Designing a Comprehensive Visual Recognition System
(Extended Abstract)

Hartmut Neven

Google Inc.
neven@google.com

Abstract. Computer vision has made significant advances during the last decade. Many capabilities such as the detection of faces or the recognition of rigid textured objects such as landmarks are now working to very satisfying levels. Across the various products and services offered by Google we are interested in analyzing an image crawled on the web in all its aspects. When designing such a comprehensive system it becomes obvious however that important abilities are still lacking. One example is object class recognition that scales to thousands or even millions of classes. Another area where we are still facing obstacles is the reliable recognition of objects that have little surface texture and which are largely contour defined. Even a seemingly simple task such as reading text in a photo is still lacking the accuracy we need. The talk describes our efforts in designing a large scale image recognition system that can analyze any given image on the web with respect to many dimensions. We report on the recognition disciplines in which we made good progress but we also call out areas which still require additional work to reach production ready solutions.

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, p. 3, 2010.