Processing Iconic Gestures in a Multimodal Virtual Construction Environment

Christian Fröhlich, Peter Biermann, Marc E. Latoschik, and Ipke Wachsmuth

Artificial Intelligence Group
Faculty of Technology, University of Bielefeld
D-33594 Bielefeld, Germany
{cfroehli,pbierman,marcl,ipke}@techfak.uni-bielefeld.de

Abstract. In this paper we describe how coverbal iconic gestures can be used to express shape-related references to objects in a Virtual Construction Environment. Shape information is represented using Imagistic Description Trees (IDTs), an extended semantic representation which includes relational information (as well as numerical data) about the objects' spatial features. The IDTs are generated online according to the trajectory of the user's hand movements when the system is instructed to select an existing or to create a new object. A tight integration of the semantic information into the objects' data structures allows to access this information via so-called semantic entities as interfaces during the multimodal analysis and integration process.

Keywords: Gesture Representation, Iconic Gestures, Virtual Construction, Multimodal Human-Computer Interaction

1   Introduction

In the description of object shapes, humans usually perform iconic gestures that coincide with speech (are coverbal). Marked by a similarity between the gestural sign and the described object, iconic gestures may easily depict content difficult to describe using words alone. Though the expressive potential of iconic gestures in human-computer communication is generally acknowledged, few applications of non-verbal modalities go beyond pointing and symbolic gestures, one of the first to mention is [7].

Using iconic gestures in virtual construction scenarios can be very useful when it comes to the description of shape-related aspects of certain construction parts. The processing of iconic gestures enables the user to specify the shape of a new object and to reference objects by their shape. To be able to do so, the system needs to have an internal representation formalism for shape-related information of the construction parts.

The next section introduces the Imagistic Description Tree (IDT) formalism [6] as an appropriate representation for shape-related information in multimodal systems. Section 3 will then describe our application scenario and some technical details about the use of IDTs in virtual construction. Section 4 will conclude with a brief outline of possible future work.

2 Defining Shape through Iconic Gestures

Based on a comprehensive corpus of speech-gesture shape descriptions acquired from an empirical study [6], the Imagistic Tree was proposed as a representation for the semantics of multimodal shape-related expressions, to the end of algorithmically interpretive operational shape descriptions from gesture and speech input modalities. It extends an earlier approach [5], which models the two factors of extent and (partial) profile information in gestures, but which has not included structured spatial organization of gesture and accompanying speech reflecting this factor.

The IDT models object extent, profile, and structure, as the salient semantic elements contained in iconic gestures. The basic level of IDT representations are object schemes, in which each object is described by a collection of up to three axes which represent the objects extents in one, two, or three spatial dimensions. Using different combinations of axes in an object schema, several basic objects can be represented, such as cubes, cylinders, etc.

Structural aspects of an object are represented in an imagistic description which can recursively embed further imagistic descriptions for object parts, to result in a tree-like structure similar to the hierarchical structure used in the Marr and Nishihara [3] model. For each such description of an object part, an object schema is used to define its overall proportions, a spatial anchor flag which signals whether the description is spatially anchored in a parent coordinate system, and a transformation matrix defining the position, orientation, and size of the object or part in relation to the parent description. The complete tree describing an object including all parts, parts of parts, etc. is called Imagistic Description Tree. A more detailed description of the formal structure can be found in [4].

Initially, the IDT model was developed as the conceptual basis to represent shape-related information acquired via gesture and speech for usage in an operational gesture understanding system. Capturing gesture (motion) data via data-gloves and motion trackers, the system is able to recognize and to conceptualize shape-related gestures and verbal expressions in a unified spatial representation of an object description and to determine target objects which most closely match the input. In our application system–called the Virtuelle Werkstatt [2]–this procedure has been used to make reference or select objects for further manipulation in the virtual environment. More information on the technical approach will be given in Section 3.

3 Technical Approach

Our application system is concerned with the processing of multimodal user input – natural language as well as deictic, symbolic and iconic gestures – to drive the modifications of a 3D-visualized virtual construction scene. The 3D-visualization is done inside a three-sided CAVE environment to give the user a high degree of immersion. The trajectories of the gestural movements are detected with the help of an optical tracking system, which uses 9 infrared cameras

