Retrieval of multi- and hyperspectral images using an interactive relevance feedback form of content-based image retrieval

Irwin E. Alber[a], Morton Farber[a], Nancy Yeager[b]
Ziyou Xiong[c], and William M. Pottenger[c,d]

[a]Boeing Integrated Defense Systems, Seal Beach, CA, 90740
[b]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign
[c]Beckman Institute, University of Illinois at Urbana-Champaign, [d]Lehigh University[1]

ABSTRACT

This paper demonstrates the capability of a set of image search algorithms and display tools to search large databases for multi- and hyperspectral image cubes most closely matching a particular query cube. An interactive search and analysis tool is presented and tested based on a relevance feedback approach that uses the "human-in-the-loop" to enhance a content-based image retrieval process to rapidly find the desired set of image cubes.

Keywords: Content-based image retrieval, relevance feedback, multi-spectral images, hyperspectral images, image data mining

1. INTRODUCTION

A significant number of government and industrial organizations are engaged in collecting and analyzing a wide range of remote sensing images for a variety of commercial and military applications. Space and earth-based sensor platforms produce an enormous variety of image data. The size of the image files is particularly large for the case of multi- and hyperspectral collections. These collections can easily generate terabyte-size files of spatial/spectral image cubes.

For remote sensing scientists and image analysts to interrogate these large image files, it has been our experience that a very efficient set of image mining algorithmic tools is needed. These tools must be able to perform, quickly and efficiently, a series of interactive search, retrieval, visual display, and analysis tasks.

In this paper we sketch our development of an image-mining algorithm and visual display tool that searches the database for stored image cubes that most closely match a particular query image. This interactive search and analysis tool incorporates the relevance feedback approach developed at the University of Illinois to improve the performance of a content-based image retrieval process[1,2,3]. The relevance feedback methodology uses the "human-in-the-loop" to aid in the process of retrieving hard-to-define multi-spectral image objects.

In section 2, we briefly review the underlying motivation and approach of the relevance-feedback-based approach and its coupling to the content based image retrieval (CBIR) portion of the algorithm. Key to the search process is the designation of a prescribed set of feature vectors for each image that are pre-calculated and stored off-line as a set of search indices. Spectral and texture feature vectors attempt to express in a compact low-level digital form, some of the higher level concepts in a given image scene (e.g. a fire and smoke event against a rain-forest background).

In section 3, we outline the overall algorithm search process. The image mining process is illustrated with examples from two sets of multi-spectral images: (1) GOES weather-satellite data[4] (5 bands of space-time images of hurricane events; 0.65, 3.9,6.7,10.7,and 12.0 microns), and (2) MODIS airborne simulator (MAS) images of fires from NASA's SCAR-B Brazilian rain forest campaign (50 bands; 0.55-14.2 microns)[5,6].

In section 4, quantitative measures are presented of relevance feedback effectiveness, referenced to a single-step CBIR search. Preliminary performance results are presented for our two multi-spectral image search experiments.

[1] William M. Pottenger gratefully acknowledges the help of his Lord and Savior, Jesus Christ, in this work.

Finally, we summarize our work and discuss future algorithm developments in section 5. Here, we indicate how the speed of the retrieval process can be considerably increased using a “smart” triangle inequality and triangle-trie algorithm[7,8,15].

2. CBIR AND RELEVANCE FEEDBACK

2.1 CBIR

Content-based image retrieval (CBIR) systems use low level feature vectors, e.g. measures of color, texture, and shape, to express the visual “content” of images[1]. Once prescribed, these features are calculated from the set of images to be searched, and then stored in a reference file. The “art”, and the inherent difficulty, in constructing an effective CBIR code is to try to devise a set of feature vectors that most succinctly represent the higher level features that are important to the user.

Given a query image, and its corresponding feature vector representation, the remainder of the CBIR procedure then becomes a search in a pre-designated feature vector space, to find a set of images whose feature vectors most closely resemble those of the query image.

A sample query might be the following. “Find me all the images in the database that display an orange sunset similar to that of this particular favorite picture of a sunset.”

The measure of similarity between the retrieved images and the query image is represented mathematically by a similarity or distance metric, e.g., Euclidean distance. The task of the search algorithm is to find those images with the smallest distance to the query.

2.2 Relevance Feedback Concepts (RF)

The dilemma faced by CBIR users relates to the imprecise nature of human perception. Many different users may interpret the same sunset image in many different ways. Stated in another way, different users will weight the feature vectors (e.g., color or texture), and particular components of these features (e.g., the orange part of the sunset), differently.

In their defining relevance feedback paper Rui, Huang, et al.[1], proposed an *interactive* retrieval approach to image data mining. This approach attempted: (1) to fill the gap between low level features and high level visual concepts, and (2) to use the subjectivity of the human perception of images, against a limited number of retrieved images (typically 20), to improve the search process. In this RF approach, dynamically updated feature and feature weights, based on user feedback in the retrieval process, capture the user’s high level query and perception subjectivity. Rui’s 1998 paper[1] employed a heuristic-based approach using color (three components), texture, and shape features based on concepts from earlier text retrieval studies. This relevance feedback formalism was incorporated into a software code called MARS (Multimedia Analysis and Retrieval System) developed at the University of Illinois. The RF code and on-line learning techniques was shown to significantly increase retrieval performance over that of similar CBIR-only retrieval systems.

In a later development of the relevance feedback scheme, Rui and Huang[2], the heuristic-based approach for determining the correct weighting parameters was set aside in favor an approach that attempts to transform the feature space more optimally using feature-space clustering and classification techniques. Please see the cited papers[1,2] for details of the mathematics of weight updating within the RF implementation scheme.

