© 2013 FIZ Karlsruhe

**io-port 05787434**

**Gennaro, Claudio; Amato, Giuseppe; Bolettieri, Paolo; Savino, Pasquale**

**An approach to content-based image retrieval based on the Lucene search engine library.**

Lalmas, Mounia (ed.) et al., Research and advanced technology for digital libraries. 14th European conference, ECDL 2010, Glasgow, UK, September 6–10, 2010. Proceedings. Berlin: Springer (ISBN 978-3-642-15463-8/pbk). Lecture Notes in Computer Science 6273, 55-66 (2010).

Summary: Content-based image retrieval is becoming a popular way for searching digital libraries as the amount of available multimedia data increases. However, the cost of developing from scratch a robust and reliable system with content-based image retrieval facilities for large databases is quite prohibitive. In this paper, we propose to exploit an approach to perform approximate similarity search in metric spaces developed by [3,6]. The idea at the basis of these techniques is that when two objects are very close one to each other they 'see' the world around them in the same way. Accordingly, we can use a measure of dissimilarity between the views of the world at different objects, in place of the distance function of the underlying metric space. To employ this idea the low level image features (such as colors and textures) are converted into a textual form and are indexed into the inverted index by means of the Lucene search engine library. The conversion of the features in textual form allows us to employ the Lucene's off-the-shelf indexing and searching abilities with a little implementation effort. In this way, we are able to set up a robust information retrieval system that combines full-text search with content-based image retrieval capabilities.

*Keywords:* approximate similarity search; access methods; Lucene
doi:10.1007/978-3-642-15464-5_8