Scheme for the automatic generation of directions to locate objects in virtual environments

Graciela Lara[1], Angélica De Antonio[2], Adriana Peña[1], Mirna Muñoz[3], Luis Casillas[1]

[1] CUCEI of the Universidad de Guadalajara,
Av. Revolución 1500, Col. Olímpica, 44430 Guadalajara (Jalisco), México.
graciela.lara@red.cucei.udg.mx, {adriana.pena, luis.casillas}@cucei.udg.mx
[2] Escuela Técnica Superior de Ingenieros Informático of the Universidad Politécnica de Madrid,
Campus de Montegancedo, 28660 (Boadilla del Monte), España.
angelica@fi.upm.es
[3] Centro de Investigación en Matemáticas
Avda. Universidad no. 222, 98068, Zacatecas, México
mirna.munoz@cimat.mx

Abstract. The automatic generation of direction in natural language, for the location of objects, is an ongoing research area heavily supported by the use of virtual environments (VEs). Important components of spatial language such as the selected reference object, along with specific features related to the situation of the scenario and the user, have to be properly combined in order to create a helpful direction to locate an object within a VE. In this paper we present a scheme, constructed upon literature review and specific empirical data, to link those different elements related to the location of objects, aimed to establish the suitable algorithms for the automatic generation of spatial language in VEs.

Keywords: virtual environments, location of objects, spatial language.

1. Introduction

Humans, animals and objects occupy a place in space; in consequence, we have developed spatial knowledge, a basic skill for the localization process. Albeit a seemingly simple task, spatial language to express where objects are located, calls for a mix of spatial knowledge and an accessible visual representation for our linguistic system [1].

The automatic generation spatial language, directions for the location of objects, represents a challenge with a number of difficulties. Take for example, the use of absolute references that might cause confusion with respect to relative references [2]; or the fact that there is not a straight

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*, Advances in Intelligent Systems and Computing 537, DOI 10.1007/978-3-319-48523-2_12

forward method to select a reference object [3]. Even though, a number of applications applying spatial language have been developed, among which are included: graphic design and drawing programs, computer games, navigations aids, robot systems, training simulators and geographic information system interfaces [3], where virtual environments (VEs) play an important role.

1.1. Related Work

Concerning VEs, in his doctoral thesis, Kelleher [4] developed a computational framework, perceptually based, for the interpretation of spatial language. The computer system is a VE for a user to navigate and manipulate objects. It contains a mechanism for the user to select different frames of reference thorough a semantic framework with locative prepositions. On it, a visual saliency algorithm is used for the integration of the speech.

Also for the interpretation of spatial language, Gorniak & Roy [5] developed a system through a model that describes objects of 3D scenes but with spatial relations interpreted in 2D. Their system has an algorithm to extract visual features of the objects, and it manages the description of spatial relations. Their descriptive spatial language contains hundreds of reference-expressions based on similar scenes, a syntactic analyzer of spoken expressions, and a composition engine managed by an interpreter with various lexical units.

For the automatic generation of spatial language in VEs, Barclay [3] developed a model for processing scene descriptions that operates in realistic environments, tested on a large set of 3D scenes representations. This project emphasizes the use of references and spatial relations for locating objects, through perceptual salience.

More recently, Trinh [6] developed a system with a tool for semantic modeling of spatial relations among objects in VEs, where the spatial relations are specified at a conceptual level. The model focuses on the spatial limitations of VEs, such as space communication difficulties. For a detailed analysis of computer systems with spatial language see Lara, De Antonio & Peña [7].

None of these systems incorporates the user modeling. In this paper we present a scheme that deals with the complexity and specific elements that are combined for the automatic generation of spatial language. The scheme includes an original algorithm for 3D objects, to select the best reference object; and it also, incorporates significant aspects of the user modeling, to give direction to a user in a VE for the location of objects.

