*IJCOT -Special Issue– The Malla Reddy National Conference on Information System and Knowledge Engineering (MRNC-ISKE 2013 )- July 2013*

# Replacement Policies for Improving Cache Performance

S.Jyothi*1, M.Harathi*2

M.Tech, Dept of CSE, SKDEC, Gooty, D.t: Anantapuram, A.P, India

M.Tech, Assistant Professor, Dept of CSE, SKDEC, Gooty, D.t: Anantapuram, A.P, India

**ABSTRACT:**

In this paper pivoting M-tree (PM-tree) is introduced, a metric access method combining M-tree with the pivot-based approach. While in M-tree a metric region is represented by a hyper-sphere, in PM-tree the shape of a metric region is determined as an intersection of the hyper-sphere and a set of hyper-rings. The set of hyper-rings for each metric region is related to a fixed set of pivot objects. As a consequence, the shape of a metric region bounds the indexed objects more tightly which, in turn, improves the overall efficiency of the similarity search. Preliminary experimental results on a synthetic dataset are included. A new access method, called M-tree, is proposed to organize and search large data sets from a generic "metric space", i.e. where object proximity is only defined by a distance function satisfying the positivity, symmetry, and triangle inequality postulates. We detail algorithms for insertion of objects and split management, which keep the M-tree always balanced - several heuristic split alternatives are considered and experimentally evaluated. Algorithms for similarity (range and k-nearest neighbors) queries are also described. Results from extensive experimentation with a prototype system are reported, considering as the performance criteria the number of page I/O's and the number of distance computations. The results demonstrate that the M-tree indeed extends the domain of applicability beyond the traditional vector spaces, performs reasonably well in high-dimensional data spaces, and scales well in case of growing files.

## I.INTRODUCTION:

Recently, the need to manage various types of data stored in large computer repositories has drastically increased and resulted in the development of *multimedia database systems* aiming at a uniform management of voice, video, image, text, and numerical data. Among the many research challenges which the multimedia technology entails – including data placement, presentation, synchronization, etc. – content-based retrieval plays a dominant role. In order to satisfy the information needs of users, it is of vital importance to effectively and efficiently support the retrieval process devised to determine which portions of the database are relevant to users' requests.

In particular, there is an urgent need of indexing techniques able to support execution of *similarity queries*. Since multimedia applications typically require complex *distance functions* to quantify similarities of multi-dimensional features, such as shape, texture, color [FEF+94], image patterns [VM95], sound [WBKW96], text, fuzzy values, set values [HNP95], sequence data [AFS93, FRM94], etc., multi-dimensional (*spatial* ) access methods (SAMs), such as R-tree [Gut84] and its variants [SRF87, BKSS90], have been considered to index such data. However, the applicability of SAMs is limited by the following assumptions which such structures rely on:

1. Objects are, for indexing purposes, to be represented by means of feature values in a *multidimensional vector space*;
2. The (dis)similarity of any two objects has to be based on a distance function which does not introduce any correlation (or "cross-talk") between feature values [FEF+94]. More precisely, an *Lp* metric, such as the Euclidean distance, has to be used. applications. Consequently, no attempt in the design of these structures has been done to reduce the number of distance computations.

A more general approach to the "similarity indexing" problem has gained some popularity in recent years, leading to the development of so-called *metric trees* (see [Uhl91]). Metric trees only consider relative distances of objects (rather than their absolute positions in a multi-dimensional space) to organize and partition the search space, and just require that the function used to measure the distance (dissimilarity) between objects is a *metric* (see Section 2), so that the *triangle inequality* property applies and can be used to prune the search space.

**II. BACKGROUND**

The Fundamentals of Similarity Search in Metric Spaces

During the last two decades, similarity searching in metric spaces has become intensively investigated research area, as documented in several excellent monographs and surveys the next paragraphs, we will shortly summarize basic fields of the content-based similarity searching using metric space model. For more details, see one of the referred publications. Content-based Image Retrieval The earliest use of the term *content-based image retrieval* in the literature seems to have been by Kato [1992], to describe his experiments into automatic retrieval of images from a database by colour and shape feature. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as colour, texture and shape) that can be automatically extracted from the images themselves. The features used for retrieval can be either primitive or semantic, but the extraction process must be predominantly automatic. Retrieval of images by manually-assigned keywords is definitely not CBIR as the term is generally understood – even if the keywords describe image content. CBIR differs from classical information retrieval in that image databases are essentially unstructured, since digitized images consist purely of arrays of pixel intensities, with no inherent meaning. One of the key issues with any kind of image processing is the need to extract useful information from the raw data (such as recognizing the presence of particular shapes or textures) before any kind of reasoning about the image's contents is possible. Image databases thus differ fundamentally from text databases, where the raw material (words stored as ASCII character strings) has already been logically structured by the author [Santini and Jain, 1997]. There is no equivalent of level 1 retrieval in a text database. CBIR draws many of its methods from the field of image processing and computer vision, and is regarded by some as a subset of that field. It differs from these fields principally through its emphasis on the retrieval of images with desired characteristics from a collection of significant size. Image processing covers a much wider field, including image enhancement, compression, transmission, and interpretation. While there are grey areas (such as object recognition by feature analysis), the distinction between mainstream image analysis and CBIR is usually fairly clear-cut. An example may make this clear. Many police forces now use automatic face recognition systems. Such systems may be used in one of two ways. Firstly, the image in front of the camera may be compared with a single individual's database record to verify his or her identity. In this case, only two images are matched, a process few observers would call CBIR. Secondly, the entire database may be searched to find the most closely matching images. This is a genuine example of CBIR.

