Question B: Paper Review

Title: A Method for Measuring the Complexity of Image Databases

Author: Aibing Rao, Rohini K. Srihari, Lei Zhu, and Aidong Zhang

Published in IEEE Transactions on Multimedia Vol. 4, No. 2. June 2002

Problem Statement

This paper is motive by the state of the art for evaluating image retrieval. Although research on CBIR has been actively pursued for more than a decade, relatively little research effort is on effective evaluation of CBIR techniques and systems. The evaluation methods are quite chaotic. Researchers design different algorithms and then test the performance on their own testbeds.

The typical procedure for evaluation is: 1) set up an image testbed; 2) select a query set as well as the ground-truth; 3) after retrieval tasks on the query set are performed, certain metrics such as precision and recall are calculated based on the relevance of the retrieval results to the images in the database and the ground-truth; 4) uses these metrics to compare the effectiveness of different approaches. One weakness of this performance evaluation is that it highly dependent on the type of the testbed and the query set selected. The second disadvantage is human subjectivity is unavoidable in this procedure. In the step 2, some researchers might only make choice in favor of their own techniques. Manually Making judgments of "relevant" or "irrelevant" in step 3 introduce intensional and unintensional errors. Beside, the metrics such as precision and recall themselves are inappropriate due to the tedious process of measuring relevance and the human subjectivity.

In summary, until now there are no highly satisfactory methods for measuring the effectiveness of the CBIR techniques. Also, there is no common testbed, and no theory on how to compare different testbeds. The lack of a uniform evaluation methodology is clearly impediment to the development of the multimedia retrieval field.

In this paper, the authors develop a method for quantitatively measuring the complexity of image databases (CID) with respect to CBIR. The objective and formal evaluation methodology established in this paper can also be applied in the design of other measurement. This paper provides a guideline to generalize text retrieval techniques into image retrieval. More research results could be obtained by this generalization.

Detail contributions:

- Conduct experiments to verify an important observation that the performance of an image retrieval system is strongly related to the type of the image testbed. This observation suggests that the types of image are needed to be distinguished before image retrieval techniques are evaluated, which is lacked in most performance evaluation.
- Identify the aspects related to the measure of CID: Four aspects are defined: (1) Homogeneity which refers to the degree of perceptual similarity among the images. (2) Heterogeneity which refers to the degree of perceptual dis-similarity among the images. (3) Content variety which refers to different complexity levels of image semantics. It is harder to retrieve images from a database where images have complex contents than from a database where images have only simple contents. (4) Cardinality which refers to the number of images in the database. These aspects are needed to be reflected in the measurement of CID.
- Present a formal framework for measuring CID: Two conditions that a measure of CID must satisfy are formally defined. The first condition states that the complexity after merging two databases together is upper-bounded by the product of the individual complexities. The second condition states that a less effective algorithm superimposes higher complexity on the same database than a more effective one does. The CID can be used to measure the degree of difficulty to retrieve images from the image database. It also can be used to distinguish between types of image databases.
- Propose Query-based approach to compute CID: For the settings that an image database is considered as a black box where only query by example is supported, an approach to compute CID is proposed. It is equivalent to the traditional precision-recall approach. The average precision is defined and CID is computed as the average precision across all queries. This approach is based on the overall performance of a set of standard queries.
- Propose Block-based approach to compute CID: For the settings that an image is considered as a list of labels of predefined keyword subimages, called blocks, a novel approach to compute CID is proposed. This approach is based on statistics of the image database and information theory. It is query-independent. This approach is a natural generalization of the techniques from statistical language modeling in text processing to statistical image database analysis. The concept of perplexity in text processing and speech recognition, which is a measure of the internal complexity of a text corpus with respect to word prediction, is generalized to the concept of CID. The CID metric is not only

independent of queries, hence it is objective but also is a comprehensive measurement which reflects the degree of homogeneity, heterogeneity, content variety, and cardinality.

- Present a statistical analysis of the image database: The N-block approach which generalizes the idea of N-gram in text domain is presented. The correlations of image blocks within an image are the focus and revealed by the probabilities of groups of codes distributed according to certain spatial configurations. The N-block approach is used to obtain the probability distribution needed by computing CID.
- Conduct experiments on both synthetic and real-world images: Comprehensive experiments are conducted on both synthetic and real-world images. The experiments demonstrate that the proposed measurement is related to the performance empirically obtained from the standard query set and consistent with the perception intuition: a simple image database has a lower complexity. Some interesting conclusions are drawn: (1) For any given technique, most of the complexities of the animal databases are higher than those of scenery databases. (2) For any given technique, most of the complexities of the object and human databases are lower than those of scenery databases. (3) For most of the databases, the complexities with different techniques are consistent. (4) Bi-block model significantly reduces the complexities in the block-based approach.

Disadvantages:

- Drawback of query-based approach: One of the disadvantages of the query-based approach is the difficulty of designing the set of queries. To compensate this disadvantage, the database itself is usually used as the standard query set in the computation. However, it requires the relevance judgment of the whole database.
- Drawback of N-block approach: Comparing to text modeling, there is a drawback in the N-block approach. Since sentences in text have the natural linear ordering constraint, it is appropriate to assume the constraint in text modeling. However, the assumption is not natural for images. To be consistent, the proposed models all assume the left-to-right, top-to-bottom ordering for conditional dependencies. The constraint is added for simplicity. It should be removed for more general cases.