2.3 Feature Vector Selection

In the joint Boeing-NCSA project reported on here, we adopt the basic elements of the latest version of the MARS relevance feedback code to the problem of retrieving multi- and hyperspectral images.

The important decision here is to select a set of feature vectors and display tools tailored to the multi-dimensional properties (x, y, and wavelength) of the image-cube database. The feature vectors initially selected for our study are:

(1) **Texture Vector**: a three component texture measure based on the local-orientation technique of Jahne[9];

(2) **Spectral Vector**: two spectral feature vectors, namely (a) the spectral mean, and (b) the spectral variance. The mean and variance of each band or wavelength is calculated from the ensemble of spectra for each x,y pixel within a local image “tile”. An example of two 50-element index vectors, for a given MODIS/MAS fire scene tile, is shown in Figure 1. Tile sizing is discussed further in section 3.

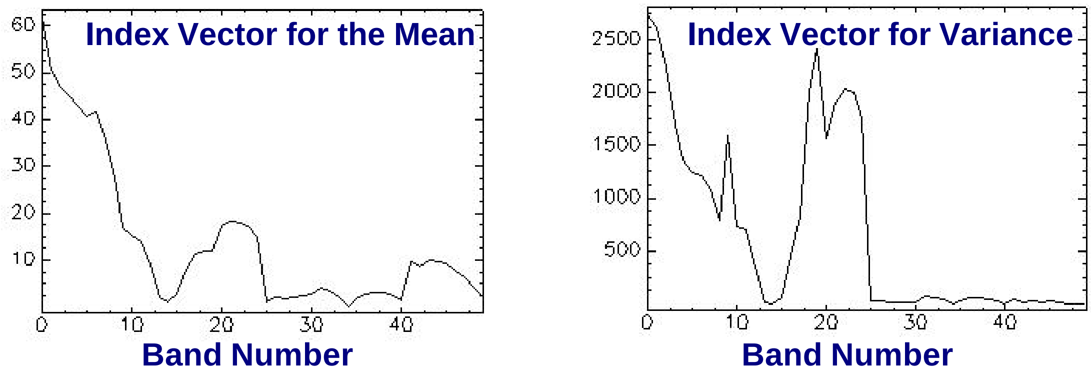


Figure 1. Index structure for two 50-element spectral vectors. MODIS/MAS data. Fire Tile(#79), August 1995 scene.

The number of components in the spectral mean and variance vector is data specific; 5 for the GOES-8 data and 50 for the MODIS/MAS data.

These low-order moments of the spectra are clearly sub-optimal, but they were chosen to provide a first-cut at this most difficult interactive search problem. Sample results are presented and discussed in sections 3 and 4.

3. THE SEARCH AND FEEDBACK PROCESS

3.1 Search and Feedback Outline

Here are the steps employed in the CBIR and RF search process algorithm and code (see interactive process flow in Figure 2). The main part of the code and user interface is written in IDL, the Interactive Data Language [10] .

Step 0: Build the pre-computed feature vector indices and store off-line. Scale the results to 3 different tile sizes.

Step 1: The user builds a query based on time, latitude/longitude, and feature selection. The user selects a specific query image tile.

Step 2: The search engine uses the CBIR module to retrieve the initial search results. The top N images closest in Euclidean distance to the query image are displayed in static or dynamic format.

Step 3. The user ranks the images based on the evaluation of those images that are "closest" to the query image.

Step 4: The RF engine re-computes the query vector. User supplied ranking provides the basis for updating vector weights.

Iteration Loop: Steps 3 and 4 continue until convergence to a set of image cube tiles best matching the initial query and the user input.

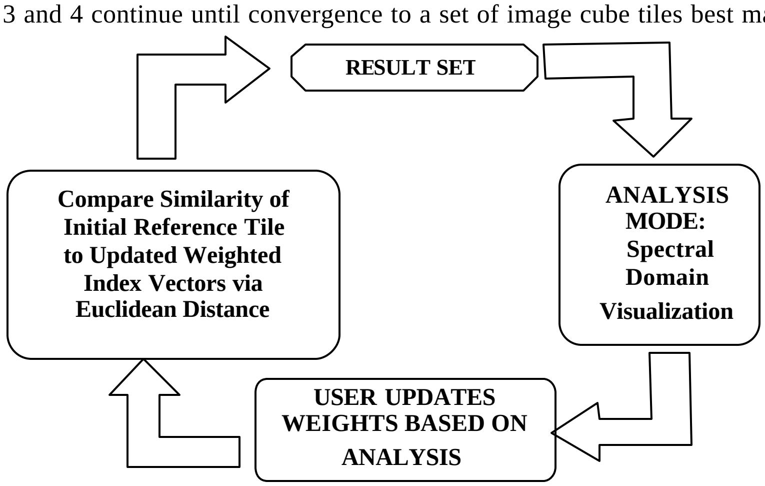


Figure 2. Relevance feedback search process

3.2 Query tile selection

For this image mining study, we chose, as part of our basic search design, to select those query images that emphasize a particular localized feature, in either space or time, viewed in context with the surrounding scene elements. For example, one might choose to select a small region encompassing the main elements of a single fire in a portion of the MODIS/MAS database, or a single hurricane feature in the GOES-8 hurricane season database. To provide the user the flexibility of selecting spatial regions on the scale of the selected event, we set up a tile structure overlaid onto a large image scene, as shown in the upper left inset of Figure 3 for the GOES–8 image. Hurricane Floyd is displayed in the lower right of the image.

