

2010

Performance Study of the Most Commonly Used Image Retrieval Evaluation Methods

Chatzichristofis, Savvas A.

ACTA PRESS

<http://hdl.handle.net/11728/10156>

Downloaded from HEPHAESTUS Repository, Neapolis University institutional repository

PERFORMANCE STUDY OF THE MOST COMMONLY USED IMAGE RETRIEVAL EVALUATION METHODS

Savvas A. Chatzichristofis and Yiannis Boutalis
Department of Electrical & Computer Engineering,
Democritus University of Thrace,
67100, Xanthi,
schatzic@ee.duth.gr, ybout@ee.duth.gr

ABSTRACT

One of the most significant problems in the field of image retrieval results from the lack of a common test bed for the evaluation of the systems. So far, many of the methods used in the field of information retrieval have been adopted in order to evaluate the retrieval results. This paper describes the most commonly used methods for retrieval evaluation and notes their weaknesses. It also proposes a new method of measuring the performance of retrieval systems. The proposed method is generic and can be used for evaluating the retrieval performance of any type of information. The proposed evaluation technique has been implemented and is used to evaluate the retrieval results of the img(Rummager)¹ system.

KEY WORDS

Image Retrieval Evaluation Methods, MAP, ANMRR, MNRO

1 Introduction

The objective of an image retrieval system is to retrieve images in rank order, where the rank is determined from the relevance to the query at hand [1]. The image retrieval process can be executed either with the use of a key word integration technique 'upon' the images (Keyword Based Image Retrieval) or with the use of low-level characteristics exported from the image's visual content (Content Based Image Retrieval). Content based image retrieval (CBIR) is defined any technology, that in principle helps to organize digital image archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR [2].

The performance of a content image retrieval system is measured using the following procedure. A certain benchmark image database is initially used. Classic examples of such databases are the Wang [3] database, the UCID database [4] and the Nister database [5]. Each database is comprised of a number of N images and includes a total of Q queries. An image that is to be used as input to a system in order to evaluate its performance is designated as query.

For each query there is a given total number of images with visual similarity which are considered as the ground truth. The system's performance is calculated using a technique to evaluate the order in which the images which form the ground truths for all the queries appear. Section 2 provides an in-depth analysis of the two most commonly used retrieval evaluation methods. The Mean Average Precision and the Average Normalized Modified Retrieval Rank [6].

This paper proposes a new technique for evaluating the order in which the retrieval results appear for each query. The proposed method, called the Mean Normalized Retrieval Order (MNRO), has been designed with the purpose of overcoming the weaknesses apparent in existing evaluation methods. The MNRO indicator grades each retrieval system with a value that fluctuates in the range of [0, 1]. Section 3 describes the proposed method and outlines its advantages over other performance evaluation techniques.

The method is experimentally tested in Section 4, based on artificial query trials and evaluations. and finally, the conclusions are drawn in Section 5.

2 Image Retrieval Evaluation Methods

The most widespread image retrieval evaluation methods with the ability to evaluate the systems using only one number are AP (Average Precision) and NMRR (Normalized Modified Retrieval Rank).

The average precision AP for a single query q is the mean over the precision scores after each retrieved relevant item:

$$AP(q) = \frac{1}{NG(q)} \sum_{k=1}^{NG(q)} P_q(R_k)$$

where R_k is the recall after the k^{th} relevant image was retrieved. The mean average precision (MAP) is the mean of the average precision scores over all queries:

$$MAP = \frac{1}{Q} \sum_{q \in Q} AP(q)$$

where Q is the set of queries q .

¹<http://www.img-rummager.com>