Inducing Frames of Reference

Christian Kray[1]

Abstract. In this paper, we investigate induced frames of reference that provide a new means to improve localisations, to increase the precision of localisations, and to compensate the lack of positional information. We first review existing approaches to classify and define frames of reference in general. We then introduce induced frames of reference and provide several examples for this concept. A prototypical implementation in the context of a mobile tourist guide demonstrates the utility of induced frames of references. We also discuss further uses and application areas as well as benefits and drawbacks of establishing frames of reference in this way.

1 INTRODUCTION

Frames of reference are a key concept in spatial reasoning and spatial language. They describe the context, in which utterances related to space can be decoded, and they provide the basis for several types of spatial relations [2]. In order to unambiguously specify the location and/or direction of objects, frequently a frame of reference is required, which structures the embedding space in a way that allows for relating to this structure. There have been several proposals on how to classify frames of reference such as according to the way in which the origin is defined [3], or depending on the current scope [9]. According to [1] a reference system or a frame of reference is specified by three characteristics: the origin of the coordinate system (which is independent of the kind of coordinate system used), its orientation and its handedness (i. e. the relation between the axes).

Frequently, three basic types of frames of reference are distinguished in the literature (e. g. [10]): deictic, intrinsic, and extrinsic frames of reference. *Deictic* frames of reference designate those frames that inherit their origin, orientation and handedness from the speaker of an utterance. *Instrinsic* frames of reference are established based on an anchor object: it determines the origin of the coordinate system as well as its orientation. Depending on the type of an object, the direction is derived from the topology, size, or shape of the object. For example, if the anchor object is a building, the orientation is often defined by a prominent front and/or by the location of the main entrance. *Extrinsic* frames of reference may also inherit their origin from an anchor object. However, their orientation and handedness is not determined by intrinsic properties but rather by external factors such as the direction of motion.

A further frequent distinction is made between allocentric and egocentric frames of reference. An *allocentric* frame of reference relies on a fixed coordinate system: Its direction and origin is imposed by external factors such as the compass points, and they are independent of the observer's or addressee's current position. Consequently, in an allocentric frame of reference, one can refer to objects in the environment from a survey perspective, e. g. "Go north across the lawn." (see, for example, [13, 14, 15]). In an *egocentric* frame of reference, the origin of the coordinate system is determined by the location of a human observer or addressee, and its orientation is established with respect to the intrinsic body axis.Therefore, egocentric frames of reference can be considered to be a special case of the intrinsic type [4]. However, due to their relevance in practical applications, it makes sense to define a distinct category. Verbal route directions, for example, often rely on an egocentric frame of reference. This view is also known as route or field perspective (cf. [12], [11]).

Levinson [6] provides an overview over different definitions in a number of disciplines, and proposes a unifying classification consisting of three different types of frames of reference. *Intrinsic* frames of reference are defined by the inherent features of an object that serves as relatum. *Relative* frames of reference rely on a viewpoint that is distinct from relatum or the object to be localised. *Absolute* frames of reference refer to a fixed direction (e. g. defined by gravity).

In this paper, we will discuss *induced frames of reference*, which introduce an additional criterion to distinguish frames of reference that is orthogonal to previous systems. Induced frames of reference also provide a new means to address several issues, which were hard to tackle using other frames of reference or which could not be solved at all. Most prominently, these include spatial constellations that are hard to describe using traditional frames of reference as well as situations where the position of the addressee is only partially known or not at all. Additionally, induced frames of reference may help to improve the average quality of localisations.

In the following sections, we will first introduce induced frames of reference. We will then present some examples as well as a prototypical implementation. In the subsequent section we will discuss several applications for this way of establishing a frame of reference and point out benefits and disadvantages. The paper concludes on a short summary of its main contributions.

2 INDUCED FRAMES OF REFERENCE

In addition to the different categories of frames of reference that we discussed above, there is another dimension, along which we can differentiate: the way in which a frame of reference is established. So far, we have only considered frames of reference that are established *directly.* Instead of relying on the immediate establishment of a frame of reference, it is also possible to use *meta-communicative acts* such as turn instructions to *induce* a frame of reference. The resulting induced frames of reference can then be defined as follows:

> An *induced frame of reference* is a frame of reference that requires the listener to first perform one or more mental or physical actions before the frame of reference is established. These actions include rotation and relocation, which may be applied to the origin and/or the orientation of an original frame of reference.