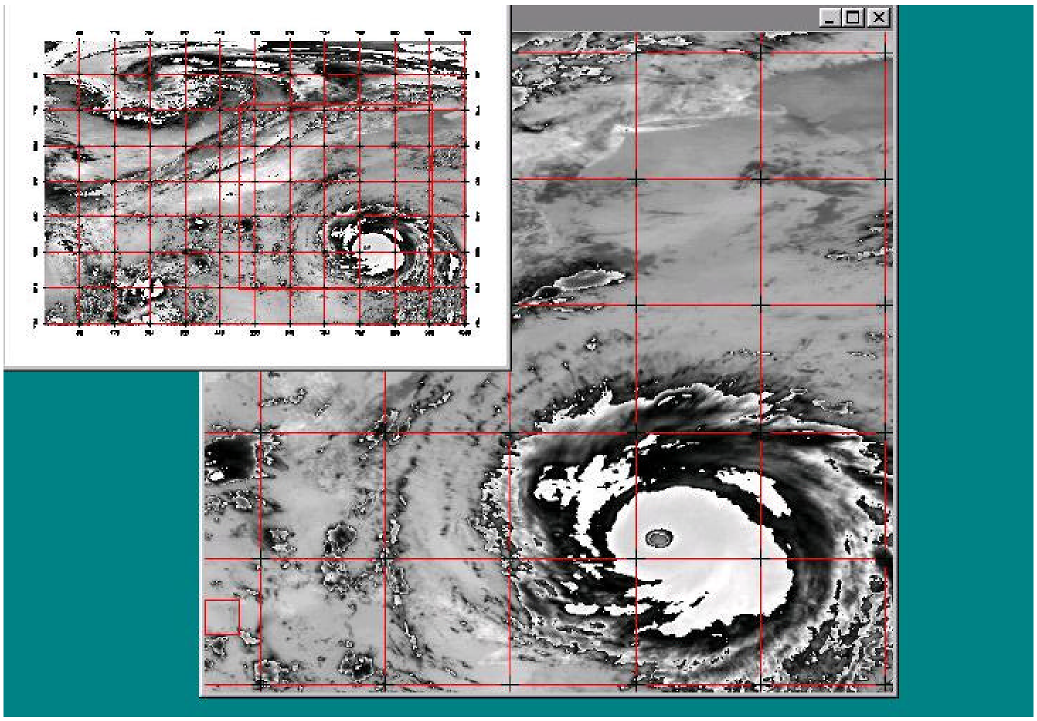


Figure 3. User selects Hurricane Floyd reference tile from GOES-8 image scene. Band 4, at 10.7 microns is displayed.

As noted in section 3.1, we calculate the pre-computed feature vector at 3 different tile sizes: 352x352, 176x176, and 88x88 pixels, respectively. This range of tile sizes gives us the flexibility to select different scales for the object of interest. Clearly, at this level of scaling, we cannot be precise in selecting those features with unusual boundaries. We leave the selection of a more object-based scale or representation to a later development of the algorithm.

The size of the stored index database, i.e., the total number of feature vectors, is a significant factor in the speed of retrieval. For the smallest tile size (88x88) applied to the 30 gigabytes MODIS/MAS database, there will be about 2.4 million feature vectors that must be stored in the index database.

Figure 4 shows the selected fire query tile (176x176) from MODIS/MAS with the thermally sensitive band 30 displayed.

3.3 Content-Based Image Retrieval Procedure

Once the query tile is selected, the search and retrieval algorithm is activated through a client-server GUI created in IDL. Next the Euclidean distances are calculated between the query image and all images in the user-selected database. The nearest 19 images are ranked and displayed along with the query image.

Figure 5 shows the CBIR results ... ensitive infrared band 30 (3.59 microns) displayed. A similar displ... e plume of the fire/smoke query and the associated smoke sensitive ... step.

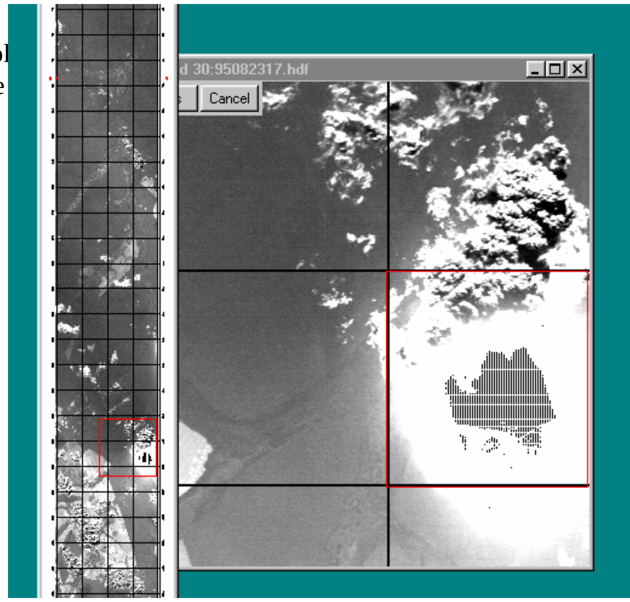


Figure 4. A user selected query tile is selected to initiate the search. Tile #79 from the full MODIS/MAS image strip on the left is shown in blown up form on the right. The 3.59-micron band of the fire event is displayed, MAS band 30.

