

**Research/Implementation project proposal:  
Efficiency Analysis of Quasi-Random Sampling Algorithm for Contents-Based  
Image Retrieval**

- ◆ Appropriate Courses: COMP8740 (AI), COMP8760 (CSci), COMP8770 (eSci)
- ◆ Status: Proposition
- ◆ Student: Jyun-Hao Huang, u4476115, COMP8740
- ◆ Supervisors: Dr. Jun Zhou, Dr. Antonio Robles-Kelly
- ◆ Research Areas: computer vision, pattern recognition, algorithms, data structure
- ◆ Technical Difficulty Level: moderate
- ◆ Conceptual Difficulty Level: moderate

**Description**

Contents-based image retrieval (CBIR) is an important problem in the areas of pattern recognition, computer vision and e-Business, 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].

Recently, Zhou and Robles-Kelly 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 vary large database, there is no theoretical and experimental analysis of the efficiency of this sampling algorithm.

The aim of this project is to analyse the efficiency of the QuaRS algorithm. It requires theoretical analysis of the complexity of the algorithm and then comparing it to other image retrieval systems. The student is required to transplant the current Matlab code of the QuaRS algorithm into C++ platform, then implement several mage retrieval algorithms on benchmark image database, for example, KD Tree and Vocabulary Tree [5], using the same image description as in the QuaRS algorithm. The experiments shall 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 by analysis of the ROC curves.

**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 Databases, 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 7th 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.