Appeared in IEEE Int. Conf. on Multimedia Computing Systems, ICMCS'97, Ottawa, Canada, June 1997

Browsing and Placement of Multiresolution Images on Secondary Storage*

Sunil Prabhakar Divyakant Agrawal Amr El Abbadi
Ambuj Singh Terence R. Smith
Department of Computer Science
University of California
Santa Barbara, CA 93106.

Abstract

Image decomposition techniques such as wavelets *are used to provide multiresolution representations of images. The original image is represented by several coefficients, one of them with visual similarity to the original image, but at a lower resolution. Several strategies are evaluated to store the image coefficients on parallel disks so that thumbnail browsing as well as image reconstruction can be performed efficiently. Disk simulation and experiments with real disks are used to evaluate the performance of these strategies. The results indicate that significant performance improvements can be achieved with as few as four disks by placing image coefficients based upon browsing access patterns.*

1 Introduction

With rapid advances in computer and communication technologies, there is an increasing demand to build and maintain large image repositories. The Alexandria Digital Library project (ADL) [3] is an example of such an application. An image repository should not only provide access to the holdings but also provide the means for users to browse the contents of the library. Browsing is useful in locating items of interest in situations where users are not sure which items match their interests and would like to navigate through a set of items to determine those that are of interest. Since most items viewed during browsing may not be of interest to the user it is not necessary to transmit the complete item, which would consume large amounts of system resources. For browsing purposes, the user is not shown the complete item, rather a summary is presented to enable the user to evaluate the suitability of the item. When items of interest are identified, the user may wish to see these items in greater detail. It is therefore important to provide the ability to browse items and view them at multiple levels of detail or resolution.

For the purposes of ADL, which is designed to handle mostly image data, the technique of wavelet decomposition [1] has been proposed for generating multiresolution images. Using this technique, an image is decomposed into a low resolution copy or icon and several coefficients. The icon is the lowest resolution copy available. Higher resolution copies can be progressively reconstructed from the icon and other coefficients. The original image can be reconstructed from the icon and all the coefficients. In order to improve the I/O performance of browsing, we have investigated efficient placement techniques on magnetic disks [2]. The latency of disk access is reduced by fetching data in parallel from multiple disks. The study focuses on browsing access patterns when images are retrieved by similarity matching, where similarity is based on the content of the images. Two orthogonal declustering schemes have been evaluated. The first scheme declusters the image icons based on their similarity. Since similar images are likely to be accessed together, it would be beneficial to place them on different disks and access them in parallel. The second scheme declusters the coefficients of each image across disks since the coefficients will be accessed together when the image is reconstructed to obtain a higher resolution copy.

2 Experimental Results

The various strategies were extensively tested using simulation as well as experiments with real disks. The results from experiments with real disks were found to be in good agreement with the simulation results. Two groups of placement strategies which use different al-

*Work supported by a research grant from NSF/ARPA-/NASA IRI9411330 and NSF instrumentation grant CDA-9421978.

gorithms to place the thumbnail coefficient of each image on the disks were studied. For the *Distance based* group, the thumbnails were declustered based upon the similarity of the image contents. For the *Round Robin* based group, the thumbnails were placed in a sequential manner with no regard to image content. Under each of these groups, four different declustering strategies for the rest of the image coefficients were evaluated. The *Interleaved* and *Zonal* strategies decluster image coefficients across all disks whereas the *Separated* strategy places the non-thumbnail coefficients together on a designated disk and the *Bundled* strategy places the thumbnail and all other coefficients on the same disk. The difference between the Zonal and Interleaved strategies is that Zonal separates all coefficients of each type into a separate zone on each disk. As a reference, a Random placement strategy which placed all coefficients and thumbnails randomly over all the disks was also considered. Thus nine placement strategies for images were evaluated. The tests were conducted under different levels of user concurrency to study the effect of contention for the resources.

In all our experiments, it was found that the placement strategies in the Distance based group always performed better than the corresponding strategies in the Round Robin group and the Random strategy. From this we conclude that declustering of the image thumbnails based upon their image content is beneficial. The gains of coefficient declustering were, however, not as consistent. The dataset used for the experiment consists of 50,000 landsat images. The experiments were conducted with varying numbers of users. Each user executed 1000 sets of queries. Each set consisted of the request for an image thumbnail chosen at random, followed by the request for a certain number of the most similar thumbnails and finally, the expansion of some of these retrieved thumbnails to a higher resolution. In Figure 1 we present some representative results of tests conducted with real disks. The x-axis gives the number of concurrent users and the y-axis gives the ratio of the average access time for a random policy to the average access time for the other placement strategies. A higher ratio indicates better performance of the strategy as compared to the random strategy. Since the Round Robin based strategies were always outperformed by the Distance based strategies, only the Distance based strategies are shown. The experiments were repeated several times and 95% confidence intervals were generated based upon the results, as shown in Figure 1. From the figure it is clear that when only 1 or 2 concurrent users are present, the strategies that decluster the coefficients perform better than those that do not decluster. As the number of concurrent users increases

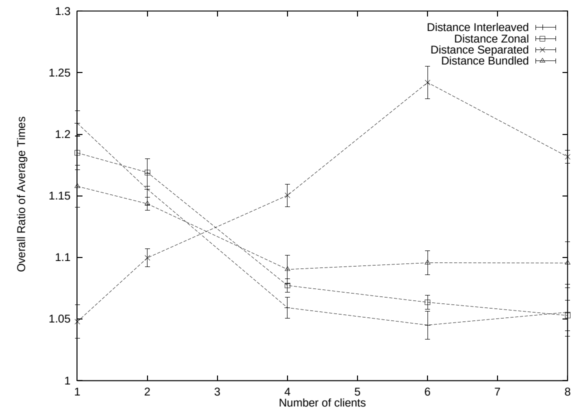

Figure 1: Comparison of declustering strategies

to 3 or more, both policies that decluster the coefficients, Interleaved and Zonal, are outperformed by the policies that do not decluster coefficients. From this we can conclude that coefficient declustering is good in a single user system, but as the number of concurrent users increases, coefficient declustering becomes harmful.

3 Concluding Remarks

It was found that although declustering images based upon their content is beneficial for content based retrieval, the benefits of declustering the image coefficients, however, depend largely upon the degree of user concurrency. The best declustering strategy therefore, depends upon the level of concurrency. For low concurrency levels, Interleaved is the best, for high concurrency Separated and Bundled are better. Significant performance improvements were achieved with as few as four disks. It should be pointed out that these placement strategies could also be applied to other types of data and retrieval patterns.

References

[1] K. R. Castleman. *Digital Image Processing.* Prentice-Hall, Inc., 1996.

[2] S. Prabhakar, D. Agrawal, A. El Abbadi, A. Singh, and T. Smith. Browsing and placement of multiresolution images on secondary storage. Technical Report TRCS96-22, Dept. of Computer Science, Univ. of California, Santa Barbara, 1996.

[3] T. R. Smith and J. Frew. Alexandria digital library. *Communications of the ACM*, 38(4):61–62, April 1995.