Reference Query Tile #79 from MODIS/MAS scene 95-08-23-17

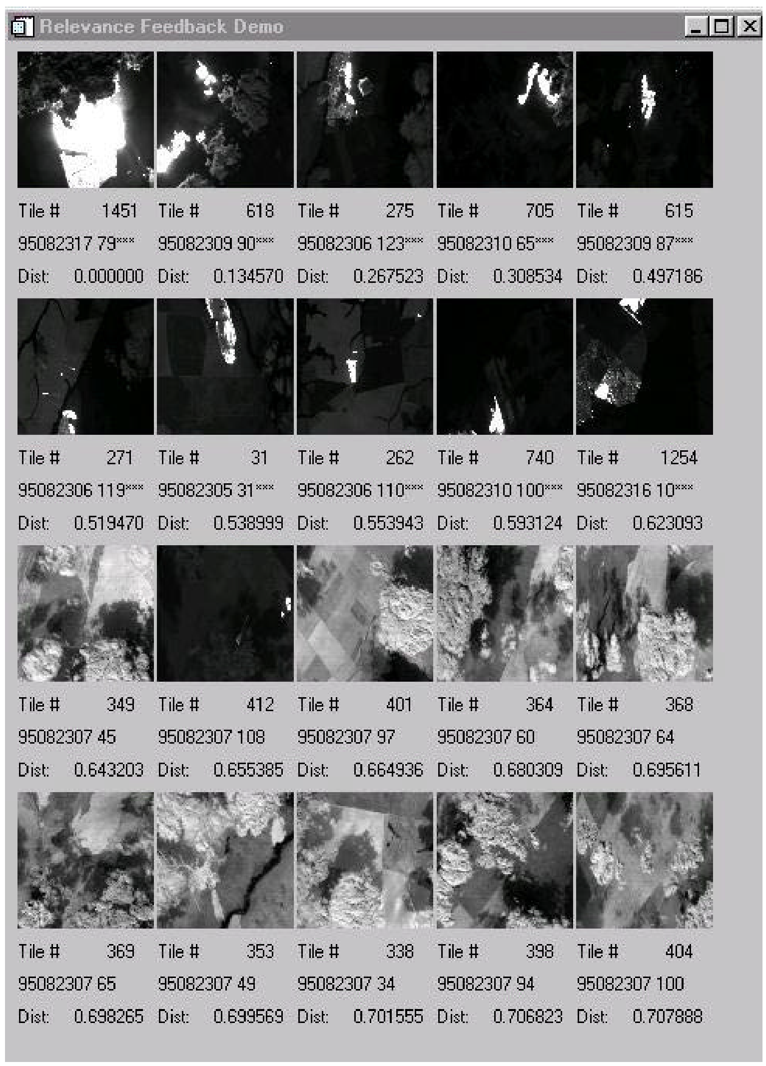


Figure 5. Results of the initial Content-Based Image Retrieval. 19 retrieved tiles + the query tile are shown, band 30.

3.4 Visualization and Analysis

In order for the user to subjectively rank the initially retrieved image cubes as to which of them are "closest" to the query cube, one must be able to individually examine/analyze all of the 19 candidate cubes in an expeditious manner. For the five band GOES-8 data, this can readily be achieved by displaying, on a single screen, thumbnail images of each of the five candidate bands adjacent to the corresponding set of bands for the query cube.

Another technique, widely used in the multi-spectral analysis community, is to use a false color composite based on three representative bands. The key here is to know in advance, using prior experience or a physics-based model, which of the three bands to select.

When the number of bands is of the order of 50 or larger, one may want to comparatively view all the bands of the candidate and query cubes in as short a time as possible. This can be accomplished by displaying the data as an animation, treating the spectral dimensional as a time sequence. Again, this particular technique has been used extensively in various codes that view high dimensional multi- and hyperspectral image data. We implemented, in IDL, a side-by-side spectral "movie" algorithm, named ViewMovie.

We found spectral animation a useful tool for viewing and evaluating the retrieved cubes, for the MODIS/MAS fire problem since we were able to quickly see whether the spectral radiance images transition from a smoke containing scene in the visible, to a hot-zone scene in the infrared. As an example, figure 6 shows two of the bands from the number one candidate cube animation (tile # 618 in Figure 5) compared to those from the reference cube (tile # 1451 at the left in Figure 5).