in conjunction with retroreflective markers. The number of cameras is sufficient to ensure that the gestures are visible to the system at most times. Hand postures and handshapes are detected using data gloves.
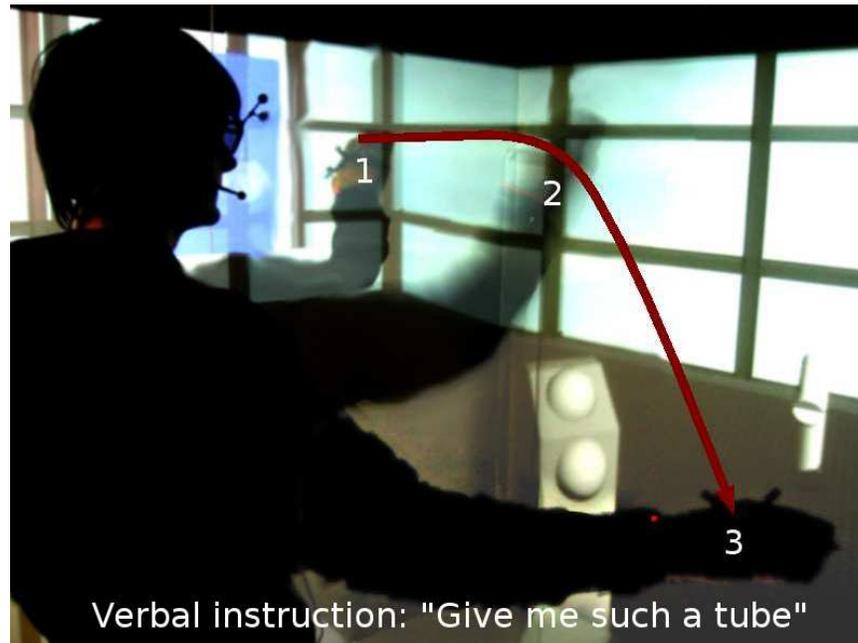

Fig. 1. Defining an object's shape via an iconic gesture, in this case a bent tube

The application domain is a virtual construction scenario. The user is able to create virtual construction parts and to modify them via gestures and speech, e.g. scale a part using a two-handed gesture. The parts are semantically enriched with necessary information about properties (e.g. shape or color) and capabilities (e.g. connection ports). This information is stored using a knowledge representation layer, which is accessed by the multimodal interpretation and analysis modules via semantic entities – virtual placeholders for the semantically enriched parts, which offer a standardized knowledge access interface. The trial task of the application is the creation of a "city mobile", which is a scooter car for handicapped and elderly people, though the general ideas and principles can be applied to a variety of virtual construction scenarios.

In this application IDTs are used to represent shape-related information about the virtual construction parts. Creation and referencing of objects is always done by speech and can be multimodally augmented by adding a gesture, which contains additional information. To create an object the user would for example say something like *"Give me a tube"*. Alternatively the user can interact multimodally with the system by specifying an object's shape with the help of

an iconic gesture, e.g. *"Give me such a tube" + iconic gesture*. The same works for selecting an object. The user can instruct the system in natural language, e.g. by saying *"Take that tube"*, and issue a gesture (deictic as well as iconic) describing the object to be selected.
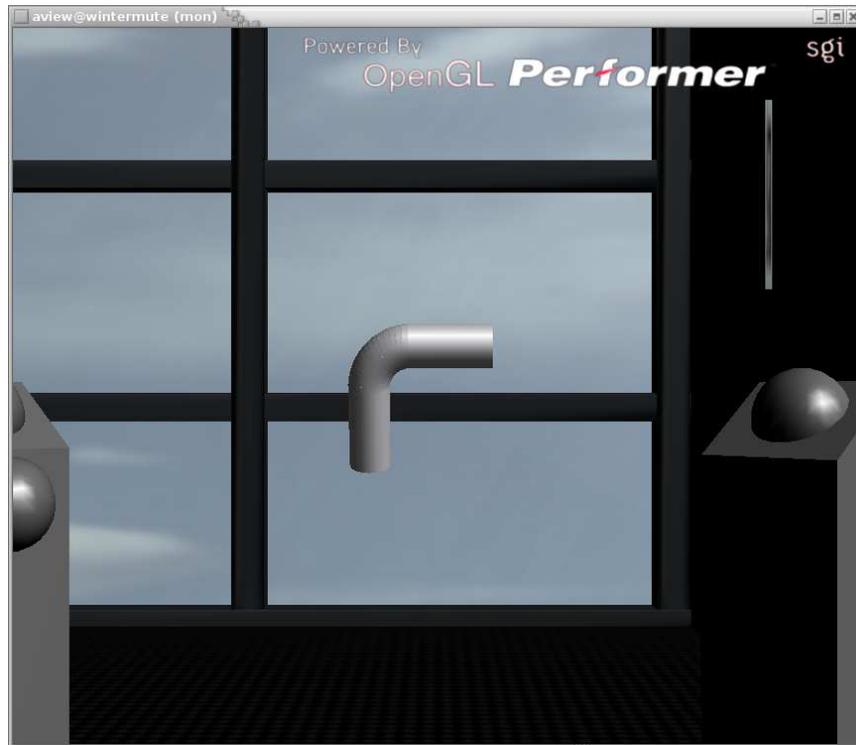

Fig. 2. Newly created construction part resulting from an iconic gesture

A typical gesture accompanying verbal input processed to create a bent object in virtual reality is shown in Figure 1. The bent tube is created by making two linear gesture movements. The system detects the linear segments and computes the angle between them. The length of the first segment is determined by the distance between points *1* and *2*, while the second segment's length is given by the distance between points *2* and *3*. The generated IDT is assigned to the virtual part as semantic information. The selection of a tube by its shape works similar to its creation. The system detects the linear segments, computes the angle and generates an IDT for the gesture. This IDT gets then compared to the IDTs of all available objects and finally the one with the best match is selected. The resulting construction part can be seen in Figure 2. The part consists of two segments and a parametrized angle between them, which can be adjusted after the part was created (for more information on the parametrization of virtual

construction parts, see [1]). The next subsection illustrates the IDT structure of such a construction part.