[1] Lancaster University, Computing Department, Lancaster, U.K., email: kray@comp.lancs.ac.uk

The following sentences provide some examples:

(1) *If you stood in front of the church*, the fountain would be to your right.

(2) *If you turn a little bit to the right*, the castle is exactly behind the church.

(3) *Standing on the market square facing the church*, the library is to your left.

From the definition, several conclusions can be drawn: First of all, the actions inducing the frame of reference can be either absolute or relative. In the former case, no information on the orientation or origin from an original frame of reference is required to perform the corresponding operation (see, for example, sentence (1)). In the latter case, the action is relative to an original frame of reference (see sentence (2)), which implies that the corresponding induced frame of reference can only be established if the original one is known. A second (related) observation is that it is also possible to establish an induced frame of reference 'out of the void', e. g. when the inducing actions include absolute reorientation and relocation (see sentence (3)).

An induced frame of reference can also help to generate 'better' relational expressions such as localisations. Usually, the set of available frames of reference in a given situation consists of the ones defined by the listener and the speaker as well as those established by the target object and all potential anchor objects. Applying the orientation of either speaker or listener to any of those objects can yield further frames of reference, but this is really only a special case of inducing a frame of reference. Even if we include the latter ones, it is still possible that there is no combination of a frame of reference, a spatial relation, and an anchor object that yields a satisfactory localisation. In this case, the induction of a frame of reference can help to improve the resulting relational expression since the frame of reference used can be 'adapted' more precisely to the corresponding spatial relation.

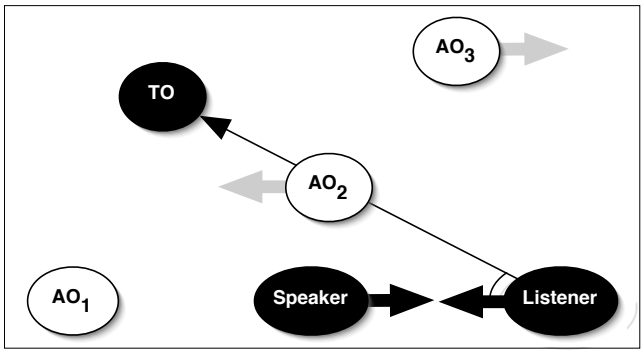

Figure 1. Using an induced frame of reference: an example. (The arrows attached to the objects indicate the orientation of the corresponding frame of reference, i. e. their intrinsic front.)

To illustrate this, consider the following example (see figure 1): A speaker wants to describe the location of the target object TO to a listener while they are facing each other. Of all objects in their environment, only the anchor objects AO_1 through AO_3 are suitable candidates for use in a relational expression, e. g. because all other objects are hard to distinguish or unknown to either the listener or the speaker. However, neither frame of reference established by the potential anchor objects, the speaker, or the listener yields a single spatial relation (such as `left-of`) that applies well to the given situation. Hence, neither deictic, intrinsic or extrinsic frames of reference nor egocentric or allocentric frames allow for a precise and easily comprehensible localisation of TO. While it is possible to introduce additional relations (such as `left-of-in-front`), not all languages provide means to easily verbalize those. However, in this situation the speaker can easily induce a frame of reference by giving a turning instruction such as "Turn towards AO_2.", which will result in very good applicability of the relation `behind(`AO_2`,` TO`)`.[2]

3 A PROTOTYPICAL IMPLEMENTATION

In order to evaluate the utility of induced frames of reference in a real-world application, we implemented the SISTO agent, a component that can generate utterances based on this type of frame of reference [5]. SISTO is a part of a mobile tourist guide [8], where it provides a number of services related to space such as incremental guidance and localisation of sights of interest. Within the context of the latter service, we enabled the use of induced frames of reference that rely on simple turning instructions such as "if you turn right...".

The process of computing a localisation involves three main steps:[3] the determination of potential anchor objects, the establishment of a frame of reference, and the evaluation of spatial relations. These steps are highly interwoven - many relations depend on a frame of reference and/or require an anchor object, some frames of reference are defined by an anchor object, etc.

