Relevance Feedback: Perceptual Learning and Retrieval in Bio-computing, Photos, and Video

Micha Haas, Joachim Rijsdam, Bart Thomee, and Michael S. Lew
LIACS Media Lab
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
mlew@liacs.nl

ABSTRACT
One of the most important characteristics about relevance feedback is that it ideally finds a set of human perceptually correlated results because the user is directly involved in the search process. In principle, relevance feedback is an iterative learning process where positive and negative examples accumulate as the user gives feedback on each new iteration of results. If we view relevance feedback as a learning problem then we can immediately grasp that there will be the associated problem of learning from a small training set. Towards a solution, we present MediaNet, which is an approach toward integrating additional knowledge sources into the relevance feedback process. The additional knowledge sources are used to shape the learning space when insufficient training samples are available. We also integrate genetic or evolutionary algorithms directly into the search process. Experiments are given on test collections in bio-computing, general photos and video.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval]**: Information Search and Retrieval - relevance feedback. I.2.6[**Computing Methodologies**]: Artificial Intelligence-*learning*

General Terms
Algorithms

Keywords
Relevance feedback, bio-computing and multimedia retrieval

MIR'04, October 15–16, 2004, New York, New York, USA.

1. INTRODUCTION
The advent of prevalent digital imagery combined with rapidly decreasing storage costs has led to a worldwide need for effective methods for finding imagery in all its forms: color photos, MRI, X-Ray, video, etc. A wide variety of retrieval systems have been presented by the research community of which these are representative examples [2-7] [9][12-15] [17-18] [23].

The early image retrieval tools started off as derivatives of either computer vision or text retrieval tools. Computer vision approaches typically began with low-level features and direct similarity measures. In text approaches, keywords were assigned to images using statistical quantifiers, where the context of an image within HTML can be described by statistically associating relevant words to the conceptual context of that image, through singular value decomposition. These tools usually did not use the image information itself, but due to the sheer size of the WWW and its textual nature they were still useful [23]. After these came SQL based retrieval systems. It is however challenging to efficiently navigate these databases. Therefore recent research in the field of image and video retrieval has focused on the actual content of the images, giving rise to Content-Based Retrieval (CBR).

It was however soon shown that it was very hard to find satisfactory results after just a single step, which has been called the Page Zero Problem. To overcome this problem, interactive techniques to guide the retrieval process were applied to these image retrieval systems. Most of them were based on the Query By Example method. And the strategy that showed the most promise was Relevance Feedback [1], originally developed for textual retrieval. It is supposed to tackle the main problem of current image retrieval applications, that of translating the semantically description that the user has in mind of a particular picture, and the automatically extracted syntactical image features that are stored alongside the large image databases.

Relevance Feedback is a process where the user can guide the retrieval by interactively updating the search query. This interactive approach moves away from the computer centric approach where retrieval was performed by fixed weight feature comparison, and tries to include the user into the loop of the retrieval process by dynamically and interactively updating the weights applied to different feature vectors. Rui et al. [6-7] describe how this interactive approach can aid image retrieval, and explain the different kinds of relevance that can be assigned to objects. Roughly, the relevance of a particular image to a query can be decomposed into two streams: positive and negative relevance feedback. Multiclass methods for relevance feedback with a performance comparison are given in Peng [4].

2. RELEVANCE FEEDBACK

The original Rocchio [1] formula attempts to move the current query point toward the estimate of the ideal query point. The iterative estimation for relevant documents, D'_R and non-relevant documents, D'_N obeys the following equation:

$$Q' = \alpha Q + \beta\left(\frac{1}{N_{R'}}\sum_{D'_R} D_i\right) - \gamma\left(\frac{1}{N_{N'}}\sum_{D'_N} D_i\right)$$

where α, β, γ are constants and $N_{R'}$ and $N_{N'}$ are the number of documents in D'_R and D'_N, respectively.

The general principle behind the Rocchio method is straightforward: move the query point toward the relevant documents and away from the non-relevant documents. We modify the usage of the Rocchio formula by adding a variable weight for the relevance or non-relevance.