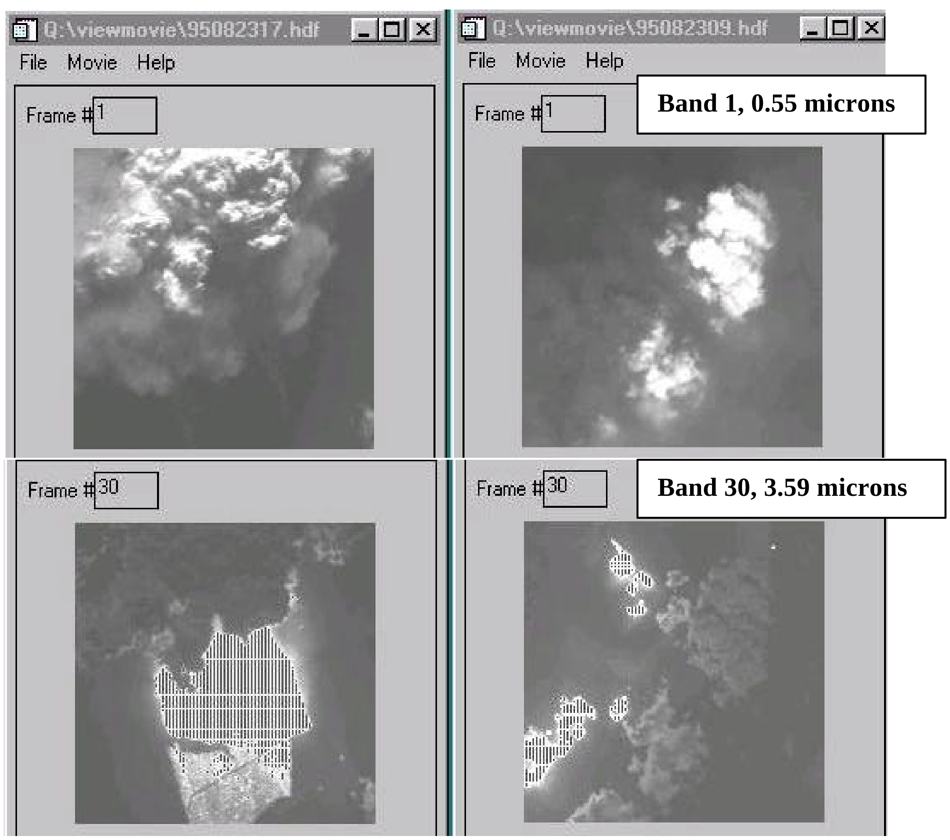


Figure 6. Animation analysis: side-by-side animations through the spectral domain. Two representative bands are displayed for MODIS/MAS data cubes. Candidate cube # 618 (right) is compared to query cube # 1451 (left).

3.5 Relevance Feedback Procedure (RF)

Once the candidate image cubes have been visualized, the user selects a score for each candidate cube. An arbitrary zero to five scale was chosen for our application based on previous RF studies. A score of five indicates that a candidate cube is, in the analyst's opinion, quite relevant or similar to the query cube. A score of zero indicates that the cube is highly non-relevant or quite dissimilar. The relevance feedback algorithm then automatically revises the weights of the query feature vectors based on the input scores. In effect a new vector weighting is selected based on the attributes of a number of image cubes that the user has determined most closely display the characteristics of the object of interest (e.g. a fire in a rainforest).

A new set of images is retrieved, using the new vector weights. Figure 7 displays the final converged set of RF retrieved images for our MODIS/MAS example after two RF iterations. The number of retrieved fire-like image cubes has increased from 10 in Figure 5 (CBIR only) to 15 in Figure 7, after two rounds of relevance feedback. Convergence may take more iterations in other examples, but the number of iterations is typically less than four or five.