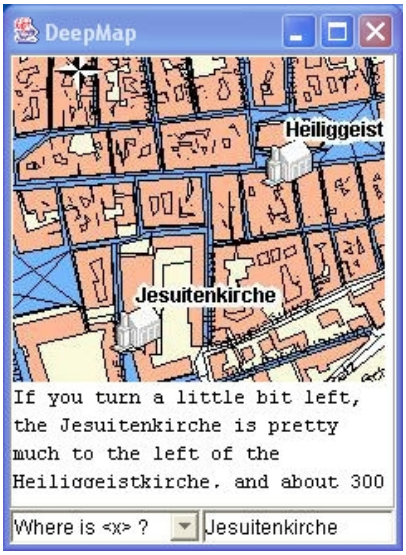

Figure 2. A localisation using an induced frame of reference: the user is standing near the 'Heiliggeistkirche' looking North while asking "Where is the Jesuitenkirche?"

[2] In this example, we assume that the speaker just wants to communicate the location of the target object. Otherwise, the occlusion of TO by AO_2 may be a problem.

[3] That is assuming that the target object is known. Oftentimes, there is the need to first identify it, for example in case the user refers to it anaphorically ("Where is it?")

In our implementation, the SISTO agent used a number of heuristics to first gather a set of potential anchor objects such as 'proximity to target object', 'proximity to user' and 'prementioned'. It then determined a number of potential frames of reference based on those objects and the position of the user as well as induced frames of reference. The engine computed the latter ones by 'rotating' the user towards each of the potential anchor objects and the target object. Finally, it evaluated the applicability of a fixed set of angular and distal relations in combination with each anchor object and each frame of reference. The SISTO agent then selected a tuple of a frame of reference, an anchor object and a relation based on the degree of applicability of the relation and the user-dependent relevance of the anchor object.[4] Figure 2 depicts an example output (without the speech synthesis) based on an induced frame of reference.

Based on our experiences with the system, we can report several initial observations from a number of lab tests and a field trial. As it is possible to enable/disable the use of induced frames of reference through a configuration file, we were able to compare the output in either case for identical situations. Generally, the average rating for the selected localisations was higher if induced frames of references were enabled. Even though these were not always selected, they significantly improved the rating in difficult cases, e. g. when there was only a small number of potential anchor objects or when no combination of a (non-induced) frame of reference and an anchor object yielded a good rating for any relation. This observation coincided with the subjective impression that the localisations generated for these difficult cases were easier to understand and map to the real world when they were using induced frames of references. In comparison, the localisations based on a non-induced frame of references tended to be less precise and left more room for interpretation.

However, we also observed an increasing response time of the system when induced frames of reference were enabled. Even though we only realised one type of induced frames of reference (rotation of the listener), the number of localisations to evaluate effectively doubled. While this was partially compensated by an internal caching system that prevented unnecessary access to the external databases used in process, it stands to expect that the inclusion of additional types of induced frames of reference further impacts the response time. This is one of the issues we discuss in the following section.

4 DISCUSSION

So far, we have focussed on how to use induced frames of reference in relational expressions, i. e. localisations. However, a different perspective that is highly relevant for mobile systems, consists of perceiving induced frames of reference as a means to address the lack of positional information such as the user's current location or orientation. In principle, induced frames of reference enable a system to compensate for the lack of any kind of positional information: When the current viewing direction is unknown, the system can select a direction, which is best suited at the moment, and then precede the actual output with a turn instruction such as "If you turn towards the fountain, ...". Even if the viewing direction is known precisely, there might be another one, which is preferable in the current situation, e. g. in case of a localisation, where an induced frame of reference may increase the degree of applicability of a angular relation.

[4] Each anchor object was evaluated according to a number of situational factors such as 'being visible', 'being of interest to the user', or 'visually salient'. For an detailed description of these factors and the evaluation process, see [5].

If the current location of the user is unknown, the system can analogously select an origin, which best suits the actual purpose, and instruct the user to (either mentally or physically) relocate, e. g. by generating instructions such as "If you stood on the corn market, ...". The same can be done in case of imprecise positional information by inducing precision through statements such as "If you stand exactly in front of the church, ...". Note that this does not require the user to perform a (physical or mental) reorientation, as only the origin of the frame of reference is affected. However, the total lack of positional information may require the combination of both relocation and reorientation (e. g. "If you stood on the corn market facing the church, ..."). Theoretically, this combination can help to address all possible situations from imprecise information on one or more constituents to the total absence of any information. In practice, this is not always feasible. For example, incremental route instructions are of little help if not tailored to the exact position of the user. Therefore, inducing location and orientation prior to giving the corresponding instruction would be infeasible. However, inducing only the orientation of a frame of reference may help to overcome the problem of not knowing where the user is looking at.