3. KNOWLEDGE AND REL. FEEDBACK

Our approach toward expanding the training set is to integrate the knowledge from additional sources toward expanding the training set and consequently improving the accuracy of the relevance feedback results. In principle, the relevance of an unknown item based on the interactions between the elements in the additional sources and the known user classified items can be expressed as

$$R_X = P(X \mid R_1, \ldots R_N, A_1, \ldots, A_M)$$

where R_x is the relevance of an unknown item X based on the influences from the known items, R_i, and the annotated items, A_k

Many different kinds of annotated databases exist today. Examples include local personal collections, Google and Altavista, Bio-computing databases, Computer-aided diagnosis databases, DVD movies with subtitles, news feeds like Routers or the AP as displayed in Figures 1 and 2. Our goal is to use any or all of these to improve search results.

We categorize the databases by their usage in the search context. A *search database* is one in which the user would like to find media. A *domain knowledge database* consists of at least two different types of media linked with each other (i.e. a text annotated image) which can be used to assist in the search process. A *bridge database* has one type of media which are linked together via relationships. A dictionary would be a common example of a bridge database because text is linked to text.

For the purposes of this work, the unannotated database is the search database and the annotated database is the domain knowledge database. The bridge database is WordNet[22] or MediaNet, which will be discussed in more detail in the next section

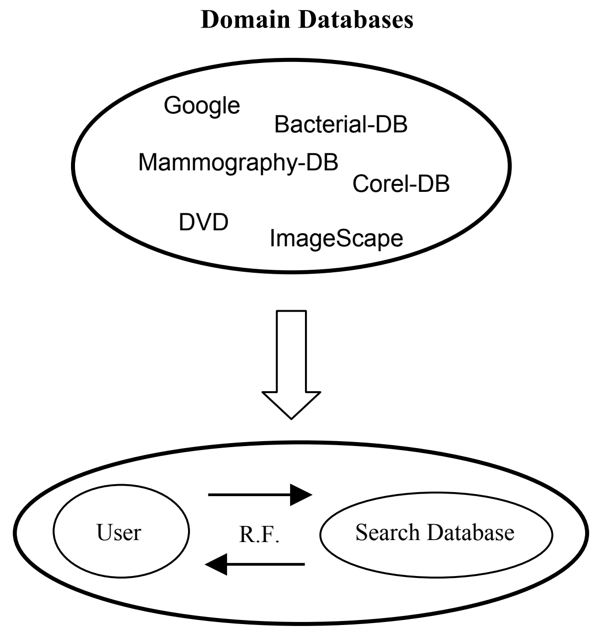

Figure 1. Use Domain Databases to assist in relevance feedback

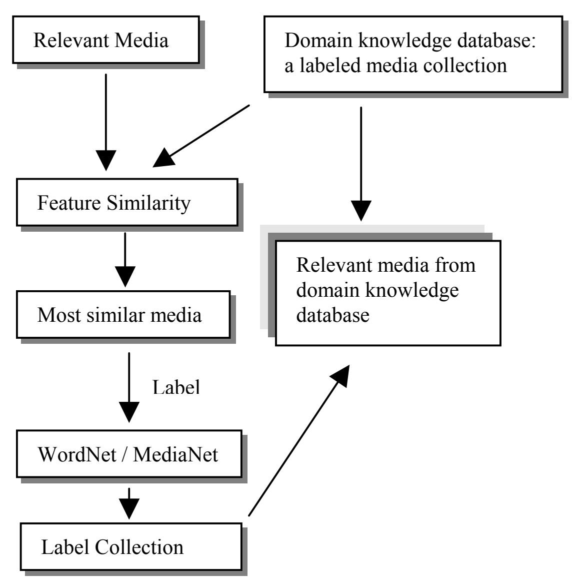

Figure 2. Using Domain Knowledge Database Information.

Let D_U be the collection of media which contains the potentially relevant media. Let D_E be the collection of other databases available as external knowledge and users choices be:

very non-relevant = -3
non-relevant = -2
somewhat non-relevant = -1
denotes uncertain = 0
somewhat relevant = 1
relevant = 2
very relevant or a perfect match = 3

For the inter-word relationships, we utilized WordNet[22], which is a free downloadable database of word definitions and synonyms. WordNet is critical in that it provides a way to link annotated images together. The assumption is that if we can find one relevant image in an additional annotated source, then with WordNet, we can locate other semantically similar images via the WordNet links.

In method *RF-Wordnet*, we take a relevant or very relevant image, search for the most similar text annotated image, search Wordnet for similar (synonyms) text, and insert the images which have the similar text into the relevance model as somewhat relevant.

RF-Wordnet has the advantage that it links the text annotated images together in a relevant manner. This helps in populating the training set with more samples.

