Image Browsing using Hierarchical Clustering

Santhana Krishnamachari
Mohamed Abdel-Mottaleb
Philips Research
345 Scarborough Road
Briarcliff Manor, NY 10510, USA
{sgk,msa}@philabs.research.philips.com

Abstract

Digital images and video clips are becoming popular due to the increase in the availability of consumer devices that capture digital images and video clips. Digital content is also growing over the Internet. The increase of the digital content creates a need for user- friendly tools to browse through large volumes of digital material. In this paper we present a clustering based browsing algorithm. Images are automatically clustered using a hierarchical clustering algorithm and users can then browse through the images by navigating the tree structure that results from the clustering. We have tested the algorithm on a large number of images.

1. Introduction

Content-based image/video browsing and retrieval are becoming important research topics in recent years. Research interest in this field has escalated because of the proliferation of video and image data in digital form. The growing popularity of the Internet, the introduction of new consumer products for digital image and video creation, and the emergence of digital standards for television broadcasting have resulted in a greater demand for efficient storage and retrieval of multimedia data.

Content-based retrieval systems for images [1-5,8] based on various image features have been presented. However, image browsing has not received as much attention as image retrieval from the research community. Most of the above mentioned systems require the user to perform search by example, by choosing one of the existing images in the database as a query image or by requiring the user to compose an image. Often users can not present to the system good examples of what they are looking for without effectively browsing through the database. Hence an efficient browsing tool is essential to introduce the user to the contents of the database. After identifying an image of interest through browsing, the user can perform content-based search to retrieve similar images from the database.

For databases with large numbers of images, it is not feasible to browse linearly through the images in the database. A desirable characteristic of a browsing system is to let the user navigate through the database in a structured manner. In [9] a content management systems for home video libraries, which automatically segments the video clips and extracts a visual table of content, was presented. The user can browse the video material through the table of content. In [10], a system was presented for parsing and browsing the video through the extracted key-frames.

In this paper we present a browsing approach where users can navigate non-linearly through the images in the database. The approach is to automatically create a hierarchical clustering of the images in the database and use the resulting tree structure to navigate through the images. This approach is efficient and does not burden the user to go through all the images in the database. This approach can be applied to key-frames extracted from video segments to enable the users to browse through their video libraries. It should be clear that the usefulness of this approach depends on the robustness of the hierarchical clustering results. If some images are grouped incorrectly, the user may not be able to find them by browsing through the tree. Quantitative evaluation of the clustering for browsing is difficult without subjective tests. We have obtained quantitative results to show the use of hierarchical clustering for efficient content-based retrieval [6,7].

The rest of the paper is organized as follows: Section 2 presents the details of image representation and similarity measure. Section 3 presents the details of the hierarchical clustering algorithm. Section 4 presents the algorithm for selecting the representative images from each cluster. Section 5 presents the use of hierarchical clustering for image browsing and presents some examples. Section 6 presents briefly the quantitative results to show the effectiveness of hierarchical clustering. Finally, Section 7 concludes the paper.

0-7695-0250-4/99 $10.00 © 1999 IEEE

2. Image representation and Similarity Measure

2.1. Image representation

Several histogram-based approaches have been proposed for image representation by color. These approaches are based on using a single color histogram for image representation. The difference between these approaches is mainly in their choice of the color space and color quantization. Since these approaches use a single histogram to calculate similarities, the results are expected to reflect only the global similarity. In this paper, we use the scheme we used in [2], which allows retrieval based on local color features. Images in the database are divided into rectangular regions. Then every image is represented by the set of normalized histograms corresponding to these rectangular regions. It should be noted here that the choice of the rectangular region size is important. In one extreme, the whole image is considered as a single region, in this case the image representation reflects the global color information. As the size of the region becomes smaller, the local variations of color are captured by the local histograms. The size of the region should be small enough to emphasize the local color and large enough to offer a statistically valid histogram. In the experiments, images were partitioned into 16 rectangular regions.