While there are a number of benefits resulting from the use of induced frames of reference, there are also some drawbacks. One of these issues is that the listener has to perform one or more mental or physical spatial operations before being able to decode the information based upon the induced frame of reference. Since, for example, mental rotations are very demanding operations in terms of cognitive resources, induced frames of reference can actually increase the 'cognitive load' of the user compared to direct establishment, and are therefore not suited when the user's cognitive resources are strained (e. g. while they are performing a secondary task). However, the trade-off between increased cognitive load and more precise localisations requires further empirical studies.

In addition, the inclusion of induced frames of references in the reasoning process also entails a much higher computational load: When considering only direct establishment, the set of potential frames of reference is restricted to those defined by the listener, the speaker, and by all suitable anchor objects. Depending on the situation (e. g. localisation in an urban environment), the resulting set may already consists of hundreds of candidates. If we take into account that some anchor objects do not have an intrinsic front, this number grows further since we can then apply the orientation of either the listener or the speaker to the corresponding objects.

Even if we do not count the resulting frames of reference as being induced – which we could as the orientation is imposed on a previously directionless frame of reference – the impact of induced frames of reference is still large: For every suitable anchor object (including the listener and the speaker), we have to consider several different orientations instead of a single one. In theory, we could generate an infinite number of candidates by applying every possible orientation. In practice, the number of potentially meaningful orientations is limited, e. g. by the listener's and the speaker's own frame of reference as well as by the target object. The spatial constellation of the target object, the listener and the speaker as well as the potential anchor object also induces some potentially meaningful orientations.

Figure 3 shows and example situation to illustrate these considerations: The intrinsic orientation of listener, speaker, and the anchor object (AO) are depicted using thick arrows, all of which are potentially useful orientations for an induced frame of reference. The two unlabeled circles indicate additional potentially useful origins for an induced frame of reference, which are implied by the constellation of the objects in the scene: these origins allow for certain

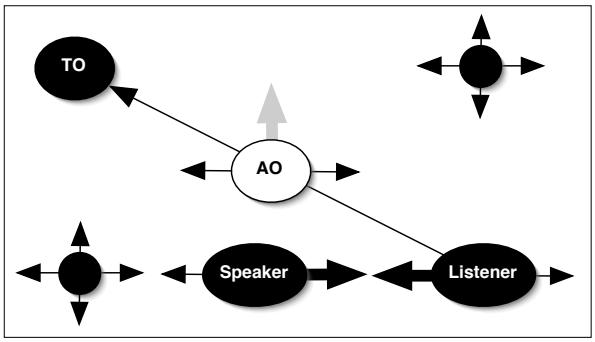

Figure 3. Improving the quality of relational expressions using an induced frame of reference. (Thick arrows indicate intrinsic orientations, thin arrows denote potential orientations for induction.)

angular relations (such as `left-of` or `in-front-of`) to apply perfectly. For example, if there is a good way to induce the origin depicted by the unlabeled circle in the upper right corner, then the speaker could describe the location of TO very precisely using the `in-front-of` relation. A corresponding linguistic realisation is given in sentence (4).

(4) If you stood at `<circle>`, TO is exactly in front of you.

Consequently, there is a need to design mechanisms that drastically reduce the number of frames of references considered by the system. In our prototypical application, we chose to only use induced frames of references resulting from uni-directional rotations of the user towards potential anchor objects. This approach was motivated by the application context: tourist visiting foreign cities have only limited knowledge about local landmarks, and localisations should therefore refer to those objects that are already familiar to them [7, 15]. In a more general setting, additional strategies could be applied such as the 'inverse' application of relations (i. e. determining the frame of reference in which an application has a high degree of applicability (see also Figure 3)). In addition to formally evaluating the acceptance and efficiency of induced frames of reference, designing further strategies to reduce the number of potential frames of reference is a major challenge for future work.

5 CONCLUSION

In this paper, we introduced the notion of inducing frames of reference, which are not established directly but instead require a mental or physical rotation and/or re-location prior to their establishment. We provided several examples, and reported initial observations from a prototypical implementation in the context of a mobile tourist guide. These seem to support the assumption that induced frames of reference may be beneficial in real-world applications.