However, we can go farther. We know that some synonyms are closer than others. We also know that some descriptive words are more ambiguous than others. Therefore, our natural extension to WordNet is to associate relevance probabilities with the words which results in MediaNet.

In the current version of MediaNet, we have approximated the relevance probabilities by making them proportional to the number of "senses" of the word, which is already within WordNet. Then we manually adjusted the relevance probabilities of approximately 1400 common descriptive words.

Thus, method *RF-MediaNet* works as follows: When the user selects a particular media as relevant or very relevant, we use the MediaNet to find the set of highly perceptually correlated in D_E, which we denote as D_S. Each element of D_S is added to the set of relevant media with a relevance factor from MediaNet. Furthermore, we also add all media linked through the WordNet text model to D'_R with appropriate relevance factors from MediaNet.

There are several different possibilities for choosing the images in the next relevance feedback iteration. The most common method, *Near-N*, which is used for methods *Rocchio* and RF-WordNet displays the N images which are closest to the query point.

In RF-MediaNet, we do not choose the images closest to the query point. Instead, we choose the next set of images by following the evolution of a set of images according to the genetic optimization algorithm, GAS [19]. Genetic or evolutionary algorithms have the ability to find multiple local maxima in high dimensional spaces. In principle, genetic algorithms function as follows: For each generation, perform crossover to generate new children with mutation factors. In MediaNet, each child is a query point in the high dimensional feature space. The fitness function is performed by the user when he manually rates the relevance of each image.

Our alteration to GAS is to show the user the N images which maximize the increasing gradient of the relevance factor over the feature space. This is one of the important reasons why we have several different levels of user relevance feedback. If there is only one level of relevance for an item, it is not possible to know the direction of increasing relevance.

In general, we assume that the user is searching for media from the *unannotated* collection or database. One or more sources composed of *annotated* imagery are used to improve the accuracy of the results.

4. PHOTOS: GENERAL

For content based image retrieval, the user is attempting to find a specific image from a large collection of unannotated imagery. For the additional source, we used a database from another project, ImageScape[23], which consists of 25 million images with weak annotation collected from the World Wide Web.

For the content similarity, we used HSV color features quantized to 4:2:2 bits for V, H, S, respectively. For texture, we used the local binary pattern (LBP) texture features as explained in [10], and for shape we used the moment invariants features from [8]. Searching for the most similar image from the ImageScape collection was accomplished using the logarithmic *kd-tree* algorithm [16].

The unannotated database consisted of 6000 images from the Corel Stock Photo Collection mixed with 20,000 images from the WWW as shown in Figure 3. The number of images shown to the user in each iteration of the relevance feedback was 12. The computational platform included a Pentium IV CPU at 3 Ghz with a Terabyte RAID.

In the experiments, 23 users were asked to find as many images as possible in the unannotated database for targets: lion, jet, flower, rabbit, landscape, beach, and celebration. The users were also asked to count the number of images shown in iterations 0, 5, 10, and 15, which they considered to be relevant or very relevant. Figure 4 displays the comparative relevance accuracy for each method.

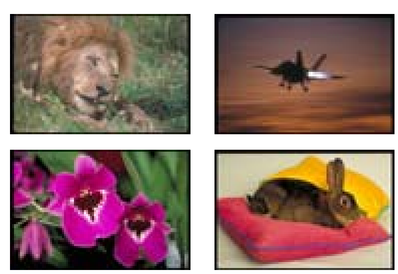

Figure 3. Examples of images in the unannotated database

5. BIO-COMPUTING

In Bio-Computing, there are a wide variety of different uses for content based retrieval. One example is the analysis, biological modeling, and understanding of bacteria properties. For example, in the case of Deinococcus bacteria, current research has found that it is extremely resilient to extended dry environments and more importantly, it can withstand extremely high levels of radiation.

