of initial descriptions (Section 3). The results show that our method performs similarly to the baselines with complete initial information. In contrast, with incomplete initial information, the method outperforms the baselines and significantly reduces the dialogue length (Section 3).

## 2 Approach

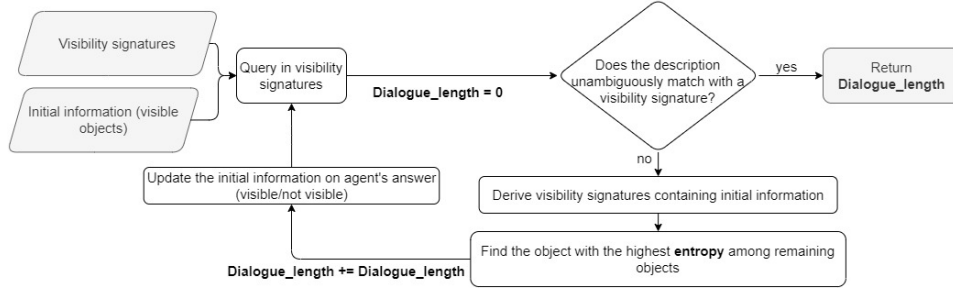
To initiate a dialogue for indoor self-localization, an initial description of the location of the *wayfinder* is required. The initial description is a set of references to visible objects in the space (e.g., I can see objects  $O_4$ ,  $O_6$ , and  $O_7$  according to the conversation in Figure 2b). This set of references is supposed to identify the wayfinder’s location in the space uniquely. Here, a location is synonymous with one of the regions identified by decomposition of space by visibility signatures of a set of visible objects (landmarks), located along the periphery of the space through which the wayfinder navigates [1]. We note that these are not point-like objects but objects with a linear extent.

In Figure 2a we show a hypothetical environment with arbitrary objects (the extent of which is shown as along the boundaries). The grey-shaded polygons are the decomposed regions from each of which a distinct set of objects is visible. The set of objects  $O_4, O_6, O_7$  is visible from region number  $R_{12}$ . If the initial description is incomplete, a dialogue is initiated to update and complete the description (see Figure 3). In each interaction between the *wayfinder* and the *guide*, the guide asks whether a particular object is visible to the *wayfinder*. Based on the answer (i.e. visible or not visible), the set of references in the initial information is updated (e.g., can you see object  $O_3$ ?) and, again, matched to the signatures (see Figure 2b).



**Figure 2** (a) An example of *visibility signatures* for decomposed regions in an E-shaped environment with ten arbitrary objects. For instance, from  $R_{12}$ , the set of objects  $O_4, O_6, O_7$  is visible, and (b) A sample dialogue between the *guide* and the *wayfinder* to localize the wayfinder in the environment depicted in Figure 2a. O: object, R: Region, Ini\_Desc is the initial description.

In each interaction, the *guide* needs to choose an object and ask about its visibility. This object can be chosen in three ways: randomly, supervised between remaining objects, or based on the *entropy* measure among the remaining objects, i.e., the objects that are not



■ **Figure 3** Overview of the entropy-based search to complement and match an initial location description to the environment.

included in the initial set of visible objects and visible from the potential regions according to the initial information. A sample dialogue based on the *entropy* measure is shown in Figure 2b. Assuming that the *wayfinder* is in region  $R_8$ , the initial description from the *wayfinder* is the set  $O_4, O_6, O_7$ . According to Figure 2a, the *wayfinder* can be in regions  $R_2, R_3, R_8$ , or  $R_{12}$ . The visibility of objects  $O_0, O_1, O_2, O_3, O_5$ , and  $O_8$  can be asked in the next interaction.

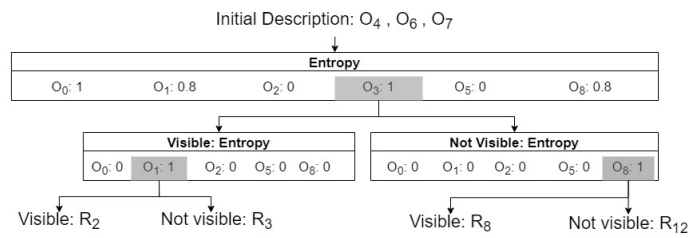
The entropy will help to identify candidate object that invalidates the most candidate regions with visibility signatures partially matching the initial description (see Equation 1). We treat these regions as candidate regions. For instance, based on the initial description in Table 2b (objects  $O_4, O_6$ , and  $O_7$ ), we can identify candidate regions based on the partially matching visibility signatures (Table 1). The entropy of remaining objects  $O_0, O_1, O_2, O_3, O_5$ , and 8 is calculated based on Equation 1 and shown in Figure 4.

$$Entropy_{O_k} = -(P(v) \log_2 P(v) + P(nv) \log_2 P(nv)) \quad (1)$$

In Equation 1,  $P(v)$  is the probability that the object  $O_k$  is visible in the decomposed region matching the initial description, and  $P(nv)$  is the probability that  $O_k$  is not visible in the decomposed region matching the initial description. The object with the highest entropy will be enquired about next, to update the initial description. Amongst the remaining objects in the example in Table 1, objects  $O_3$  and  $O_0$  have the highest entropy. Choosing one of these objects means that only one other question is required to derive the wayfinder's location, regardless of the answer. Different possible combinations of answers and questions are shown in Figure 4.