Based on the observations made with this prototype, we discussed the benefits and costs of using induced frames of reference. On the one hand they can help to improve spatial expressions and to address the lack of positional information. On the other hand, the complexity of computations is increased and requires strategies to limit the number of frames of reference that are evaluated. We demonstrated one such strategy in the context of the example application, and sketched out several others. However, further empirical studies will be required to investigate the trade-offs between increased precision of localisations and increased cognitive load when employing induced frames of reference.

ACKNOWLEDGEMENTS

The research reported here was funded by the Klaus Tschira Stiftung (KTS) in the context of the project SISTO, and by the EPSRC within the RELATE project.

REFERENCES

[1] Andrew U. Frank, 'Formal models for cognition - taxonomy of spatial location description and frames of reference', in *Spatial cognition – An interdisciplinary approach to representation and processing of spatial knowledge*, eds., C. Freksa, C. Habel, and K.F. Wender, 293–312, Springer, Berlin, Heidelberg, New York, (1998).

[2] K.-P. Gapp, 'Basic meanings of spatial relations: Computation and evaluation in 3D space', in *Proceedings of AAAI-94*, pp. 1393–1398, Seattle, WA, (1994).

[3] A. Herskovits, *Language and Spatial Cognition - An Interdisciplinary Study of the Prepositions in English*, Cambridge University Press, Cambridge, UK, 1986.

[4] R.L. Klatzky, 'Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections', in *Spatial cognition – An interdisciplinary approach to representation and processing of spatial knowledge*, eds., C. Freksa, C. Habel, and K.F. Wender, 1–17, Springer, Berlin, Heidelberg, New York, (1998).

[5] Christian Kray, *Situated Interaction on Spatial Topics*, number 274 in DISKI, Akademische Verlags-Gesellschaft AKA, Berlin, 2003.

[6] S. C. Levinson, 'Frames of reference and molyneux's question: Cross-linguistic evidence', in *Language and Space*, eds., P. Bloom, M. Peterson, L. Nadel, and M. Garrett, 109–169, MIT Press, (1996).

[7] K. L. Lovelace, M. Hegarty, and D. R. Montello, 'Elements of good route directions in familiar and unfamiliar environments', in *Spatial Information Theory (Proceedings of COSIT 99). Springer, Berlin*, eds., Christian Freksa and David M. Mark, pp. 65–82, Berlin, Heidelberg, New York, (1999). Springer.

[8] Rainer Malaka and Alexander Zipf, 'Deep Map - challenging IT research in the framework of a tourist information system', in *Information and communication technologies in tourism 2000*, eds., D. R. Fesenmaier, S. Klein, and D. Buhalis, 15–27, Springer, Berlin, Heidelberg, New York, (2000).

[9] Dan R. Montello, 'Scale and multiple phychologies of space', in *Spatial Information Theory: A theoretical basis for GIS*, eds., A. Frank and I. Campari, pp. 312–321, Berlin, Heidelberg, New York, (1993). Springer.

[10] Gudula Retz-Schmidt, 'Various views on spatial prepositions', *AI Magzine*, **9**(2), 95–105, (1988).

[11] K. Schweizer, T. Herrmann, G. Janzen, and S Katz, 'The route direction effect and its constraints', in *Spatial Cognition – An interdisciplinary approach to representation and processing of spatial knowledge*, eds., C. Freksa, C. Habel, and K.F. Wender, 19–38, Springer, Berlin, Heidelberg, New York, (1998).

[12] B. Tversky, 'Cognitive Maps, Cognitive Collages, and Spatial Mental Models', in *Spatial Information Theory. A Theroretical Basis for GIS, COSIT'93*, pp. 14–24, (1993).

[13] B. Tversky and P. U. Lee, 'How Space Structures Language', in *Spatial cognition – An interdisciplinary approach to representation and processing of spatial knowledge*, eds., C. Freksa, C. Habel, and K.F. Wender, 157–177, Springer, Berlin, Heidelberg, New York, (1998).

[14] B. Tversky and P. U. Lee, 'Pictorial and Verbal Tools for Conveying Routes', in *COSIT 99: Spatial Information theory: cognitive and computational foundations of geographic information science*, eds., C. Freksa and D.M. Mark, pp. 37–50, Berlin, Heidelberg, New York, (1999). Springer.

[15] Steffen Werner, Bernd Krieg-Bruckner, Hanspeter A. Mallot, Karin Schweizer, and Christian Freksa, 'Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation', in *Informatik '97 GI Jahrestagung*, ed., M. Jarke, pp. 41–50, Berlin, Heidelberg, New York, (1997). Springer.