VISUAL INFORMATION RETRIEVAL ON THE WEB

Esen OZKARAHAN    Adil ALPKOCAK

Dokuz Eylul University
Department of Computer Engineering
Bornova 35100
Izmir / TURKEY

{esen,alpkocak}@cs.deu.edu.tr

Keywords: Content-based image retrieval, Web searching, image indexing and searching, image retrieval, digital libraries, distributed information retrieval.

Abstract: Today on the Internet there is a wide variety of text based search engines, however the same is not true for searching visual information placed in Internet Web pages. There is increased activity and research for querying such databases especially for content based visual querying. The heterogeneous, distributed and transient nature of visual information, lack of interoperable retrieval systems and the limited bandwidth of Web environment presents bottlenecks for such efforts. In this study the difficulties of visual information retrieval on the Web are highlighted and a visual information retrieval system in such an environment is presented.

One of the most important tasks of a visual search engine on Web is the collection of visual data from Web pages. In our system, the visual data collection process is carried out by a series of automated software agents that traverse the Web detecting visual data from HTML tags. The system builds several indexes for visual data based upon its internal and external attributes. Internal attributes are feature vectors extracted from the original image such as color, texture and shapes. Retrieval process of internal attributes requires a special access mechanism. External attributes are external descriptors such as width, height and format, and the textual descriptions extracted automatically from the HTML documents and their URLs containing the image. All the attributes and thumbnail images are stored in the database along with the collected visual data. Thumbnails are used to reduce the amount of data transferred in presentation of the query result.

Another important part of the system is visual query interfacing which is a current research topic. Visual queries are grouped into three major groups: query by example, navigational query and expressive queries. Recent research on this field has shown that information is meaningful only when it can be retrieved through an expressive query. To achieve this, we are developing tools and interfaces using JAVA language to help users in expressing their visual queries.