

# EMR: A Scalable Graph-based Ranking Model for Content-based Image Retrieval

## ABSTRACT

*Graph-based ranking models have been widely applied in information retrieval area. In this paper, we focus on a well known graph-based model - the Ranking on Data Manifold model, or Manifold Ranking (MR). Particularly, it has been successfully applied to content-based image retrieval, because of its outstanding ability to discover underlying geometrical structure of the given image database. However, manifold ranking is computationally very expensive, which significantly limits its applicability to large databases especially for the cases that the queries are out of the database (new samples). We propose a novel scalable graph-based ranking model called Efficient Manifold Ranking (EMR), trying to address the shortcomings of MR from two main perspectives: scalable graph construction and efficient ranking computation. Specifically, we build an anchor graph on the database instead of a traditional  $k$ -nearest neighbor graph, and design a new form of adjacency matrix utilized to speed up the ranking. An approximate method is adopted for efficient out-of-sample retrieval. Experimental results on some large scale image databases demonstrate that EMR is a promising method for real world retrieval applications.*

---

## **EXISTING SYSTEM:**

- ❖ Most traditional methods focus on the data features too much but they ignore the underlying structure information, which is of great importance for semantic discovery, especially when the label information is unknown.
- ❖ Many databases have underlying cluster or manifold structure. Under such circumstances, the assumption of label consistency is reasonable. It means that those nearby data points, or points belong to the same cluster or manifold, are very likely to share the same semantic label. This phenomenon is extremely important to explore the semantic relevance when the label information is unknown. In our opinion, a good CBIR system should consider images' low-level features as well as the intrinsic structure of the image database.

## **DISADVANTAGES OF EXISTING SYSTEM:**

- ❖ It has expensive computational cost, both in graph construction and ranking computation stages.
- ❖ Particularly, it is unknown how to handle an out-of-sample query efficiently under the existing framework.
- ❖ It is unacceptable to recompute the model for a new query. That means, original manifold ranking is inadequate for a real world CBIR system, in which the user provided query is always an out-of-sample.

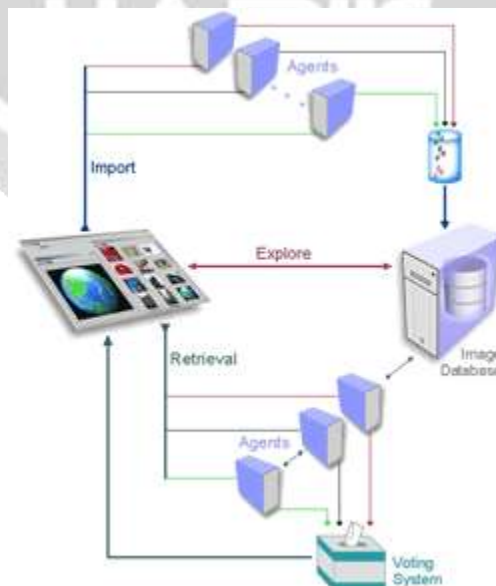
## **PROPOSED SYSTEM:**

- ❖ In this paper, we extend the original manifold ranking and propose a novel framework named Efficient Manifold Ranking (EMR).
- ❖ We try to address the shortcomings of manifold ranking from two perspectives: the first is scalable graph construction; and the second is efficient computation, especially for out-of-sample retrieval.
- ❖ Specifically, we build an anchor graph on the database instead of the traditional k-nearest neighbor graph, and design a new form of adjacency matrix utilized to speed up the ranking computation.
- ❖ The model has two separate stages: an offline stage for building (or learning) the ranking model and an online stage for handling a new query.
- ❖ With EMR, we can handle a database with many images and do the online retrieval in a short time. To the best of our knowledge, no previous manifold ranking based algorithm has run out-of-sample retrieval on a database in this scale.

## **ADVANTAGES OF PROPOSED SYSTEM:**

- ✓ We show several experimental results and comparisons to evaluate the effectiveness and efficiency of our proposed method EMR on many real time images.
- ✓ We can run out-of sample retrieval on a large scale database in a short time.
- ✓ Our model EMR can efficiently handle the new sample as a query for retrieval. In this subsection, we describe the light-weight computation of EMR for a new sample query. We want to emphasize that this is a big improvement over our previous conference version of this work, which makes EMR scalable for large-scale image databases.

## **SYSTEM ARCHITECTURE:**



## SYSTEM SPECIFICATION

### Hardware Requirements:

- System : Pentium IV 3.4 GHz.
- Hard Disk : 40 GB.
- Monitor : 14' Colour Monitor.
- Mouse : Optical Mouse.
- Ram : 1 GB.

### Software Requirements:

- Operating system : Windows Family.
- Coding Language : J2EE (JSP,Servlet,Java Bean)
- Data Base : My Sql.
- IDE : Eclipse - Galileo
- Web Server : Tomcat 5.0/6.0
- Web Designing : Dream Viewer
- Documentation : MS Office