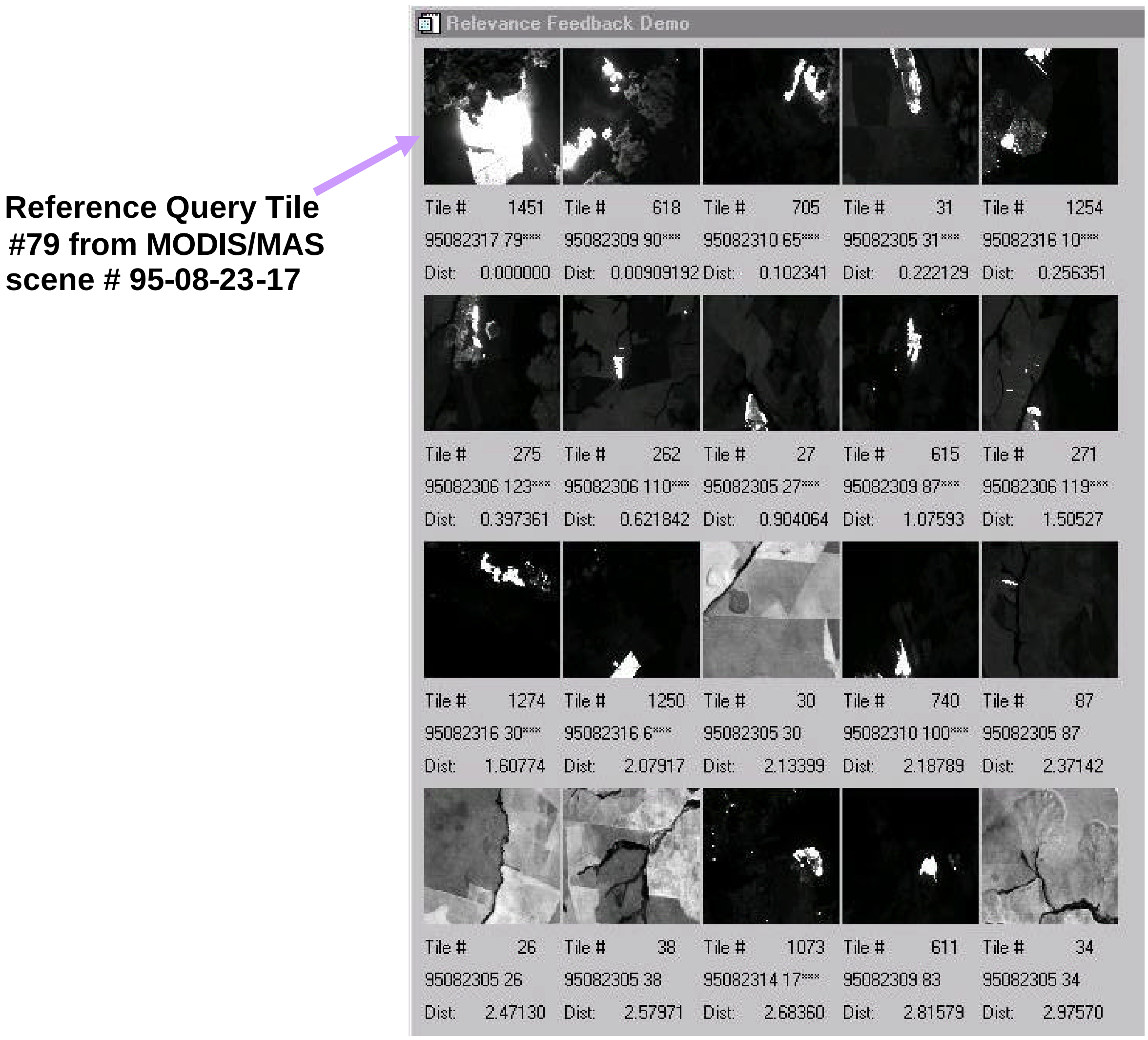


Figure 7. Results of the converged Relevance Feedback process for MODIS/MAS fire query. The query tile + 15 retrieved fire tiles are displayed, Band 30.

4. RELEVANCE FEEDBACK EXPERIMENTAL EFFECTIVENESS

4.1 Performance Measures

The effectiveness or performance of a CBIR or RF scheme can be specified by several objective measures[2]. A standard pair of performance measures is precision (Pr) and recall (Re). They are defined as follows:

$$\mathrm{Pr} = \frac{N(\text{Number of retrieved relevant objects})}{N_T(\text{Number of total retrieved objects})}$$

$$\mathrm{Re} = \frac{N(\text{Number of retrieved relevant objects})}{M_T(\text{Number of total relevant objects})}$$

These two measures are usually plotted against one another to form a parametric precision-recall curve. The parameter governing a point on the curve is the distance threshold d_T, i.e. an image is considered retrieved if its distance, d, from the query image satisfies $d \leq d_T$. If the distance threshold is set quite high, most values of d will satisfy the inequality. Thus, Pr will approach zero (because N_T will be very large). In addition, for this high threshold case, Re will be close to one (because N will approach M_T). In contrast, when the distance threshold is quite low, Re will asymptote to zero.

A somewhat different measure, currently being adopted for image retrieval systems when the recall is consistently low, is the "precision-scope" curve[2]. Scope, or Sc, specifies a fixed number of images returned to the user, e.g. the top 20 images as in our Figures 5 and 7. The revised precision measure as a function of scope, Pr(Sc), is then defined by:

$$\mathrm{Pr}(Sc) = \frac{N(\text{Number of retrieved relevant images})}{Sc(\text{Fixed number of total returned images})}$$

Since in our numerical experiments, Sc remains fixed at 20 (the query image + 19 returned images), then Pr(Sc) is basically a function only of N, the number of retrieved "relevant" images. We use our understanding of the spectral physics of the image features (e.g. what constitutes a fire) or some other objective measure devised by phenomenological experts in the field to define which images are "relevant" and which are not.

4.2 Performance Results

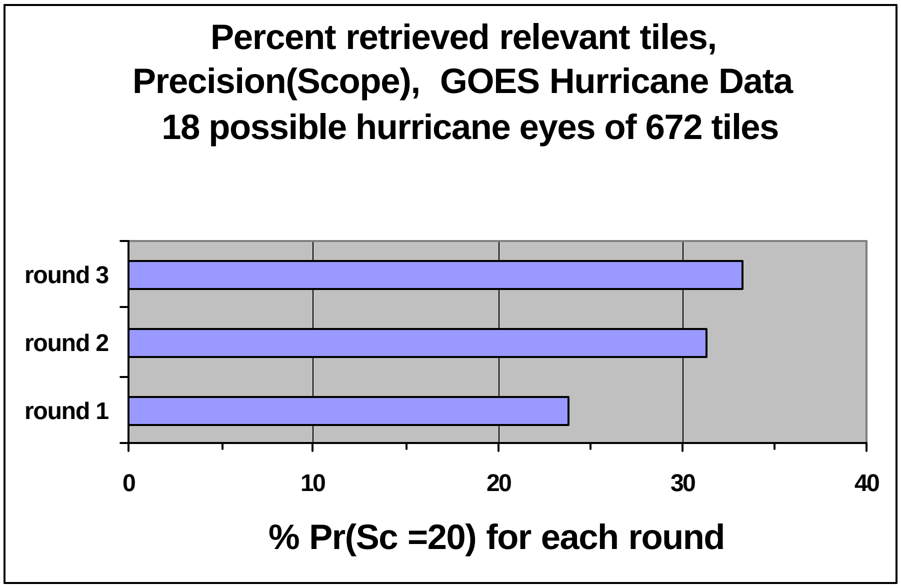
The effectiveness of the relevant feedback procedure was evaluated for two numerical image retrieval experiments, one for each set of data.

4.2.1 GOES -8 Experiment

The 5 band GOES -8 hurricane data was tested using a data set of 672 tiles, containing 18 possible hurricane eyes. For this experiment, the primary two feature vectors selected were: (a) the mean spectrum in a tile, and (b) tile texture based on the local orientation technique. All 5 bands were selected to be included in the CBIR and RF retrieval problem.

Three basic retrieval rounds were evaluated, with round 1 being the CBIR step. The query vector was changed 18 times, using a different hurricane eye query for each retrieval experiment. The ensemble recall results are presented in the bar chart of Figure 8. Note that in round 1, the average number of relevant tiles retrieved in round 1 was 4.77 out of the 20 retrieved tiles shown to the user. This coverts to a precision (percent of relevant retrieved tiles) of 23.9% as shown in Figure 8. An improvement of the precision as a function of the scope metric, Pr(Sc=20), is shown for the user assisted rounds 2 and 3. At the end of the third round the average number of retrieved relevant tiles increased to 6.67 (almost 40% more than in the CBIR round) and the precision level increased to 33.4%.