2.2. Similarity measures

To enable non-linear browsing, similar images are automatically grouped together using an appropriate similarity measure. The similarity between two images is measured by calculating the similarity between the histograms of the corresponding rectangular regions. Then a single measure of similarity between the two images is calculated by adding the individual similarities between the corresponding regions. We have used the histogram intersection measure to compare the individual normalized histograms (probabilities). Given two normalized histograms, $p = \{p_1, p_2, ..., p_m\}$, $q = \{q_1, q_2, ..., q_m\}$, the similarity measure is defined by

$$s_{p,q} = \sum_i \min(p_i, q_i)$$

3. Hierarchical Clustering

Let n be the number of images in the database, the similarity between all pairs of images is pre-computed. The hierarchical clustering [11] is performed as follows:

1. The n images in the database are placed in n distinct clusters. These clusters are indexed by $\{C_1, C_2, ..., C_n\}$. For the kth cluster, the set E_k contains all the images contained in that cluster and N_k denote the number of images in the cluster. $E_k = \{k\}$ and $N_k = 1$ for $k=1,2,...,n$.
2. Two clusters C_k and C_l are picked such that the similarity measure $S_{k,l}$ is the largest. (The similarity measure between two clusters is defined in the following subsection). These two clusters are merged into a new cluster C_{n+1}. This reduces the total number of unmerged clusters by one. E_{n+1} is updated to $E_{n+1}=\{E_k \cup E_l\}$ and N_{n+1} is updated to $N_{n+1} = N_k+N_l$. Out of the two children of C_{n+1}, one is referred to as the right child $RC_{n+1} = k$ and the other as the left child $LC_{n+1} = l$. The similarity measures between the new cluster C_{n+1} and all the other unmerged clusters are computed as discussed below.
3. Steps 1 and 2 are repeated until only one cluster is left out.

Figure 1 shows a simple example of hierarchical clustering with 8 images.. For this example, N_{14}=5, N_{12}=3, E_{14}={1,2,3,4,5} and E_{12}={6,7,8}. Also, RC_{14}=5 and LC_{14}=13. The tree structure obtained from hierarchical clustering can be effectively used for browsing. The clustering algorithm presented here does not directly depend on the nature of similarity measure. In [6], we have presented a performance evaluation of the clustering algorithm for different similarity measures.

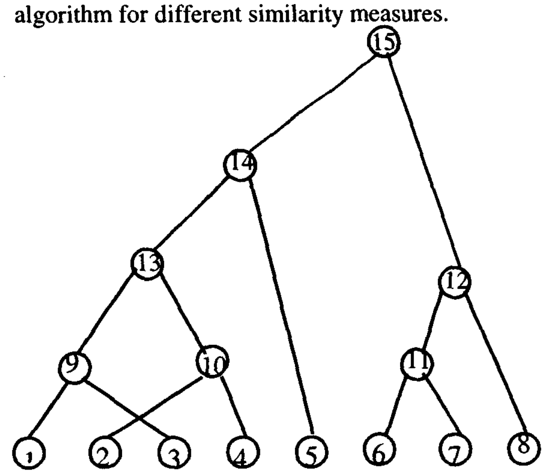

Figure 1: Cluster merging process in hierarchical clustering.

The similarity measure between two images is defined in the previous section. The measure of similarity, $S_{k,l}$, between two clusters, C_k and C_l, is defined in terms of the similarity measures among the images that are contained in those clusters as follows:

$$S_{k,l} = \frac{\sum_{i,j \in \{E_l \cup E_k\}_{i \neq j}} s_{i,j}}{P_{(N_k+N_l)}} \tag{1}$$

where $s_{i,j}$ is the similarity measure between images i and j, E_k is the set of images present in the cluster C_k and N_k is the number of images in cluster C_k. P_n is the number of pairs of images in a cluster with n images: $P_n = (n-1)n/2$

In other words, $S_{k,l}$ is defined to be the average similarity between all pairs of images that will be present in the cluster obtained by merging C_k and C_l. Since the similarity between clusters is defined in terms of the similarity measures between the images in the clusters, there is no need to compute the cluster centers every time two clusters are merged.

