A dual layered linkage system for combining visual and text-based queries for image analysis

Adalbert Wilhelm
Jacobs University Bremen,
Visiting Professor at George Mason University, Fairfax, VA

Joint work with Iulian Ilies[1], Arne Jacobs[2], and Otthein Herzog[2]
[1]Jacobs University Bremen {i.ilies, a.wilhelm}@jacobs-university.de
[2]TZI, Universität Bremen {jarne, herzog}@tzi.de

The World-Wide-Web hosts a constantly expanding collection of images. For fast access and efficient usage of this wealth of visual data, new indexing and retrieval methods are needed, particularly since most of the existing approaches do not exploit all available information. The prevalent approach in current web search engines is to associate images with text such as file names, captions, manually-added tags, or keywords extracted from the surrounding articles. This method restricts the set of searchable images to those associated with text, and can lead to errors if the existent associations are incomplete or incorrect. To enable more efficient queries based on visual information, alternative procedures relying on image processing have been proposed. In semantic search, images are annotated automatically by object or scene recognition algorithms. While this strategy is certainly appealing, progress within the field of image understanding remains limited, with algorithms being successful only in restricted settings such as face recognition. A different strategy is the pure visual search, which relies on low-level image features such as color or textures. However, this type of search is of limited usefulness in textual queries, unless the extracted features represent real concepts.

The goal of the work presented in this talk is to combine the advantages of these different methods into an integrated framework for image search and retrieval that would support both visual queries – e.g. finding images similar to, or keywords relevant for, a given image –, as well as standard text-based queries – e.g. finding images associated with the input keyword, either directly or through similar images. Our approach focuses on the establishment of a dual-layered linkage system between different images based on common keywords and on similar visual features, or, equivalently, on the definition of a co-occurrence-based association scheme between keywords and visual features