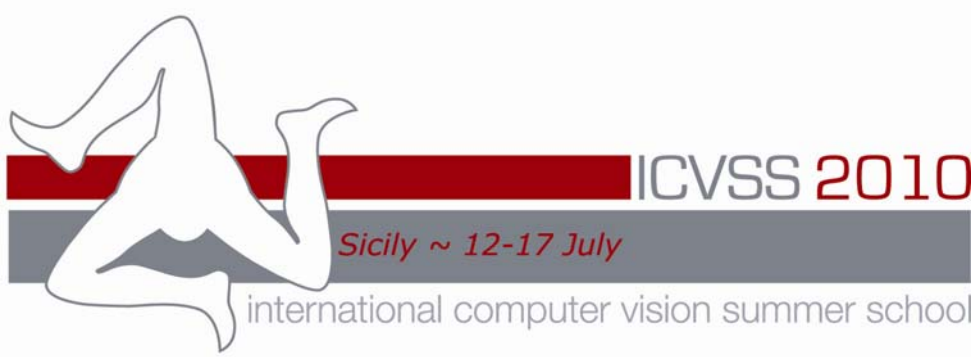

Visual Search and Image Matching

Kristen Grauman

Department of Computer Science
The University of Texas at Austin

Abstract

Algorithms to rapidly search large image or video collections are critical for content-based retrieval and recognition applications. This lecture will overview methods for matching, indexing, mining, and retrieval with large-scale visual data.

How one measures “nearness” in the space of visual descriptors is crucial to the success of many tasks, whether retrieving similar exemplars, mining for noisy instantiations of a central theme, or constructing classifiers between object categories. Generic off-the-shelf metrics are well-studied in computational terms, but can fall short in terms of accuracy. For example, a simple differencing of two images' pixel intensities is easy to compute and scalable to index, but is much too brittle to reveal their semantic similarity. For this reason, researchers create sophisticated representations and associated distance functions to compare them, aiming to capture what is relevant within the complex image or video signals.

First we will discuss some effective representations and distances, such as interest point detectors, local features, visual vocabularies, pyramid matching, and other useful kernel functions. Then we will examine the efficient data structures, embedding functions, and search strategies that can accommodate them, in some cases offering firm guarantees about the number of data points that must be compared to the query (i.e., assuring sub-linear time retrieval). Methods include hashing algorithms, tree-based structures, and branch-and-bound strategies. We will especially emphasize recent work showing how to integrate learning into the indexing approach, thereby specializing the search to better respect learned relationships between objects, scenes, or activities.

Having introduced these tools, we will then explore novel applications of large-scale retrieval and clustering that exploit massive collections in interesting ways---such as using image queries issued from a cell phone to identify a visual match, mining the Internet for iconic images, recognizing geographic locations by example, or discovering rare objects. Finally, we will summarize current benchmark collections and associated evaluation procedures.

Keywords: Image search, matching, hashing, content-based retrieval, mining large-scale visual data