2. Elements for the automatic generation of spatial language

As mentioned, spatial language involves different concepts and elements. Because the location of an object is inherently relative, to place it, a *frame of reference* has to be established. The spatial frames of reference provide a structure to specify the object's spatial composition and position; a coordinate system to give directions from different points in space or a mental representation of positions such as up, down or side [8]. In the adoption of a specific frame of reference, an object can be pointed in relation to: the observer, the environment, its own intrinsic structure, or other objects in the environment [9].

A common practice to give direction for the location of an object is the use of *reference objects*; in fact, it might be difficult to state the position of an object without referring to another [8]. The selection of a *reference object (RO)* conveys the recognition of its prominent features, that is, its *perceptual saliency*. Early mentioned by Titchener [10], this key concept for the location of objects has been described as those features of the object that somehow draw our attention [11, 12, 13], mainly probable because they are rare or just different in the scenario [14,15]. However, as Gapp [16] stated out, in some cases, the selection might obey only to the distance between the *object to be located (OL)* and the RO. The use of a RO implies to establish a spatial relation between the OL and the RO, see Figure 1.

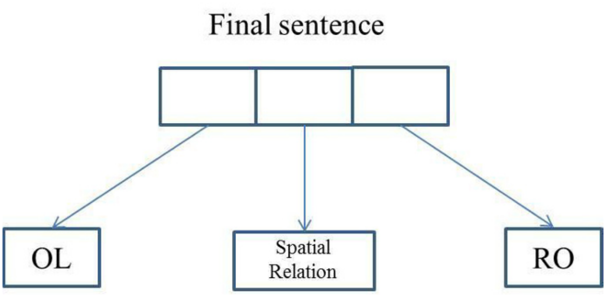

Fig. 1. Structure of a spatial language sentence

In addition, the good perception of an object depends on factors such as visual acuity, clarity of vision and the viewpoint of the observer, as well as on the social and psychological impact that these features might have on the viewer.

A first approach to give directions to a person for the location of an object is to take into account if the OL is within the *field of view* of that person [17]. If the OL is not in the users' field of view, it might or not be the case that they are in the same room (or space). In both cases, extra directions are required to place the user in a position in which his/her field of view reaches the OL; and both are out of the scope of this paper.

The scheme here presented is then focused to the particular case that the OL is within the user's field of view. This leads to diverse situations, from which were included in the scheme the next cases: A) The object is salient by itself; B) The object is somehow occluded; and C) The object is not salient and therefore a RO is required.

The elements for the automatic generation of spatial language situations will be next described, and in the next sections how to use those elements for the three aforementioned situations, including others derived from them, are discussed.

2.1. Syntactic structure of the spatial language

Kelleher [4] proposed a linguistic structure proper for the automatic generation of spatial language, the one shown in Figure 2. In which the syntactic structure is divided into a nominal syntagm that refers to the OL, and a verbal syntagm and adverbial syntagm. The nominal syntagm includes the article "The", + an optional (Object_feature) + the OL; an example is: "The yellow pen". The verbal sytagm specifies the nature of the problem, in this case to indicate the positional situation, which is always "is". Finally the adverbial syntagm contains both the spatial relation and the RO; its structure comprehends the "spatial relation" + the article "the" + an optional (Object_feature) of the RO + the RO. It is worth to mention that the adverbial sytagm changes when it includes several references object, take for example this adverbial syntagm: "…in front of the desk, between the red ball and the bicycle, to the right of the second printer."

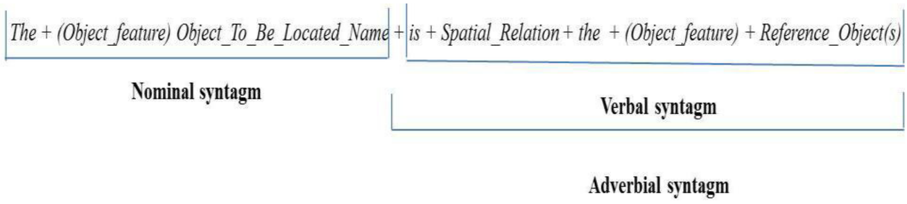

Fig. 2. Syntactic structure of a spatial language sentence

In order to generate in an automatic form this structure, we need to establish a reference object and then select the proper spatial relation.

