Scene Understanding

Martial Hebert
Carnegie Mellon University, USA

Abstract

In scene understanding we are interested in describing the scene in terms of labeled regions and objects. Many approaches have been proposed that formalize the problem as a labeling problem (pixels, regions, or features) in the image space. After reviewing some of the key techniques, we will identify some of the key challenges involved in developing these approaches further. In particular, we will discuss how the standard classification/learning approaches could be augmented with structural, symbolic, or domain information. To address these challenges, we'll investigate the use of additional information beyond image features.

First, we will review progress in using information about scene geometry and show how, even very weak and partial information can be used effectively for refining scene interpretation.

Geometry, defined broadly to include any partial representation of shape, is critical for relating models of objects and scenes with perceived images. However, it is still difficult to use geometric information effectively. I will discuss several aspects of using geometric information for recognition and scene analysis at different levels, from weak shape features to more complete 3D data. Weak geometric features include contour fragments describing objects' shapes and relative locations of local point features. The difficulty in using this type of information is that it is very ambiguous. Techniques based on targeted use of quantization and on matching approaches can be used to deal with such partial data. Intermediate geometric representation include scene geometric context in which major surfaces are identified and labeled, thus constraining the interpretation of the rest of the image. We will illustrate recent progress in this area through examples in 3D reconstruction of indoor environments, scene parsing, and object recognition.

Beyond geometry, we will explore the use of other sources of information that help us guide the interpretation process. In particular, we will show how knowledge about expected 3D relationships between objects and surfaces can be used effectively to derive more accurate scene models from images. We will also show how scene understanding relates to knowledge about activities in 3D scene space, and how these relationships can be used to generate new information from the image, such as predictions, and to increase the accuracy of the result. Illustrative examples will be provided from recent work from the last couple of years.

Syllabus: Scene Understanding, Recognition in Context