3.1 Example: IDT Structure of a Virtual Construction Part

As our basic construction parts are rather simple in their spatial structure, their corresponding IDTs are too. Figure 3 shows a virtual part (right) along with the structural outline of its IDT (left).

The bent tube in the figure has a simple tree structure. It has an imagistic description for the whole part, which contains information about the shape of the complete tube itself. Shape meaning in this case extent, position (relative position of the segements towards each other) and also possible verbal descriptions of the object. It also contains two child imagistic descriptions, one for each segment of the bent tube. The segment descriptions have object schemes themselves, which again contain information about their shape. The relevant information which is needed for interacting with them are their relative position towards each other (their transformation matrices) together with their lengths. The angle between the two segments is computed through their transformation matrices. It is parametrized and can be adjusted through speech or gesture later on. The angle itself is not explicitly expressed in the IDT, but it is implictly contained in the relative transformation matrices of both subparts.
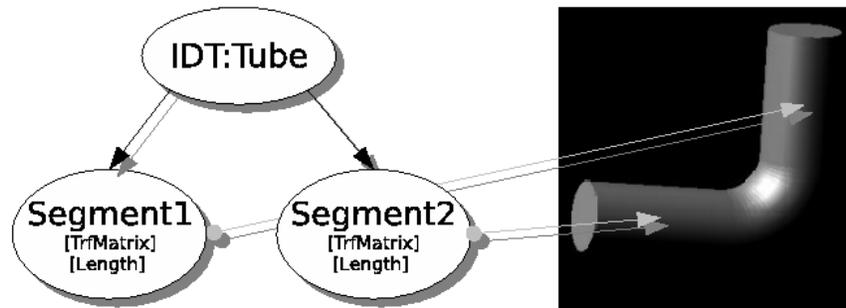

Fig. 3. Simplified IDT structure of a virtual construction part

4 Conclusion and Future Work

In this paper we have presented a way to utilize iconic gestures inside virtual reality environments. We showed how to make use of the Imagistic Description

Tree formalism to enhance virtual construction parts with shape-related information. This enables the user to specify and reference construction parts with the help of iconic gestures. The IDT formalism has turned out to be a very powerful way to represent shapes for virtual construction parts, and we expect that further exploration of the formalism and the use of more of its features will introduce a broader range of interaction capabilities.

Future work consists of integrating more general shape-related concepts like properties such as longish or thin, so the user can for example instruct the system to create an abstract thin object with its corresponding IDT. Furthermore we plan to model and also automatically generate IDT structures for more complex construction parts – like tires or car-seats – to be able to interact with them using iconic gestures as well. It would also be possible to exploit information derived from the handshape when it comes to specifying and referencing of an object.

References

1. P. Biermann, C. Fröhlich, M. E. Latoschik, and I. Wachsmuth. Semantic information and local constraints for parametric parts in interactive virtual construction. In *Proceedings of the 8th International Symposium on Smart Graphics 2007*, pages 124 – 134, Kyoto, Japan, 2007. Springer.
2. P. Biermann, B. Jung, M. Latoschik, and I. Wachsmuth. Virtuelle Werkstatt: A Platform for Multimodal Assembly in VR. In *Proceedings Fourth Virtual Reality International Conference (VRIC 2002)*, pages 53–62, Laval, France, June 2002.
3. H. Marr, D. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. In *Proceedings of the Royal Society*, B 200, pages 269–294, 1978.
4. T. Sowa. Towards the integration of shape-related information in 3-D gestures and speech. In *Proceedings of the Eighth International Conference on Multimodal Interfaces*, pages 92–99, New York, 2006. ACM Press.
5. T. Sowa and I. Wachsmuth. Interpretation of Shape-Related Iconic Gestures in Virtual Environments. In I. Wachsmuth and T. Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction*, LNAI 2298, pages 21–33, Berlin, 2002. Springer.
6. T. Sowa and I. Wachsmuth. Coverbal Iconic Gestures for Object Descriptions in Virtual Environments: An Empirical Study. In M. Rector, I. Poggi, and N. Trigo, editors, *Proceedings of the Conference "Gestures. Meaning and Use"*, pages 365–376, Porto, Portugal, 2003. Edições Universidade Fernando Pessoa.
7. C. J. Sparrell and D. B. Koons. Interpretation of coverbal depictive gestures. In *AAAI Spring Symposium Series*, pages 8–12. Stanford University, March 1994.