■ **Table 1** *Visibility signatures* of the potential regions based on the initial description of the conversation in Table 2b.

Region <sub>id</sub>	$O_0$	$O_1$	$O_2$	$O_3$	$O_4$	$O_5$	$O_6$	$O_7$	$O_8$
$R_2$	1	1	0	1	1	0	1	1	1
$R_3$	1	0	0	1	1	0	1	1	1
$R_8$	0	0	0	0	1	0	1	1	1
$R_{12}$	0	0	0	0	1	0	1	1	0

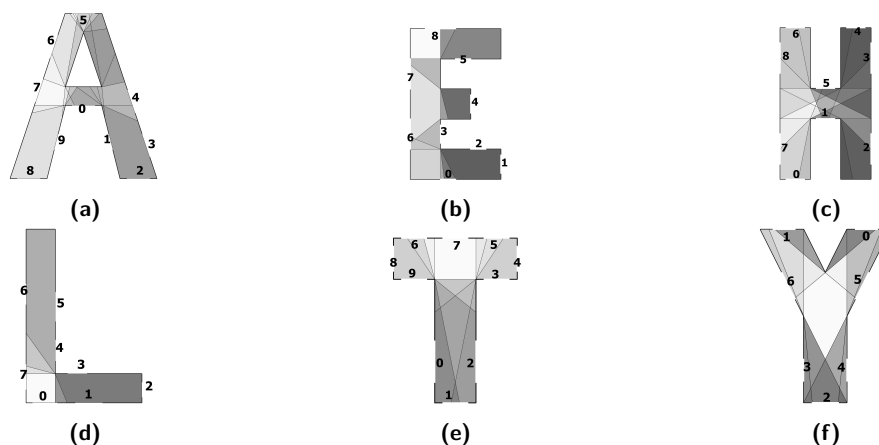


■ **Figure 4** Choosing the questions based on the entropy of remaining objects in the conversation shown in Figure 2b.

To show the effect of entropy on the reduction of the dialogue length, the entropy-based model is compared with two baseline models enabling to update the initial description: references chosen randomly and a guided supervised search among the set of objects that are not in the initial description, i.e., remaining objects that includes *only* the objects that are visible from each candidate region but not all of them. For instance, the next answer object in our scenario can be selected randomly among all the remaining objects in the environment, i.e., objects  $O_0, O_1, O_2, O_3, O_5,$  and  $O_8$ . Alternatively, since objects  $O_2$  and  $O_5$  cannot be seen from the regions identified as candidates by the initial description (Table 1), the next object may be chosen in a supervised (guided) manner among objects  $O_0, O_1, O_3,$  and  $O_8$ . All three models are tested with different numbers of references in the initial description and combinations of initial descriptions to illustrate the effect of the initial description on the number of interactions.

### 3 Results

The three methods have been applied to six different hypothetical environments, generated based on the font outlines of san-serif characters from the capitalised Roman alphabet (Figure 5). For each decomposed region set as the *wayfinder's* location, different combinations of visible objects have been introduced as the initial description (a subset of the visibility signature). Next, the dialogue was initiated, and an additional object's visibility has been enquired about among the remaining objects.

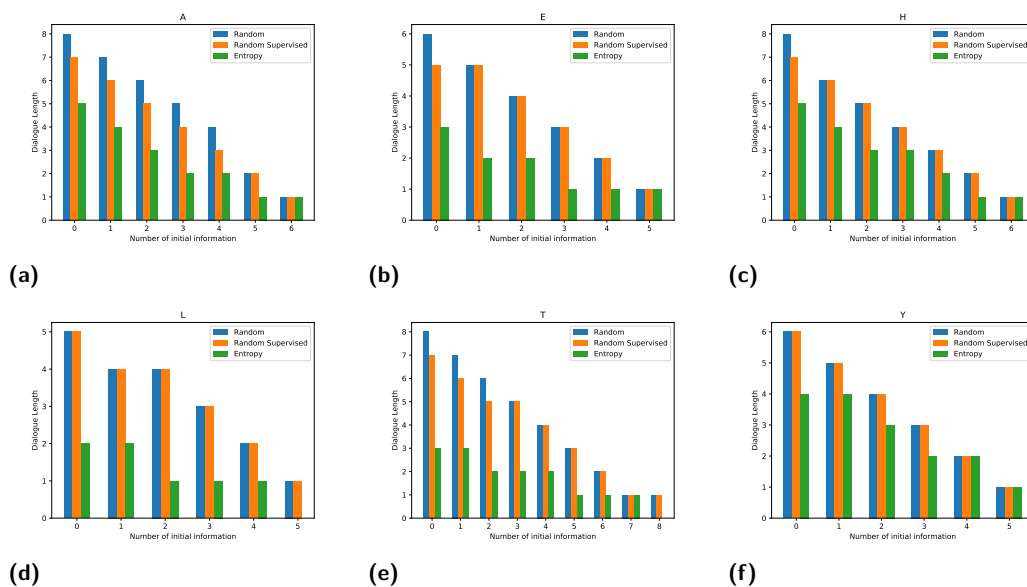


■ **Figure 5** Different environments with distinct visible objects along the boundary. The discrete gray shaded polygons in the environments are the decomposed regions. A set of objects is visible from each decomposed region [1].

