

Content Based Retrieval on Very Large Visual Document Archives

© Giuseppe Amato

ISTI-CNR,
Pisa, Italy

firstname.lastname@isti.cnr.it

This work was partially supported by the VISITO Tuscany project, co-funded by the Tuscany Region under the POR CREO FESR program and by the ASSETS project, co-funded by the European Commission within the ICT Policy Support Programme (ICT PSP).

This tutorial will discuss the issues related to content based retrieval in very large dataset of visual documents. Content based retrieval typically is not performed using the visual content itself, rather visual features are extracted and retrieval is performed searching by similarity on the extracted features. Similarity search is a difficult task because efficient techniques to process database or text queries cannot be applied here. Therefore in the last decades researcher have investigated techniques for executing similarity search efficiently and in a scalable way. One popular way to compare similarity between visual documents is the use of global visual features and to measure their similarity (or dissimilarity) by using a similarity (or distance) function. Various indexing strategies and search algorithms based on distance function were defined during the last decade. A relevant research direction has been that of the tree-based access methods, that allow search algorithms just to inspect a small portion of the dataset.

Limitations of tree-based techniques were addressed by defining techniques for approximate similarity search, where

significant improvement boost is obtained at the expense of some minor imprecision in the search result. Recently permutation-based methods, where documents are represented as permutations of a set of reference objects, have been defined. In these methods, similarity between documents is approximated by comparing permutations. Permutations-based indexes allow retrieval to be executed very efficiently, in datasets containing hundred millions images. A new very active field of research is that related to the use of local visual features to compare and retrieve images. Local features offer much higher retrieval quality, however, the efficiency issue is orders of magnitude more difficult. Currently, most techniques are based on a quantization of local features as Bag-of-Word and the use of inverted files (as for instance Lucene). However, the association of words with local features is still a difficult task. Recent new challenging research directions also include the study of techniques for answering to keyword based queries just using the visual content of documents. For instance, suppose you want to retrieve pictures of the "Pisa Leaning Tower", without using any metadata. In this case the problem is twofold. Techniques that offer the same time high accuracy and efficiency should be investigated. A very similar problem is that of automatically annotating pictures. For instance consider the scenario where pictures taken with mobile phones are automatically annotated as soon as they are acquired.

The tutorial will offer an overview of the state of the art in this topic and will discuss open research directions.