2.2. Algorithm for the Selection of the Best Reference Object

Three criteria were considered for this algorithm: *visual saliency* of the object, and from the user: *prior knowledge* [16, 18], and the *probability to remember*. The visual saliency of an object might relay on a number of factor. However, because VEs are predominantly visual and based on a literature review, we proposed in [19], the use of the more prominent objects' features related to the human vision, that is: color, size and shape.

For the automatic interpretation of these features in a VE, in [20] we proposed a weighted combination of these three measures of the features of the objects, and how to get them, in order to calculate the saliency by each object in the scenario, to obtain an ordered relation by perceptual saliency of all of them.

The second criterion for the algorithm is the prior knowledge of the user. It refers to the areas of knowledge to which the objects belong and the user's probability to posses them, based on his/her previous training or experience. It represents the level of familiarity that a user possesses with a certain area of knowledge and the objects associated with this area. For example, a computer engineer should have certain knowledge regarding computer input/output devices. This requires for the system to manage a user modeling [21], with a test from which a degree of familiarity related to the analyzed object can be established.

The last criterion for the algorithm is the probability to remember, that is, the probability for a user to remember a previously seen object in the VE. For it, three factors can be considered: the perceptual saliency of the object in question, the user's ability to remember objects' location and the history of the user visibility of the object within the VE.

The first factor, the perceptual saliency of the object, can be calculated as in [20]. For the user's probability to remember the location of an object, the Rey-Osterrieth complex figure text [22] provides a scale for each particular user, which in turn will be part of the user modeling.

As of the history of vision, it involves several concepts: the maximum viewed surface of the object inside the scene from any of the past user's points of view, the clarity of vision or clarity of perception, the time of vision exposure of an object by the user, and the time of oblivion (the involuntary action of stopping remembering or keeping in memory the information of the objects observed in a scene). All of which imply a number of calculation, regarding the history of the different user's fields of view during his/her navigation in the VE.

This algorithm will proportionate the best RO, and by including in it the OL, it will also be able to establish if the OL is salient by itself.

A third part of the syntactic structure (see Figure 1) is the spatial relation. An algorithm to select it is next described.

2.3. Algorithm for the Selection of the Spatial Relation

Gapp [23, 24] divided spatial relations into two classes: topological, those that refer to a region proximal to an object (e.g. "at", "near"); and projective, that take into account the relation between those objects (e.g. "in front of", "behind" or "above"). The relation "between" has an exceptional position in the group, because it refers to two objects, its basic meaning is defined by the structure of its region of applicability; the location with the highest degree of its applicability is the midway between two ROs.

Different criteria for the automatic choice of a particular spatial relation can be applied. A set of selected relations and the criteria to use them is here briefly described. The set includes some of the most commonly used, and that we consider cope most of the situations.

On/Under are spatial relations suitable when the OL is in contact with the RO, therefore a collision in the VE has to be detected, in the "Y" axis.

Close to is a proper relation when the distance between the OL and the RO does not exceed certain threshold.

Left/Right/Above/Below/In front of/Behind are spatial relations mutually exclusive. The key to choose one is to establish, from the point of view of the user, the nearest objects to obtain their roominess. The roominess will evaluate the objects' points that fall down inside or out of the distance, from the user to the objects in his/her field of view.

Inside requires first determining if the OL is inside another one. In this case can be applied the ray-casting technique; by evaluating the hit object or objects through several rays directed from the center inside of an object, it can be inferred an outsider object.

Between can be used once is evaluated if there are several ROs candidates. The two closest objects of reference (RO1 and RO2) are then identified and their distance to the OL is calculated. A criterion based on the object with highest distant, lower than the nearest distance to the OL with a threshold difference, can be applied to use this spatial relation.