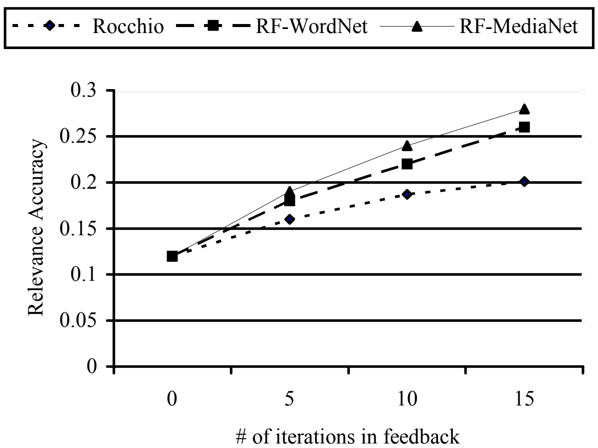

Figure 4. Image retrieval relevance accuracy

Some biologists consider it to be the "toughest lifeform in existence" or "Conan the Bacterium." Another current challenge is in "bio-defense." This includes finding dangerous spores such as anthrax. Thus, the analysis and biological modeling of viruses is useful for both biologists and medical practitioners. Examples are displayed in Figures 5 and 6.

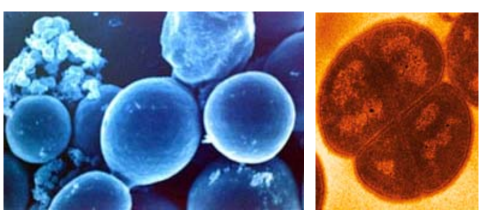

Figure 5. Understanding aspects of Deinococcus bacteria

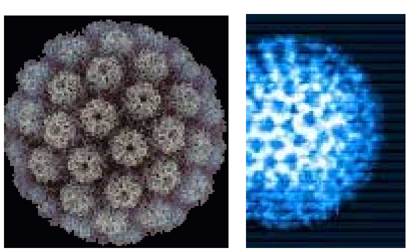

Figure 6. Identification of viruses

The content similarity was measured using the same features as in the Section 4. For these tests, we used databases from the National Institutes of Health (NIH) and Harvard University. The annotated database consisted of 7100 images of bacteria and viruses from confocal and electron microscopy. The unannotated database held 2300 images. The goal for the users was to find relevant images to a target from the unannotated database.

In this experiment, 14 users were asked to find 9 varieties of bacteria and viruses in the non-annotated database. They were asked to count the number of relevant images in iterations 0, 5, 10, and 15. *RF-MediaNet* refers to using the bacteria/virus annotated database as the additional source. Figure 7 displays the comparative relevance accuracy for each method.

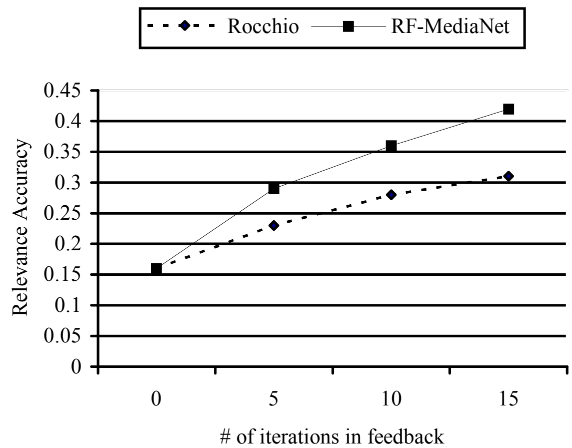

Figure 7. Bio-Computing: Identification of bacteria and virus relevance accuracy

In Computer Aided Diagnosis (CAD), the computer is used to assist a medical practitioner in diagnosing radiology imagery. Early diagnosis of cancer allows for a wider set of treatments and a greater likelihood of survival for the patient.

In this area, we used the digital mammography database from USF [21], which has labeled examples of cancerous growths and normal tissue. Relevance feedback is used to improve the automatic diagnosis of new images and also to explain why a particular diagnosis was made. Figure 8 contains 2 examples of a cancerous growth.

The annotated collection consisted of 9200 images. The unannotated collection consisted of 1800 images.

For this set of experiments, 14 users were given 5 unannotated images and asked to find similar examples from the annotated databases. Each user was asked to count the number of relevant or very relevant images in iterations 0, 5, 10, and 15 for methods Rocchio and RF-MediaNet.

In our experiments, RF-MediaNet refers to using the USF database as the annotated additional source. Figure 9 displays the results of the tests.