Research and development issues in CBIR cover a range of topics, many shared with mainstream image processing and information retrieval. Some of the most important are:

- understanding image users' needs and information-seeking behaviour
- identification of suitable ways of describing image content
- extracting such features from raw images

- providing compact storage for large image databases
- matching query and stored images in a way that reflects human similarity judgements
- efficiently accessing stored images by content
- providing usable human interfaces to CBIR systems

Key research issues in video retrieval include:

- automatic shot and scene detection
- ways of combining video, text and sound for retrieval
- effective presentation of search output for the user.

## III. Replacement Policies

### 3.1 Image use in the community

It is a truism to observe that images are currently used in all walks of life. The influence of television and video games in today's society is clear for all to see. The commonest single reason for storing, transmitting and displaying images is probably for recreational use, though this category includes a wide variety of different attitudes and interaction styles, from passively watching the latest episode of a soap opera to actively analysing a tennis star's shots in the hope of improving one's own game. Images are increasingly used to convey information, in areas as diverse as map-making, weather forecasting and mail-order shopping, and to persuade or convey a mood, as in advertising. They can also be appreciated in their own right, as works of art. A detailed sociological study of image use would be out of place in this report, particularly as there is currently little evidence for the existence of different user communities with different needs. Most individuals interact with images in different ways at different times, perhaps spending an hour in an art gallery one day, and watching a sports video the next. Trying to categorize such behaviour by user type does not seem very useful.

### 3.2 Professional groups making use of images

In the realm of professional image use, the situation is rather different. While there are certainly differences in style between individual design engineers, for example, the nature of the design process imposes a number of inescapable constraints within which all engineers must work. Hence it is possible to generalize to some extent about the way images are used by different professions. Since this report is primarily concerned with image storage and retrieval, it makes sense to limit our discussion by concentrating on uses which involve stored collections of images in some way.

Some groups of people use images in their job on a daily basis, such as graphic designers and illustrators, whilst others may never be required to use them, such as bank managers and accountants. There is a wide range of professions lying between these two extremes, including medicine and law. Other groups of workers, such as librarians and museum curators, may be required to find images on behalf of clients rather than for themselves. It is impossible to give a full picture here of the uses being made of visual information. The following examples should be interpreted as being merely a snapshot of the situation:

## IV. Applications

**4.1 Crime prevention** The police use visual information to identify people or to record the scenes of crime for evidence; over the course of time, these photographic records become a valuable archive. In the UK, it is common practice to photograph everyone who is arrested and to take their fingerprints. The photograph will be filed with the main record for the person concerned, which in a manual system is a paper-based file. In a computer-based system, the photograph will be digitised and linked to the corresponding textual records. Until convicted, access to photographic information is restricted and, if the accused is acquitted, all photographs and fingerprints are deleted. If convicted, the fingerprints are passed to the National Fingerprint Bureau. Currently, there is a national initiative investigating a networked Automated Fingerprint Recognition system involving BT and over thirty regional police forces. Other uses of images in law enforcement include face recognition [e.g. Wu and Narasimhalu,, DNA matching, shoe sole impressions [e.g. Ashley, 1996], and surveillance systems. The Metropolitan Police Force in London is involved with a project which is setting up an international database of the images of stolen objects

**4.2 Medicine:** The medical and related health professions use and store visual information in the form of X-rays, ultrasound or other scanned images, for diagnosis and monitoring purposes. There are strict rules on confidentiality of such information. The images are kept with the patients' health records which are, in the main, manual files, stored by unique identifier (NI number). Visual information, provided it is rendered anonymous, may be used for research and teaching purposes. Much of the research effort related to images is undertaken in the medical physics area. Aspects of concern include effective image processing (e.g. boundary/feature detection) systems which aid the practitioner in detecting and diagnosing lesions and tumours and tracking progress/growth.

