# RELEVANCE FEEDBACK METHODS IN CONTENT BASED RETRIEVAL AND VIDEO SUMMARIZATION

*Micha Haas, Ard Oerlemans, Michael S. Lew*

LIACS Media Lab, Leiden University, The Netherlands

## ABSTRACT

In the current state-of-the-art in multimedia content analysis (MCA), the fundamental techniques are typically derived from core pattern recognition and computer vision algorithms. It is well known that completely automatic pattern recognition and computer vision approaches have not been successful in being robust and domain independent so we should not expect more from MCA algorithms. The exception to this would naturally be methods which are human-interactive or not automatic. In this paper, we describe some of the recent work we have done in multimedia content analysis across multiple domains where the fundamental technique is founded in interactive search. Our novel algorithm integrates our previous work from wavelet based salient points and genetic algorithms and shows that the main contribution and improvement is from the user feedback provided by the interactive search.

## 1. INTRODUCTION

Automatic pattern recognition and computer vision algorithms have been effective at a wide variety of tasks but typically have been found to be brittle in the sense of working accurately across multiple areas, which we refer to as domain brittleness.

We believe that the natural solution is to focus on human-interactive methods whereby the human input can address the brittleness of the recognition and vision algorithms.

In particular, there has been a recent surge of interest in the area of interactive algorithms. The popularized name for these in the domain of content based retrieval (CBR) is relevance feedback (RF). Although relevance feedback does not encompass all of the interactive methods, the term is typically used synonomously for human-interactive adjustments or refinements in CBR algorithms.

The fundamental principle behind relevance feedback methods is that the user can give feedback (i.e. good or bad; or positive or negative) on intermediate results. The feedback is used in iterative learning algorithms toward improving the accuracy of the results.

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 usage of different feature vectors. It is beyond the scope of this article to give a complete listing of the relevance feedback literature. A good survey which covers how relevance feedback has been used in iterative adjustment, estimation approaches, and classification can be found in Zhou and Huang [1] and some interesting work by Rui, et al. [7], Muller, et al., [8] and Peng [9]

## 2. RELEVANCE FEEDBACK

The general principle behind the Rocchio method is straightforward: move the query point toward the relevant documents and away from the non-relevant documents. The original Rocchio [2] 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 $\alpha$, $\beta$, $\gamma$ are constants and $N_{R'}$ and $N_{N'}$ are the number of documents in $D'_R$ and $D'_N$, respectively.

The Rocchio technique has the main advantages of simplicity, being intuitive, and has been reported to work well and effectively using small training sets or user interaction.

It should be clear that with each iteration of user feedback, the algorithm accumulates information which can be regarded as positive and negative examples. Furthermore, these examples can be used as input to more sophisticated learning algorithms toward improving the accuracy of the next set of results.

*0-7803-9332-5/05/$20.00 ©2005 IEEE*

## 3. GENETIC ALGORITHMS AND SALIENCY

One of the important unsolved problems in CBR and RF is subimage based matching. Currently, it is computationally infeasible to search through every possible scale and block of every image in a large database. Therefore, we must find a good subset of the set of all blocks/subimages. This is where salient points/regions becomes important.

In our RF setup, we have the following choices which allow the user to express gradations of relevance factors for each block:

very non-relevant = -1.0
non-relevant = -0.6
somewhat non-relevant = -0.3
denotes uncertain = 0
somewhat relevant = 0.3
relevant = 0.6
very relevant or a perfect match = 1.0

### 3.1. Wavelet based Salient Regions

A salient point is in principle an "interesting" point, that is, a point which might contain pictorial information which would be of interest to the user. For this work, we implemented both contrast based interest operators and wavelet based salient points as reported in our previous work in Tian, et al. [3]. The principle of the wavelet based saliency is that the Daubechies wavelet coefficients correlate to interesting pixels or regions of an image. Examples of images and the locations of the wavelet based salient points are shown in Figure 1. After preliminary experiments, we found that the wavelet based saliency to be closer to human psycho-visual saliency than the contrast based operators.

The important contribution of the salient points is toward reducing the total number of blocks which must be searched. The wavelet based salient point approach typically reduces the number of potentially interesting pixels from hundreds of thousands to just hundreds, that is, a thousandfold reduction in computational effort.