4.2.2 MODIS/MAS Experiment



The 50 band MODIS/MAS fire data set was tested using the full set of tiles in the MODIS/MAS database, containing 22 distinct fire events. A number of experiments were conducted, some using all 50 bands and some with a limited number of bands, centered on the infrared band at 3.6 microns.

Figure 8. Precision (Scope=20) for 5 band GOES hurricane retrieval experiment; CBIR (round 1) and relevance feedback (rounds 2,3)

The best results were found when the selected spectral mean and variance feature vector was limited to five bands (28-32), with the center band (band 30) centered at 3.6 microns. The texture vector provided little or no gain in performance.

In round 1, the average number of retrieved relevant tiles was 8.0 out of the 20 retrieved tiles shown to the user. This coverts to a precision (percent of relevant retrieved tiles) of 40% as shown in Figure 9. In round 2, the average number of retrieved tiles increased about 10%, to 8.8, with a corresponding precision level of 44.1%. No additional improvement in performance was found for the subsequent rounds.

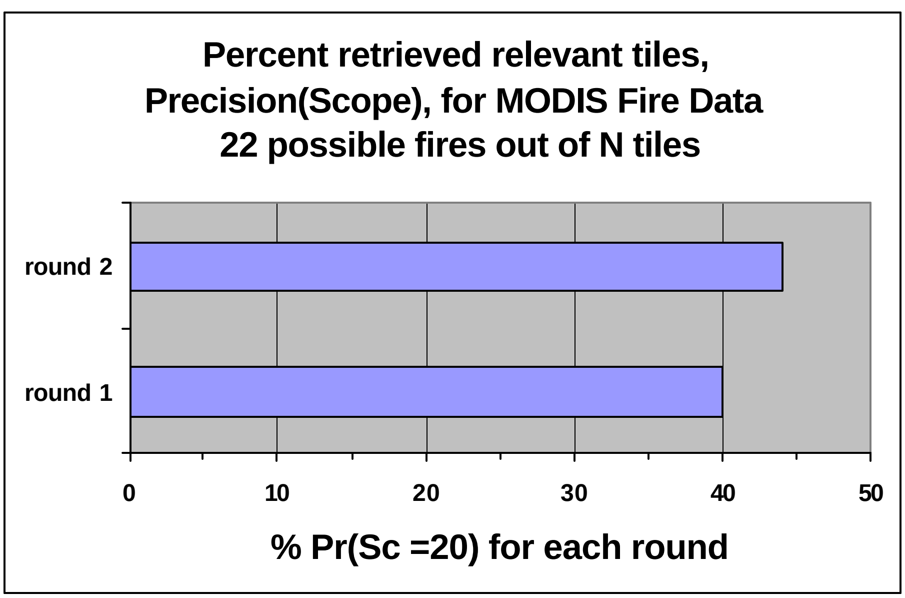


Figure 9. Precision (Scope=20) for MODIS/MAS fire retrieval experiment; CBIR (round 1) and relevance feedback (round 2). Feature vector limited by user to 5 infrared bands centered on 3.6 microns.

An important behavior observed in this MODIS/MAS retrieval experiment was that the biggest improvements in single round 2 testing (the RF round) was found when the round 1 retrieval (the CBIR round) under-performed. For example, when round one retrieved only 2 relevant images, round 2 increased the retrieval to between 4 and 10 images. In contrast, when round one performed very well (10 – 14 images retrieved), round 2 either showed no further increase, or a slight loss. While our experiments were limited, it appears that the RF mode works best in situations where the chosen feature vectors minimally represent the object of interest. In that case the human-in-the-loop provides a significant assist to the retrieval process.

5. SUMMARY AND FUTURE ALGORITHM CHALLENGES

In this paper, we have shown that the analysis of physical events, as seen by multi- and hyperspectral remote sensing imagery, can be significantly aided by the use of a set of image mining algorithms designed to find those image cubes most closely resembling a physical query of interest. Because of the need to adequately express physical objects, such as hurricanes or fires, in terms of low level feature or index vectors, we showed that the content-based image retrieval of images of interest is enhanced through the additional use of the human-in-the-loop relevance feedback procedure. The algorithms and the code results presented here represent but a first step towards developing a robust and efficient search engine for the kind of remote sensing problems that we have examined in this study.

A number of important challenges still must be addressed if this work, or a derivative of this work, is to develop into a set of fully useful tools for remote sensing analysts. The following is a list of the key challenges that must be addressed to further improve our image-mining algorithm.

Improved feature vectors geared to event and background physics. An effective algorithm must be capable of better representing the physical properties of the events of interest. For example, Healey and Jain[11] presented a CBIR search strategy for retrieving a set of multispectral satellite images based on a model of the physical properties of ground materials using radiometric reflectance properties that are invariant to sensor, atmospheric, and illumination characteristics.

Complex object/feature extraction algorithms. In several remote-sensing CBIR studies, the stored feature vectors are based on the properties of pre-calculated material classes, e.g. water, grass, ground material type, etc, or are a direct indexing of the classes themselves. To obtain these complex objects, a significant amount of pre-retrieval computation is required, using state-of- the-art spectral classification and image segmentation codes, such as the MathSoft's GeoBrowse code developed by Marchisio and Li[12]. These computations would need to be performed before the corresponding image database could be searched.

