



**Understanding and Describing Scenes**  
**Larry Zitnick**  
**Microsoft Research Redmond, USA**

**Abstract**

In recent years the ability of computers to recognize objects has greatly advanced. The availability of large training datasets, new learning algorithms and scalable compute platforms have all contributed to this remarkable progress. However, numerous problems still remain. The first part of this lecture discusses the progression of research in object recognition with an emphasis on why certain algorithms succeeded or failed. Results from current state-of-the-art algorithms are analyzed to gain insight into their limitations and areas for future research.

The success of object recognition algorithms has recently enabled the exploration of areas previously thought too difficult. The second part of the lecture explores the increasingly popular area of semantic scene understanding, i.e., how can we describe a scene using a written sentence? The fundamental questions in the area, along with the current state-of-the-art algorithms are described. Interestingly, the problems faced by semantic scene understanding have led to the growing awareness of the need to consider computer vision research within the larger context of artificial intelligence. The lecture concludes by exploring other related and growing research areas including the gathering of commonsense knowledge and dynamic scene understanding.

**Keywords**

Object recognition, object detection, semantic scene understanding, commonsense knowledge