**4.3 Fashion and graphic design:** Imagery is very important for graphic, fashion and industrial designers. Visualization seems to be part of the creative process. Whilst there will be individual differences in the way designers approach their task, many use images of previous designs in the form of pictures, photographs and graphics, as well as objects and other visual information from the real world, to provide inspiration and to visualise the end product. 2-D sketches, and, increasingly, 3-D geometric models are used to present ideas to clients and other colleagues. There is also a need to represent the way garments hang and flow.

**4.4Publishing and advertising.** Photographs and pictures are used extensively in the publishing industry, to illustrate books and articles in newspapers and magazines. Many national and regional newspaper publishers maintain their own libraries of photographs, or will use those available from the Press Association, Reuters and other agencies. The photographic collections will be indexed and filed under, usually, broad subject headings (e.g. local scenes, buildings or personalities as well as pictures covering national and international themes). Increasingly, electronic methods of storage and access are appearing, alongside developments in automated methods of newspaper production, greatly improving the speed and accuracy of the retrieval process. Advertisements and advertising campaigns rely heavily on still and moving imagery to promote the products or services. The growth of commercial stock photograph libraries, such as Getty Images and Corbis, reflects the lucrative nature of the industry.

**4.5 Architectural and engineering design.** Photographs are used in architecture to record finished projects, including interior and exterior shots of buildings as well particular features of the design. Traditionally these photographs will be stored as hardcopy or in slide format, accessible by, say, project number and perhaps name, and used for reference by the architects in making presentations to clients and for teaching purposes. Larger architects' practices with more ample resources, have introduced digital cameras and the electronic storage of photographs.

The images used in most branches of engineering include drawings, plans, machine parts, and so on. Computer Aided Design (CAD) is used extensively in the design process. A prime need in many applications is the need to make effective use of standard parts, in order to maintain competitive pricing [Bradley et al, 1994]. Hence many engineering firms maintain extensive design archives. CAD and 2-D modelling are also extensively used in architectural design, with 3-D modelling and other visualization techniques increasingly being used for communicating with clients (see, for example, [Ross, 1998] and [Evans, 1996]). A recent survey of IT in architectural firms [Fallon, 1998] emphasized the dominance of CAD (especially 2-D) in the design process, though it concluded that object-based, intelligent 3-D modelling systems will become more important in the future.

**4.6 Historical research.** Historians from a variety of disciplines – art, sociology, medicine, etc. – use visual information sources to support their research activities. Archaeologists also rely heavily on images. In some instances (particularly, but not exclusively, art), the visual record may be the only evidence available. Where access to the original works of art is restricted or impossible, perhaps due to their geographic distance, ownership restrictions or factors to do with their physical condition, researchers have to use surrogates in the form of photographs, slides or other pictures of the objects, which may be collected within a particular library, museum or art gallery. Photographic and slide collections are maintained by a wide range of organisations, including academic and public libraries.

## V.CONCLUSION

The M-tree is an original index/storage structure with the following major innovative properties:

•*I*t is a paged, balanced, and dynamic secondary memory structure able to index data sets from generic metric spaces;

• *S*imilarity range and nearest neighbor queries can be performed and results ranked with respect to a given query object;

• query execution is optimized to reduce both the number of page reads and the number of distance computations;

• it is also suitable for high-dimensional vector data.

## VI.REFERENCES

1. C. B¨ohm, S. Berchtold, and D. Keim. Searching in High-Dimensional Spaces – Index Structures for Improving the Performance of Multimedia Databases. ACM Computing Surveys, 33(3):322–373, 2011.

2. E. Ch´avez, G. Navarro, R. Baeza-Yates, and J. Marroqu´ın. Searching in Metric Spaces. ACM Computing Surveys, 33(3):273–321, 2012.

3. P. Ciaccia, M. Patella, and P. Zezula. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In Proceedings of the 23rd Athens Intern. Conf. on VLDB, pages 426–435. Morgan Kaufmann, 1997.

4. L. Mico, J. Oncina, and E. Vidal. A new version of the nearest-neighbor approximating and eliminating search (aesa) with linear preprocessing-time and memory requirements. Pattern Recognition Letters, 15:9–17, 1994.

5. M. Patella. Similarity Search in Multimedia Databases. Dipartment to di Elettronica Informatica e Sistemistica, Bologna, 1999.

6. T. Skopal, J. Pokorn´y, M. Kr´atk´y, and V. Sn´aˇsel. Revisiting M-tree Building Principles.

S.Jyothi,
M.Tech,
SRI KRISHNA DEVARAYA ENGINEERING COLLEGE
,JNTU Ananthapuram.

M.Harathi M.Tech,
Assistant Professor, Department of CSE, SRI KRISHNA DEVARAYA ENGINEERING COLLEGE,
Gooty, Ananthapuram(Dt).