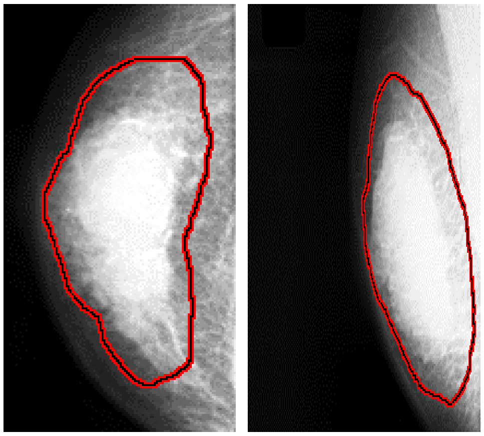

Figure 8. Mammography Example of Cancer

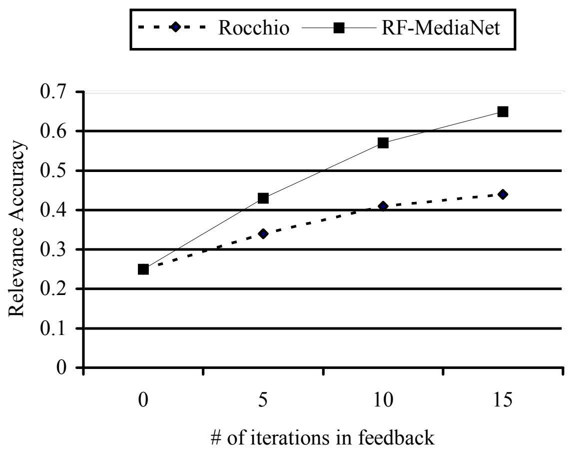

Figure 9. Bio-Computing: CAD relevance accuracy

6. VIDEO

Our third set of experiments were performed on a set of 1000 full-length movies encoded in DivX AVI format. For this topic area we examined the problem of thumbnail extraction for video summarization. Examples are given in Figure 10. The problem can be stated as follows: Given a full length movie, determine the best N thumbnails for summarizing the video.

Figure 10. Examples of thumbnails from the Phantom Menace

For the content similarity features, we used the features in the image retrieval section appended with inter-frame motion vectors of a 9x9 grid. The motion vectors for each grid point were found by searching in a 40x40 window using the L_1 distance metric.

In these experiments, 23 users were asked to find "good" representative thumbnails for 10 movies. Each user was asked to count the number of relevant or very relevant images in iterations 0, 5, 10, and 15. Figure 11 displays the results.

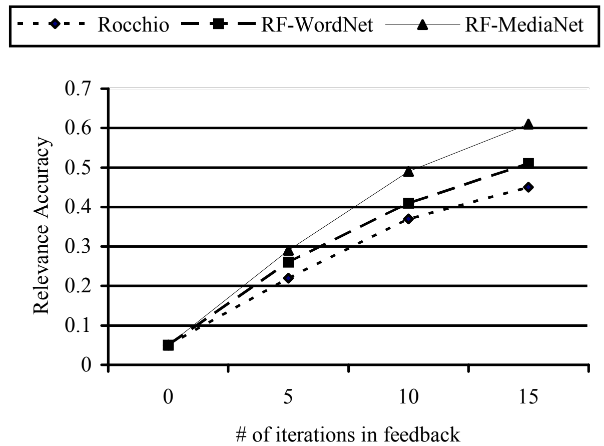

Figure 11. Video Retrieval: Summarization accuracy

7. DISCUSSION & CONCLUSIONS

Our main contribution was in integrating domain knowledge sources with genetic algorithms into the relevance feedback search process. We tested our method on four different problems: image retrieval, identification of bacteria and viruses, computer aided diagnosis, and video summarization.

The traditional Rocchio method has the advantage of low computational complexity and an intuitive query method - all media (images or video) are either relevant or non-relevant. In RF-MediaNet we expanded the feedback to 7 possibilities.

Expanding the feedback had the significant advantage of integrating additional knowledge sources while allowing higher levels of relevance for the user specified media. By having multiple levels of feedback, we could also integrate a genetic algorithm into the search process by treating the search problem as finding points in feature space which maximized the relevance. RF-MediaNet was computationally intensive compared to the Rocchio method, but also consistently yielded improved retrieval accuracy

In our experiments, the most significant gain in retrieval accuracy was found in the bio-computing areas. We expect this is due to the fact that the additional annotated databases have more accurate annotation and are focussed on the correct subject.

In future work we intend on investigating other models for exploiting knowledge such as combining multiple classifiers or inserting domain knowledge rules into the query system.

Comparing other genetic algorithms is also a promising area toward finding the local maxima.