When two clusters, C_k and C_l, are merged to form a new cluster, C_m, then it is necessary to compute the similarity between this cluster and all other clusters as given in Eq. (1). This computation is cumbersome as shown below. For any cluster C_t, the similarity measure between C_m and C_t is:

$$S_{m,t} = \frac{\sum_{i,j \in \{E_l \cup E_k\}, i \neq j} s_{i,j} + \sum_{i,j \in E_t, i \neq j} s_{i,j} + \sum_{i \in E_l \cup E_k, j \in E_t} s_{i,j}}{P_{(N_l+N_k+N_t)}} \tag{2}$$

and $S_{m,m}$ is set equal to $S_{k,l}$.

A simple recursive method to achieve the same can be obtained using the fact that the first term in Eq. (2) is equal to $P_{(Nl+Nk)}S_{l,k}$, the second term is equal to $P_{Nt}\, S_{t,t}$, and the third term is equal to $P_{(Nl+Nt)}\, S_{l,t}+P_{(Nk+Nt)}S_{k,t}-P_{Nl}S_{l,l}-P_{Nk}S_{k,k}-2P_{Nt}S_{t,t}$. Thus the similarity $S_{m,t}$ can be obtained from, $S_{l,k}$, $S_{l,t}$, $S_{k,t}$, $S_{t,t}$, $S_{l,l}$, and $S_{k,k}$. The following equation requires far less computation compared to the one above.

$$S_{m,t} = \frac{[P_{(N_l+N_k)}S_{l,k} + P_{(N_l+N_t)}S_{l,t} + P_{(N_k+N_t)}S_{k,t} - P_{N_l}S_{l,l} - P_{N_k}S_{k,k} - P_{N_t}S_{t,t}]}{P_{(N_l+N_k+N_t)}} \tag{3}$$

In Eq. (2), $S_{m,t}$ is computed by summing up the similarity measures of all pairs of images in C_m and C_t, and hence the computation grows as the square of number of images present in the two clusters. The computation in Eq. (3) is independent of the number of images in the clusters. At the beginning of clustering, for all the clusters $S_{i,j}$ is set equal to $s_{i,j}$ and $S_{i,i}$ is set equal to zero.

4. Selection of Representative Images

After clustering, a set of representative images is selected from each cluster. These representative images are used for navigation of the database. The tree structure that is obtained as a by-product of the clustering algorithm can be effectively used to select the representative set of images. In Figure 1, let us consider selecting R representative images for cluster C_{14}. From the tree structure it can be inferred that the images 1 and 3 belong to cluster 9 and images 2 and 4 belong to cluster 10. Hence a good selection of representative images, for R=3, is to select one from {1,3}, another from {2,4} and 5. If R=2, then it is apt to select one from {1,2,3,4} and 5 as representatives. Similarly for C_{12}, it is better to select 6 and 8 or 7 and 8 instead of 6 and 7. Such a selection would result in a representative set that captures the diversity of images present in the cluster.

Let C_i be a cluster for which a set of representative images is to be selected. A set of R nodes is chosen from the tree associated with C_i and from each of these nodes a representative image is selected, resulting in R representative images. The following steps explain this procedure:

1. Set n=0 and form a set R_n={i}. If R=1, then go to Step 5.
2. Each element in R_n is an index of one of the clusters in the sub-tree associated with C_i. Find an element k in R_n such that N_k is the largest.
3. Form a new set R_{n+1} by copying all the elements of R_n except k and by adding RC_k and LC_k.
4. Repeat steps 2 and 3 until the number of elements contained in R_n is equal R.
5. Now R_n contains R nodes from the tree associated with C_i. From each of these nodes a representative image is chosen. If k is an element of R_n , and N_k=1 then the selection is straightforward, i.e., the image associated with the node is selected. If N_k>1, then it is necessary to select a single image representative from E_k. This is done by selecting the one that has the maximum average similarity measure with the other N_k-1 images in E_k.