Improved texture algorithm. The numerical experiments presented in this study showed that the simple local-orientation texture algorithm adopted here for our baseline algorithm only marginally aided our image retrieval performance. A wavelet-based algorithm, as reviewed in the paper by Haralick et al[13], has been found to improve the performance of a number of image retrieval algorithms, and has been applied in several RF studies by Rui and Huang[1]. These texture wavelet methods basically follow a three-step process: (1) wavelet transform of the original image to create 10 sub-band images, (2) construction of a feature vector consisting of the variances of the 10 sub-band images, and (3) a normalization of the feature vectors.

Fast, efficient image retrieval via "smart" index structures. In order for any image retrieval code to be regularly utilized by the analyst community, the code must perform its functions quickly and efficiently. A single retrieval round should be concluded in a matter of a few seconds at most, not minutes, even for very large image cube files. The principal factor in determining retrieval speed is the time consumed in calculating the distances (e.g. Euclidean) between the query vector and each of the image vectors in a large database.

Berman and Shapiro[7,8] have utilized the triangle inequality technique, first introduced by Burkhard and Keller[14], to significantly reduce the number of direct distance calculations needed for an efficient search algorithm. The basis of this technique is that the distance between the query and image index vectors cannot be less than the absolute value of the difference between: (a) the distance between the image vector and a designated object known as a key, and (b) the distance between the query vector and the same key or keys. Using the triangle inequality allows one to discard images in the search that are found to be too far from the query image to provide a possible image cube match. Berman and Shapiro also showed that further search efficiencies are possible by taking advantage of a data structure called the *triangle trie* to reduce the number of operations. Their work suggests that "by using a relatively short trie and by storing additional *key* distances in the leaves, we can obtain the best of both worlds". The Berman, Shapiro algorithm (called FIDS[7]) is able to search a database of 37,000 images in an average time of less than one second, depending on the query.

Recent work performed at the University of Illinois, by Xiong, Pottenger, and Huang[15], has adopted the triangle inequality algorithm approach of Berman and Shapiro to speed up a U of I CBIR system with relevance feedback. The feature-weight-updating part of the RF scheme was modified to allow for the incorporation of the triangle inequality algorithm. Preliminary results applied to the GOES-8 hurricane retrieval problem are encouraging.

Based on Xiong's work, estimates have been made of the speed-up in performance of a triangle inequality algorithm (augmented by a triangle trie extention). For a poorly structured index file, it is estimated that it would take over two minutes to calculate about 60,000 distance measures. Using the triangle inequality algorithm, it is estimated that that time can be reduced to approximately 1 second.

These estimates of speed-up would make our CBIR/RF retrieval algorithm quite effective for the interactive retrievals of large image cube databases. This is the goal for the next evolution of our code.

REFERENCES

1. Y. Rui, T. S. Huang, M. Ortega, and S. Mehotra, " Relevance Feedback: A power tool for interactive content-based image retrieval", *IEEE Transaction of Circuits and Systems for Video Technology*, Vol. 8, No.5, Sept., pp. 644-655, Sept. 1998.
2. Y. Rui, T. S. Huang, "Optimizing learning in image retrieval", *Proceedings. IEEE Int. Conference On Computer Vision and Pattern Recognition (CVPR),* Hilton Head, SC, June 2000.
3. Y. Rui, T. S. Huang, and S. Mehotra, " Content-based image retrieval with relevance feedback in MARS", *Proceedings IEEE International Conference on Image Processing, ICIP 97,* October 1997.
4. GOES-8 weather satellite website: http://www.cira.colostate.edu/ramm/newgoes/goesover.htm
5. M.D. King, S.C.Tsay, S.A. Ackerman, and N.F.Larsen, "Discriminating heavy aerosol, clouds, and fires during SCAR-B: Application of airborne multispectral MAS data", Journal of Geophysical Research, Vol.103, No. D24, pp. 31,989-31,999, December 27, 1998.
6. Y.J. Kaufman, R.G. Kleidman, and M.D.King, "SCAR-B fires in the tropics: Properties and remote sensing from EOS-MODIS, Journal of Geophysical Research, Vol.103, No. D24, pp. 31,955-31,968, December 27, 1998.
7. A. Berman and L.G. Shapiro, " A flexible image database system for content-based retrieval", *Computer Vision and Image Understanding*, Vol. 75, pp. 175-195, July/August 1999.
8. A. Berman and L.G. Shapiro, " Triangle-inequality-based pruning algorithms with triangle tries", *Proceedings of IS&T and SPIE Storage and Retrieval of Image and Video Databases VII*, San Jose, CA, Jan 1999.
9. B. Jahne, *Digital Image Processing*, Springer-Verlag, chapter 11, 1997.
10. IDL website: http://www.rsinc.com/idl/index.cfm
11. G. Healey and A. Jain, "Retrieving multispectral satellite images using physics-based invariant representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No.8, August 1996.
12. G. Marchisio and A.Q. Li, " Intelligent system technologies for remote sensing repositories," in *Information Processing for Remote Sensing*, C.H. Chen editor, World Scientific, 1999.
13. R.Haralick, K.Shanmugam, and I.Dinstein, "Textural Features for Image Classification", in *Systems, Man, and Cybernetics*, November 1993.
14. W.A.Burkhard and R.M.Keller, "Some approaches to best-match file searching", *Comm. ACM* Vol. 16 No.4, pp. 230-236, 1973.
15. Z. Xiong, W.M. Pottenger, and T.S. Huang, "Speeding up relevance feedback in image retrieval with triangle inequality algorithm", submitted for publication September 2000.