

The Australian National University

COMP8740 Artificial Intelligence Project  
Research/Implementation Project Plan

---

# Efficiency Analysis of Quasi-Random Sampling Algorithm for Content-Based Image Retrieval

Uni ID: u4476115 (MComp)  
Name: Jyun-Hao Huang  
Supervisors: Dr. Jun Zhou  
Dr. Antonio Robles-Kelly

## Background

Content-based image retrieval (CBIR) is an important problem in the areas of pattern recognition, computer vision and e-business systems, which has derived into a large body of research. Several commercial systems have been proposed and implemented, such as QBIC (Query By Image Content) [1], FourEyes [2] and SQUID (Shape Queries Using Image Databases) [3].

## Problem/Task description

A common way of building CBIR systems is using different image processing techniques to extract content information from the images, known as Image Description, and the similarity between two images is obtained by calculate the distance between their corresponding Image Descriptions. Consider a dataset which contains  $n$  images, and a query image has been given. In order to find the best-match image to the query image from the dataset, we need to compare the query image to each image in the dataset, that is, we need  $n$  comparisons. This is extremely inefficient since the calculation of the similarity between images is a very time-consuming task.

Recently, Dr. Jun Zhou and Dr. Antonio Robles-Kelly from NICTA have proposed a novel quasi-random sampling (QuaRS) approach for CBIR [4]. This approach uses EM algorithm to organize the images in the database into compact clusters, then compares the similarity between the query and the clustered images to govern the sampling process within clusters. The sampling can be viewed as a stratified sampling one which is random at the cluster level and takes into account the intra-cluster structure of the dataset. This approach leads to a measure of statistical confidence that relates to the theoretical hard-limit of the retrieval performance. Although this approach is expected to be efficient when applied to very large database, there is no theoretical and experimental analysis of the efficiency of this sampling algorithm.

The main task of this project is to transplant the current MATLAB<sup>®</sup> code of the QuaRS algorithm into C++ platform under the operating system chosen, then implement other image retrieval systems on benchmark image database, for example KD Tree and Vocabulary Tree [5], using the same image description as in the QuaRS algorithm. The experiments will be performed on two aspects. Firstly, the time efficiency of the algorithm shall be compared. Secondly, the quality of the retrieval performance shall be compared.

## Plan / Schedule

The following is the proposed schedule

### Project Schedule

| ID | Job Description                 | Start      | Finish     | Period | August, 2008 |      |     |      |      |      |      | September, 2008 |      |      |      |      | October, 2008 |       |  |
|----|---------------------------------|------------|------------|--------|--------------|------|-----|------|------|------|------|-----------------|------|------|------|------|---------------|-------|--|
|    |                                 |            |            |        | 7/20         | 7/27 | 8/3 | 8/10 | 8/17 | 8/24 | 8/31 | 9/7             | 9/14 | 9/21 | 9/28 | 10/5 | 10/12         | 10/19 |  |
| 1  | Identify Task and Requirement   | 2008/7/21  | 2008/8/1   | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |
| 2  | Collecting Information          | 2008/7/28  | 2008/8/8   | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |
| 3  | Coding Basic Tools (Windows)    | 2008/8/4   | 2008/8/15  | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |
| 4  | Implement QuaRS                 | 2008/8/18  | 2008/8/29  | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |
| 5  | Implement KD TREE               | 2008/9/1   | 2008/9/12  | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |
| 6  | Perform and Analyze Experiments | 2008/9/15  | 2008/9/26  | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |
| 7  | Writing Reports                 | 2008/9/29  | 2008/10/10 | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |
| 8  | Prepare Presentation            | 2008/10/13 | 2008/10/24 | 2w     | ■            |      |     |      |      |      |      |                 |      |      |      |      |               |       |  |

References:

- [1] W. Niblack. The QBIC project: Querying images by content using color, texture and shape. In Proc. SPIE Conference on Storage and Retrieval of Image and Video Database, pages 173-187, 1993.
- [2] R. W. Picard. Light-years from Lena: Video and image libraries and the future. In International Conference on Image Processing, volume 1, pages 310-313, 1995.
- [3] S. A. M. Farzin and J. Kittler. Robust and efficient shape indexing through curvature scale space. In Proceedings of the 7<sup>th</sup> British Machine Vision Conference, volume 1, pages 53-62, 1996.
- [4] J. Zhou and A. Robles-Kelly. A quasi-random sampling approach to image retrieval. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.
- [5] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006.