Once the best reference objects and a spatial relation are established then the next step is to generate spatial language, this algorithm is next described.

3. Algorithms to Generate Spatial Language

Let us now have a close look to the syntaxes for the different mentioned situations, when the OL is in the field of view of the user.

A) **The OL is salient on itself**; this situation does not require a RO to generate its location direction. In this case, for simplicity, we proposed to use the user as frame of reference, which means that the word "you" will be used instead of the RO. The structure is then (as described in section 2.1):

"The", + an optional (Object_feature) of the OL + the OL + "is" + the "spatial relation" + "you".

Example: "The yellow box is in front of you"

B) **The OL is somehow occluded**; this could be because the object is inside another one, or because another object occludes it. In any case the object that occludes the OL becomes a secondary object to be located (OL2) and a second spatial relation is required to indicate the relation between the OL and the OL2, that in turn requires a secondary RO (RO2). The structure, in this case is:

"The", + an optional (Object_feature) of the + the OL + "is" + the "spatial relation" + "the" + the OL2 + "that is" + "spatial relation" + "the" + an optional (Object_feature) of the RO2 + the RO2

Example: “The blue ball is inside the white box that is on the brown desk”

C) **The OL is not salient and it requires a RO**, in this case the RO might or not have a high probability of being remembered by the user, both cases are treated separately. In the case that there is a high probability that the user remembers a RO, inside a certain radius near to the OL, then the spatial relation between them is determined to generate the directions with the same structure used in the A) situation.

If there is not a high probably for the user to remember any of the ROs within a radius of the OL, then the perceptual saliency of the objects is applied to them, and the sentence with directions is structured as in the situation A). These situations are summarized in the next Table 1.

Table 1. Criteria for different situations for the automatic generation of directions.

The OL is in the field of view of the user	
Situations	Criteria
A) The OL is salient on itself	• The directions include a spatial relation of the intrinsic type (listener centred).
B) The OL is somehow occluded	• The directions consider the object that contains the object to be located, or the object that occludes it, and it is transformed into a secondary object to be located OL2.
C) The OL is not salient and it requires a RO	• In this situation is considered the users’ probability to remember the RO.
C.1) The RO has a high probability of being remembered by the user	• The direction is generated, considering the RO with the maximum probability of being remembered. • The user has previous knowledge about the RO. • The spatial relation is determined between the OL and the RO.
C.2) The RO does not have a high probability of being remembered by the user	• The RO selected is the one with maximum value of saliency, the highest probability of being remembered by the user, and the user’s prior Knowledge.

4. Conclusions and Future Work

The automatic generation of spatial language is a complex task that requires the incorporation of a number of factors. All the objects' features in a VE can be explored by the computer system to categorize them by their saliency and their proximity to the user. It can also be established the field of view of the user at each moment of his/her navigation in the scenario. And, it can be included a user modeling to personalize, to some extend, the spatial language. A number of elements that require schematization, in order to automatically generate a proper direction for the user to locate an object in a VE. In this paper we proposed such a scheme, including the algorithms to select a better reference object, to select a proper spatial relation and to generate the spatial language.

Other features can be included in the future to this scheme. Regarding the user modeling, cognitive and emotional perceptions that might probably have an influence in the selection of a reference object. It can be also included other visual features of the objects, for example its texture. Finally, the directions to place the object in the field of view of the user can be included.