For the example shown in Figure 1, finding a representative set of images for C_{15} with R = 3, begins with R_0 = {15}, R_1 = {14, 12}, and since $N_{14} > N_{12}$, R_3 = {13,5,12}. The iteration stops here as R_3 contains three elements already. Now, since C_5 has a single element, image 5, is chosen as a representative image. Another representative is selected from C_{13} that contains four images {1,2,3,4} by calculating the average similarity of each image with the other three images and then choosing the one with the maximum average similarity. Assuming that image 2 has the largest average similarity, it is selected as the representative image. Similarly another representative image is selected from C_{12}.

5. Image Browsing

We have developed an image browsing system accessible through the Internet. Two screen shots of our system are presented later. The hierarchical clustering was performed on a database of 3856 images, 200 of these images are taken from two collections of COREL Professional Photo CD-ROMs, the Sampler II - Series 400000 and the Sampler - Series 200000. The rest of the images were obtained from the Department of Water Resources, California. The images are of widely varying colors and scene content.

The hierarchical clustering was performed as discussed in Section 3 by recursively merging the clusters. For each cluster (node in the tree) twenty-five representative images were selected. The tree structure and the set of representative images for each cluster is stored and then used for browsing.

From the nature of the hierarchical clustering algorithm, the clusters at the lower levels of the tree contain predominantly similar images. Whereas the clusters at the higher levels of the tree contain more dissimilar images reflecting the diversity of the image content in the database. Thus, when the user begins to browse at the top level of the tree, many dissimilar images are displayed to the user. But as the user traverses down the tree, the displayed images are more similar to the image of the user's interest.

At the top level of browsing, the user first sees the representative images corresponding to the root node of the hierarchical clustering tree. Figure 3 shows the screen shot of the representative images in the top level. Each of these images was selected from the clusters in the R_n set as discussed in section 4. The number below each representative image shows the number of images present in that cluster. The user can navigate through the database by selecting one of the displayed images that is similar to what he/she is looking for. Once the user makes the selection, the cluster containing the image selected by the user is identified and the representative images of that cluster are displayed. Figure 4 shows the screen shot when the user selects the image in the fourth row and the second column in Figure 3. Here it can be seen that the images are predominantly blue-sky images. Thus the user navigates down the tree in browsing for the image of interest. It is also possible to navigate upwards through the tree by selecting the up-arrow as shown in Figure 4.

6. Quantitative Evaluation

The hierarchical clustering achieves the same goal for browsing and retrieval, namely to avoid looking at all the images in the database in case of browsing and to avoid comparing a query with all the images in case of retrieval. This is achieved by grouping similar images in the same cluster. In case of retrieval we can quantitatively evaluate the effectiveness of clustering by comparing the results of retrieval with clustering against the results obtained by comparing the query exhaustively with every image in the database. These quantitative results can give us a good indication of the effectiveness of clustering for browsing, because while browsing a user selects an image and the system selects (or retrieves) the representative images from the cluster that contains the selected image.

We clustered all the images in the database using hierarchical clustering as discussed in Section 3. Instead of repeating Steps 1 and 2 in Section 3 until only one cluster is left out, we stop the clustering when the average number of images per cluster reaches a threshold. In our experiments the average number of images per cluster was chosen to be 22 images. The smallest cluster contained 5 images and the largest cluster had 80 images. For each of these clusters, the set of representative images is obtained as discussed in Section 4. The histograms of the representative images are averaged to obtain a cluster-histogram for each of the clusters.

To retrieve similar images for a given query, the query image histogram is initially compared with all the cluster-histograms. Then a subset of clusters that have the largest similarity to the query image is chosen and all the images in these clusters are compared with the query image. The images that exhibit the largest similarity measure to the query are then displayed to the user. This comparison scheme obviates the need to compare the query image exhaustively with each individual image in the database.