In the random model, the object was chosen randomly amongst the remaining objects in the environment. In the supervised model, the object has been chosen among the objects in the candidate regions (i.e., partially matching the original signature). Finally, in the entropy-based model, the object was chosen from the same set matching candidate regions but in order of decreasing entropy.

Figure 6 illustrates the effect of the amount of initial information and the model used on the length of the dialogue for the test environments (Figure 5). Using the entropy-based model leads to a shorter dialogue than in the random or supervised models. Further, the more objects are referred to in the initial description; the lesser is the effect of the model (incl. entropy model) on the dialogue length. In other words, the entropy model outperforms in minimal information situations, but the improvement is less notable for comprehensive initial descriptions.

Moreover, in some cases, such as in Figure 5b, d and e, the number of interactions (dialogue length) remains constant with the increasing amount of initial information when using the entropy-based model. It means that the spatial relation of objects and the configuration of the space can affect the number of interactions in the dialogue, which will be investigated in future works. Thus, the information provided by references to additional objects may be excessive.



■ **Figure 6** Comparison of the dialogue length in different environments based on the amount of initial information and using baseline approach and entropy approach.

## 4 Conclusion and Future Works

In this paper, we introduced a dialogue-based approach to localize wayfinders in indoor environments based on the visibility of objects (e.g., landmarks). We show that to minimize the number of interactions in a localization dialogue (dialogue length) between the wayfinder and a guide, the proposed entropy-guided method can be used effectively. Our approach significantly reduces the required dialogue length compared to a random or supervised baseline, even with minimal initial information.

In future work, we will test the entropy-based approach in real indoor environments and evaluate its performance with human wayfinders in realistic scenarios. We aim to enhance the approach to derive additional information from each interaction. The current approach relies only on prompted verification questions (i.e., can you see object  $O_4$ ?), while other types of questions (e.g., disjunctive questions: can you see object  $O_3$  or  $O_4$ ?) may be used for more informative interaction. Using diverse types of questions [4], we expect to gain improvements and thus further reduce dialogue length. Moreover, we did not consider the intrinsic differences among the landmarks (e.g., some are more salient than others). However, in a real-world scenario, such differences are essential, and their differences should be considered for a dialogue-based localization. Finally, the complexity of the environment and its impact on dialogue length is not comprehensively investigated in this preliminary study and provides promising avenues for future work.

---

### References

- 1 Kimia Amoozandeh, Stephan Winter, and Martin Tomko. Space decomposition based on visible objects in an indoor environment. *Environment and Planning B: Urban Analytics and City Science*, 49(3):883–897, 2022. doi:10.1177/23998083211037347.
- 2 Harm de Vries, Kurt Shuster, Dhruv Batra, Devi Parikh, Jason Weston, and Douwe Kiela. Talk the walk: Navigating new york city through grounded dialogue. *arXiv preprint*, 2018. arXiv:1807.03367.
- 3 Vojtech Gintner, Jan Balata, Jakub Boksansky, and Zdenek Mikovec. Improving reverse geocoding: Localization of blind pedestrians using conversational ui. In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 000145–000150. IEEE, 2017.
- 4 Arthur C. Graesser and Natalie K. Person. Question asking during tutoring. *American Educational Research Journal*, 31(1):104–137, 1994. doi:10.3102/00028312031001104.
- 5 Hua Hua, Peng Zhang, and Jochen Renz. Qualitative place maps for landmark-based localization and navigation in gps-denied environments. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 23–32, 2019.
- 6 Germán Martín Mendoza-Silva, Joaquín Torres-Sospedra, and Joaquín Huerta. A Meta-Review of Indoor Positioning Systems. *Sensors*, 19(20), 2019. doi:10.3390/s19204507.
- 7 Daniel R. Montello. Navigation. In *The Cambridge Handbook of Visuospatial Thinking.*, pages 257–294. Cambridge University Press, 2005. doi:10.1017/CB09780511610448.008.
- 8 Urs-Jakob Rüetschi and Sabine Timpf. Modelling Wayfinding in Public Transport: Network Space and Scene Space. In *Spatial Cognition IV. Reasoning, Action, Interaction*, pages 24–41, 2005.
- 9 Paolo Sernani, Sergio Vagni, Nicola Falcionelli, Dagmawi Neway Mekuria, Selene Tomassini, and Aldo Franco Dragoni. Voice interaction with artworks via indoor localization: A vocal museum. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 66–78. Springer, 2020.
- 10 C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.
- 11 Stephan Winter, Martin Tomko, Maria Vasardani, Kai-Florian Richter, Kouros Khoshelham, and Mohsen Kalantari. Infrastructure-Independent Indoor Localization and Navigation. *ACM Comput. Surv.*, 52(3), June 2019. doi:10.1145/3321516.