References

1. Landau B, Jackendoff R: "What" and "Where" in Spatial Language and Spatial Cognition. *Behavioral and Brain Sciences* 1993, 16(217 - 265).
2. Pederson E, Danziger E, Wilkins D, Levinson S, Kita S, Senft G: Semantic Typology ans Spatial Conceptualization *Linguistic Society of America JSTOR Language* 1998, 74(3):557 - 589.
3. Barclay M: Reference Object Choice in Spatial Language: Machine and Human Models. University of Exeter. PhD. Thesis; 2010.
4. Kelleher JD: A Perceptually Based Computational Framework for the Interpretation of Spatial Language Dublin: Dublin City University. PhD. Thesis; 2003.
5. Gorniak P, Roy D: Grounded Semantic Composition for Visual Scenes. *Journal of Artificial Intelligence Research* 2004, 21:429 - 470.
6. Trinh T-H: A Constraint-based Approach to Modelling Spatial Semantics of Virtual Environments. Université de Bretagne Occidentale. PhD. Thesis; 2013.
7. Lara G, De Antonio A, Peña A: Computerized spatial language generation for object location. *Springer Virtual Reality* 2016:1 - 10.
8. Mou W, McNamara TP: Intrinsic Frames of Reference in Spatial Memory. The American Psychological Association Journal of Experimental Psychology: Learning, Memory, and Cognition 2002, 28(1):162 - 170.
9. Wraga M, Creem SH, Proffitt DR: The influence of spatial reference frames on imagined object and viewer rotations *Elsevier Acta Psychologica* 1998, 102:247 - 264.
10. Titchener EB: Lectures on the Elementary Psychology of Feeling and Attention. In. New York: The MacMillan Company; 1908.

11. Frintrop S, Rome E: Computational Visual Attention Systems and their Cognitive Foundations: A Survey. *ACM Journal Name* 2010, 7(1):1 - 46.
12. Vargas ML, Lahera G: "Asignación de relavancia": Una propuesta para el término inglés "salience". In: *Actas Esp Psiquiatría.* vol. 39. España; 2011: 271 - 272.
13. Lahera G, Freund N, Sáin-Ruíz J: Asignación de relevancia (salience) y desregulación del sistema dopaminérgico. *Elsevier Doyma Revista de Psiquiatría y Salud Mental* 2013, 6(1):45 - 51.
14. Hall D, Leibe B, Schile B: Saliency of Interest Points under Scale Changes. In: *British Machine Vision Conference (BMVC'02).* Cardiff, UK; 2002: 646 - 655.
15. Röser F, Krumnack A, Hamburger K: The influence of perceptual and structural salience. In: Cooperative Minds: Social Interaction and Group Dynamics In Proceedings of the 35th Annual Meeting of the Cognitive Science Society: 2013; Austin, TX. USA; 2013: 3315 - 3320.
16. Gapp K-P: Object Localization: Selection of Optimal Reference Objects. In. *Fed. Rep. of Germany*: *Universität des Saarlandes* 1995: 1-18.
17. Harrington DO, Drake MV: Los campos visuales: texto y atlas de perimetría clínica. In.: Ediciones Científicas y Técnicas; 1993.
18. Gapp K-P: Selection of Best Reference Objects in Objects Localizations. In: In Proceedings of the AAAI Spring Symposium on Cognitive and Comutational Models of Spatial Representations: 1996b; Stanford, CA.; 1996b: 23 - 34.
19. Lara G, Peña A, De Antonio A, Ramírez J, Imbert R: (in press) Comparative analysis of shape descriptors for 3D objects. *Multimedia Tools Applications* 2016a:1 - 48.
20. Lara G, De Antonio, Peña A: A computational model of perceptual saliency for 3D objects in virtual environments; 2016b (in review).
21. Kobsa A: Generic User Modeling Systems. *Springer User Modeling and User-Adapted Intraction* 2001, II:49 - 63.
22. Osterrieth PA: Le test de copie d'une figure complexe. *Arch Psychol* 1944, 30:206-356.
23. Gapp K-P: From vision to language: A cognitive approach to the computation of spatial relations in 3D space. In., Künstliche Intelligenz - Wissensbasierte Systeme. Bericht Nr. 110 edn. Fed. Rep. of Germany: Universität des Saarlandes; 1994: 1 - 19.
24. Gapp K-P: Angle, distance, shape, and their relationship to projective relations. In: In Proceedings of the 17th Annual Conference of the Cognitive Science Society: 1995a; Cognitive Science Society Mahwah, NJ.; 1995a: 112 - 117.