To quantitatively evaluate the use of clustering for retrieval, we chose a set of 300 query images (that are not present in the database) and for each of the query image, we performed an exhaustive search and retrieved the top N similar images. Then, for each of query we performed a clustering-based search and obtained the top N similar images. The ratio of the number of common images in clustering based search and the exhaustive search to N is defined as the retrieval accuracy. The retrieval accuracy is then averaged over all query images. Figure 2 shows the plot of the average retrieval accuracy against the average number of the similarity comparisons required. The plot consists of 8 points, obtained by examining the top 3,4,7,10,13,19,25, and 31 clusters. The leftmost point corresponds to the result obtained by examining 3 clusters and the rightmost point corresponds to the result obtained by examining 31 clusters. From the figure it can be seen that for N=10, *i.e.*, the user is interested in the top 10 retrieved images, 90% retrieval accuracy can be obtained with 540 similarity comparisons as opposed to 3856 comparisons that would be required for exhaustive search.

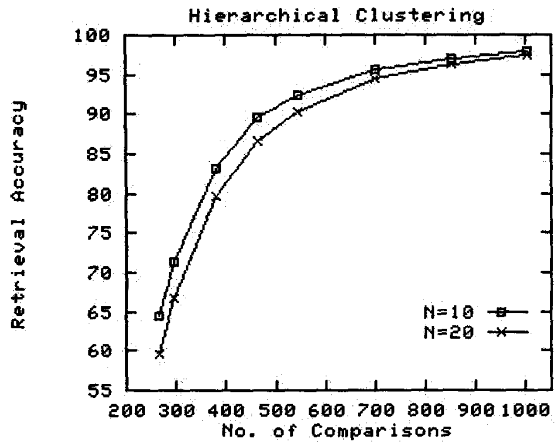

Figure 2: Retrieval accuracy with hierarchical clustering

Our experimental evaluation shows that this clustering based indexing technique offers high retrieval accuracy with a considerable reduction in the number of images being examined. For detailed results the reader is referred to [6,7].

7. Summary and Conclusions

In this paper we have presented an algorithm for nonlinear image browsing based on image clustering, where images are represented by local color histograms and are grouped using the hierarchical clustering technique. The hierarchical clustering produces a tree structure. For each node in this tree a set of representative images is selected. The representative images for the nodes at the higher levels of the tree show the diversity of image content in the database thus enabling the user to view the variety of images. The representative image set at the lower levels contains very similar images. The hierarchical tree structure can thus be efficiently used to navigate through large volumes of digital images.

7. References

[1] M. J. Swain and D. H. Ballard, "Color Indexing", *Intl. J. of Computer Vision*, 7(1), pp. 11-32, 1991.

[2] M. Abdel-Mottaleb, N. Dimitrova, R. Desai, and J. Martino, "CONIVAS: CONtent-based image and video access system", *Proc. of ACM Intl. Multimedia Conference*, Nov. 1996.

[3] W. Niblack, et.al., "The QBIC Project: querying images by content using color, texture and shape." *In Storage and Retrieval for Image and Video Databases I*, Vol. 1908, *SPIE Proceedings*, Feb. 1993.

[4] J. Smith and S.-F. Chang, "A Fully Automated Content-based Image Query System", *Proc. of ACM Intl. Multimedia Conference*, Nov. 1996.

[5] H. Zhang, Y. Gong, C. Y. Low and S. W. Smoliar, "Image retrieval based on color features: an evaluation study", *Proc of SPIE*, Vol. 2606, pp. 212-220, 1995.

[6] M. Abdel-Mottaleb, S. Krishnamachari, and N.J. Mankovich, "Performance Evaluation of Clustering Algorithms for Scalable Image Retrieval", *In IEEE Computer Society Workshop on Empirical Evaluation of Computer Vision Algorithms*, Santa Barbara, 1998.

[7] S. Krishnamachari and M. Abdel-Mottaleb, "A Scalable Algorithm for Image Retrieval by Color", *In Proc. of Intl. Conf. on Image Processing*, 1998.

[8] W.Y. Ma and B.S. Manjunath, "NeTra: A Toolbox for Navigating Large Image Databases", pp 568-572, *Proc. of Intl. Conf. on Image Processing*, Vol. 1, 1997.

[9] N. Dimitrova, T. McGee, H. Elenbaas, and J. Martino, "Video Content Management in Consumer Devices", To appear in *IEEE Transactions on Knowledge and Data Engineering*, Nov. 1998.