8. ACKNOWLEDGMENTS

We would like to thank Harvard University and the National Institutes of Health (NIH) for their expertise and assistance in the bio-computing area. We appreciate the support from the Netherlands National Science Foundation (NWO), Leiden University, Senter, and the Cyttron project.

9. REFERENCES

[1] J. J. Rocchio, Relevance Feedback in Information Retrieval, *The SMART Retrieval System, Experiments in Automatic Document Processing*, Prentice Hall, Englewood Cliffs, New Jersey, USA, pp. 313-323, 1971.

[2] M. La Cascia, S. Sethi, and S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. *IEEE Workshop on Content-based Access of Image and Video Libraries*, June 1998.

[3] Z. Stejic, Y. Takama, and K. Hirota. Genetic algorithm-based relevance feedback for image retrieval using local similarity patterns. In *Information Processing and Management, vol. 39, Jan. 2003, pp. 1-23, 2*003.

[4] J. Peng. Multi-class relevance feedback content-based image retrieval. In *Computer Vision and Image Understanding,* vol. 90, issue 1, pp. 42-67, 2003.

[5] I.J. Cox et al. Hidden annotation in content-based image retrieval. *IEEE Workshop on Contentbased Access of Image and Video Libraries*, pages 76–81, June 1997.

[6] Y. Rui et al. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Video Technology*, 8(5):644–655, Sept. 1998.

[7] Y. Rui et al. Relevance feedback techniques in interactive content-based image retrieval. Technical report, Beckman Institute and Dept. of Computer Science, 1998.

[8] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRA Trans. Information Theory*, Feb., pp. 179-187, 1962.

[9] X. Zhou and T. S. Huang. Relevance Feedback in Image Retrieval: A Comprehensive Review. *Multimedia Systems Journal, 8(6),* pp. 536-544, 2003.

[10] T. Ojala, M. Pietikainen and D. Harwood, A Comparative Study of Texture Measures with Classification Based on Feature Distributions, *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[11] L. Liu and S. Sclaroff. Index trees for efficient deformable shape-based retrieval. In *proceedings IEEE workshop on Content-based Access of Image and Video Libraries*, pp. 83–87, June 2000.

[12] S.MacArthur, C. Brodley, and C. Shyu. Relevance feedback decision trees in content-based image retrieval. In *proceedings IEEE workshop on Content-based Access of Image and Video Libraries*, pp. 68–72, June 2000.

[13] W. Niblack and R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos and G. Taubin, The QBIC project: Querying images by content using color, texture and shape. In *Proceedings SPIE Storage and Retrieval for Image and Video Databases*, pp. 173-181, 1993.

[14] A. Pentland, R. Picard, and S. Sclaroff, Photobook: tools for content-based manipulation of image databases. In *SPIE Proceedings Series*, volume 2185, pp. 34-47, 1994.

[15] J.R. Bach, C. Fuller, and A. Gupta,. The Virage image search engine: An open framework for image management. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, pp. 76-87, Feb 1996.

[16] R. Egas, D. P. Huijsmans, M. S. Lew, N. Sebe, "Adapting k-d Trees to Visual Retrieval," Proceedings of VISUAL'99, Springer-Verlag, Berlin, pp. 533-540, 1999..

[17] Y. Rui, T. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in MARS. In *Proceedings of the IEEE International Conference on Image Processing*, pp. 815-818, 1997.

[18] D. McG Squire, S. Wolfgang, and M. Henning. Relevance feedback and term weighting schemes for content-based image retrieval. In *Third International Conference On Visual Information Systems*, pp. 2–4, June 1999.

[19] M. Jelasity and J. Dombi, GAS, A Concept on Modeling Species in Genetic Algorithms, Artificial Intelligence, vol. 99, no. 1, pp. 1-19, 1998..

[20] N. Vasconcelos and A. Lippman. Bayesian relevance feedback for content-based image retrieval. In *proceedings IEEE workshop on Content-based Access of Image and Video Libraries*, pp. 63–67, June 2000.

[21] K. Bowyer, D. Kopans, W.P. Kegelmeyer, R. Moore, M. Sallam, K. Chang and K. Woods, The Digital Database for Screening Mammography, *Proceedings of the Third International Workshop on Digital Mammography*, Chicago, June 1996.

[22] C. Fellbaum, **WordNet: An Electronic Lexical Database**, Bradford Book, May, 1998.

[23] M. Lew, Next Generation Web Searches for Visual Content**,** *IEEE Computer*, November, pp. 46-53, 2000.