

- [Home](#)
- [About the CVR](#)
- [News](#)
- [Members](#)
- [Seminar Series](#)
- [Conference](#)
- [Resources](#)
- [CVR Summer School](#)
- [Research Labs](#)
- [Training at the CVR](#)
- [Partnering with the CVR](#)
- [Contact Us](#)

- Friday, October 3, 2003

Verbal interpretations and visual perception

An essential aspect of everyday communication is the ability of humans to ground verbal interpretations into visual perception. However, many innovating computer applications, like image database search or robotics, suffer from implementations which realize a too narrow human-computer interface. They especially lack of dealing with erroneous intermediate results and individual objects/words not seen before.

My talk will focus on the topic of relating verbal and visual descriptions of the same object. In the first part, I will describe a method that takes advantage of redundant verbal object descriptions in order to infer a correct description of the visual scene despite of speech and object recognition errors. This is part of my thesis work in Bielefeld which is embedded in a bigger project: "situated artificial communicators". The work aims at realizing a human-robot interface in a constructionscenario.

The second part of my talk outlines a framework for learning visual semantics of object nouns. We apply a statistical translation model (IBM model 1) on text-annotated images (e.g. furniture catalogs) and extend it in order to deal with over-segmentations and part-structured objects. The caption of an image is pre-processed by Abney's chunker and segmented regions are characterized by a shape description based on shockgraphs. This part is joint work with Sven Dickinson and Suzanne Stevenson.

Sven Wachsmuth

U Toronto, Comp. Sci. & Bielefeld University, Faculty of Technology, Germany