[10] H.J. Zhang, Y.L. Chien, and S.W. Smoliar, "Video Parsing and Browsing Using Compressed Data," *Proc. Multimedia Tools and Applications*, Vol. 1, No. 1, pp. 89-111, 1995.

[11] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, New Jersey, 1988.

Acknowledgements

We wish to thank Dave Kearney of the Department of Water Resources, California and Ginger Ogle of UC Berkeley for their assistance in providing us with the image database.

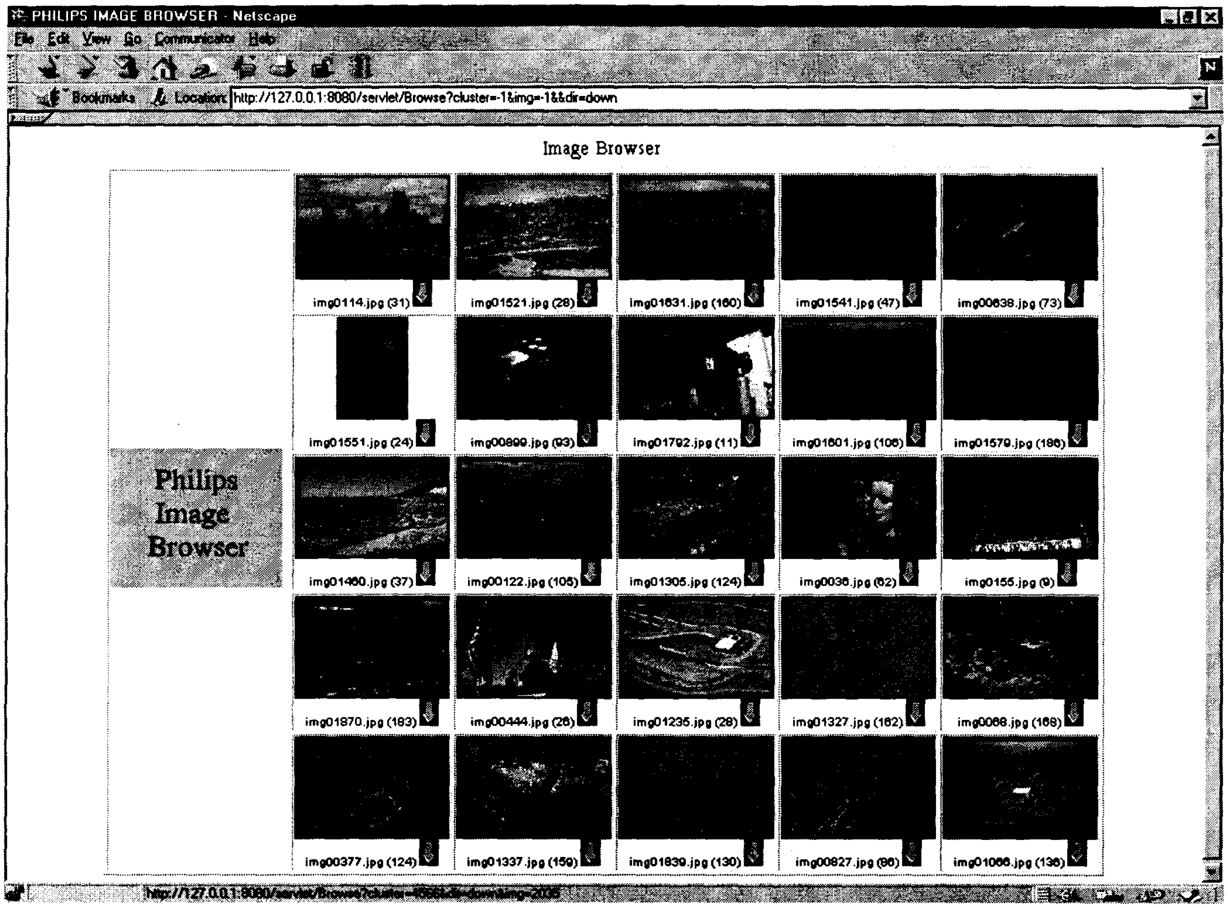

Figure 3: Top level of the image browser. Representative images of the root node are displayed. The number below each image represents the number of images present in that cluster.

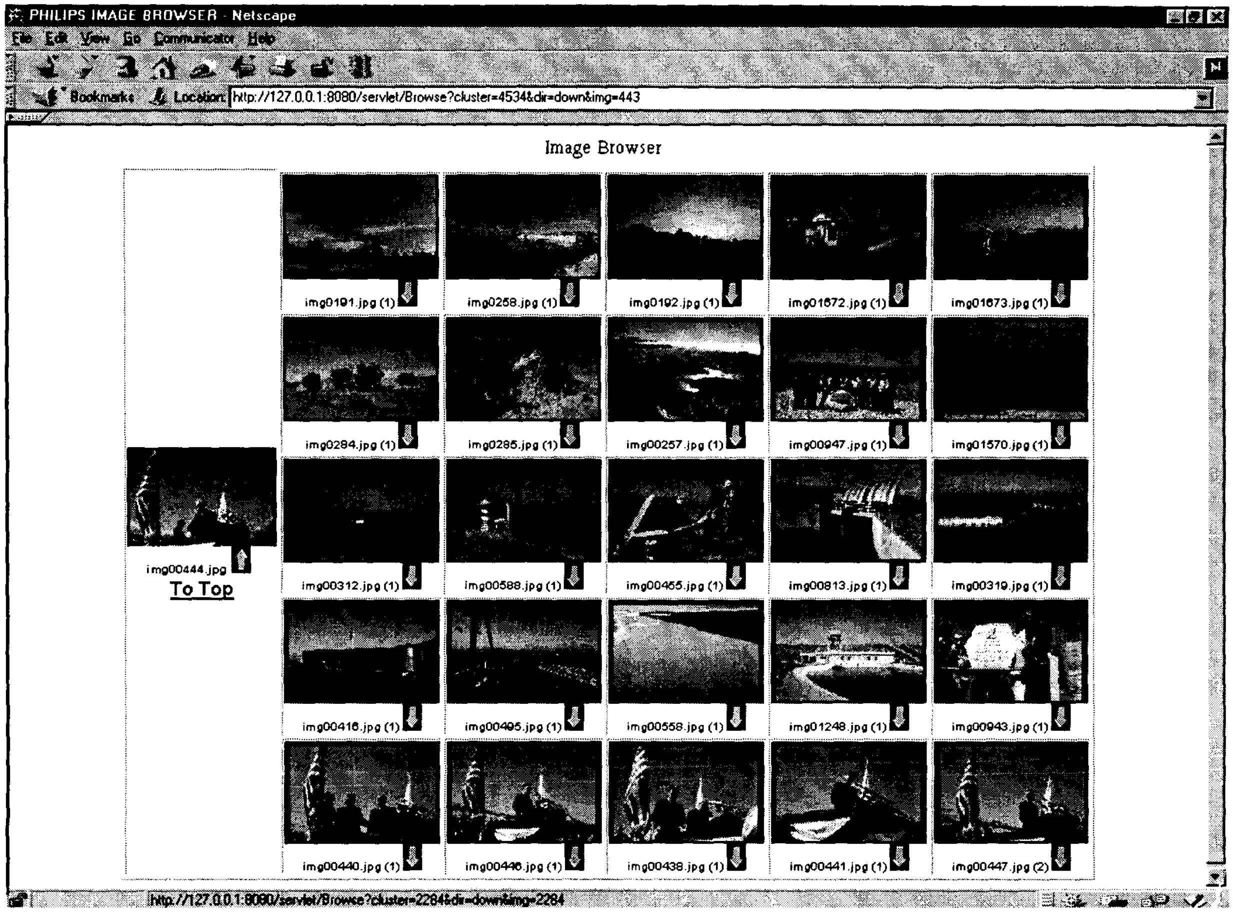

Figure 4: The set of images displayed after the user selected an image from the top-level representative images.