### 3.2. Genetic Algorithm Refinement

As opposed to the Rocchio approach, 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 paradigm of genetic algorithm based optimization. Genetic or evolutionary algorithms (for example, see [4]) 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 our approach, 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 block or subimage.

In the typical genetic algorithm approach, the predecessors would be of a previous generation and be removed. In our approach, we retain all of the blocks from all previous generations which is why we denote it as the *Ancestral Genetic Support* (AGS). For an image I which contains blocks, $\{b \in I\}$, and given the set of user feedback blocks, $\{D'_R, D'_N\}$, the AGS is defined as the support given by all of the similar blocks, $b_{sim}$, to $\{D'_R, D'_N\}$.

In contrast to the typical genetic algorithm, we show the user the N images which maximize the AGS and the increasing gradient of the AGS over the feature space. Modeling the AGS based on human perceptual similarity is another ongoing project. *Specifically, we feel the mapping from M blocks in an image to the global relevance of the image is an open research problem.*

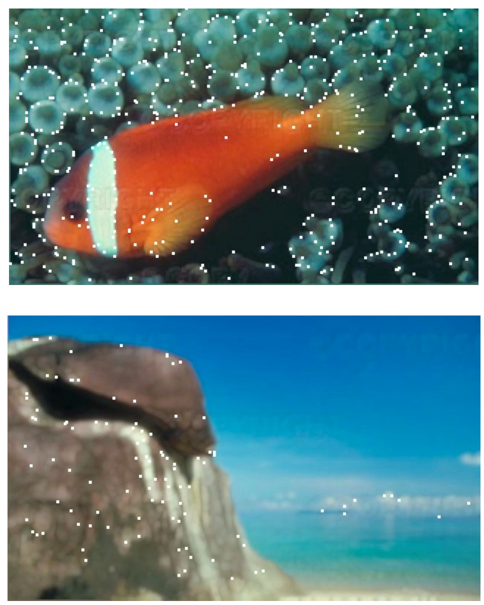

Figure 1. Images with salient (wavelet) points depicted as white squares.

## 4. AGS-GENE ALGORITHM

In constructing the working system, several issues had to be addressed. First, after the salient points algorithm returns a list, what regions should be used for the feature information around the points? Our fundamental intuition is that we need a block containing the location of the salient point which both has sufficient pixel information to be meaningful for features such as color and texture histograms, but does not step beyond the local information. Blocks which are too large will bias the feature information toward the main global features of the whole image.

Our method of addressing this issue is to introduce the notion of a "*Stable Block*" which is defined as a block of minimum size 15x15 centered at the salient point which is grown in size until one of two conditions occur: (i) any major feature changes by more than $T_{SB}$ percent or (ii) the block intersects with a different salient point. From a practical perspective, 15x15 gives a minimal amount of statistical samples for our histograms, and we should be able to increase the size of the block and consequently the samples as long as the major features do not change significantly. So, each image in our database is processed with the wavelet salient point detector and then the feature vectors based on the stable blocks are stored for indexing.

The query system can be described as follows:

(1) Display 20 random images to the user
(2) The user sets the relevance factor in the interface and then clicks on 1 or more relevant pixels in the "Image Results" list. A "Stable-Block" is found at each clicked location and then added to the complete list of feedback blocks.
(3) The next set of "Image Results" is found using the AGS genetic algorithm as described in Section 3.2. using a summation of the relevance factors.
(4) Go to Step (2)

## 5. EXPERIMENTS

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 [5], and for shape we used the moment invariants features from [6]. $T_{AGS}$ and $T_{SB}$ were both set at 10 percent.

In our experiments, we used 3 different domains and problems, namely the Corel test set (6000 images), a biological virus database (7100 images), and a video database (10 movies and TV shows) which are shown in Figures 2, 3, and 4, respectively. All of these domains were analyzed with Rocchio as a benchmark, AGS-Gene1 (repopulate with most relevant images), and AGS-Gene2 (repopulate with summation of most relevant and highest gradient relevant images). On a 3Ghz P4 with 2 GB RAM, the average response time was less than 1 second per iteration.

In the content based retrieval experiments, 21 users were asked to find as many similar images as possible for known 15 targets in the Corel and Virus databases. Example targets were aircraft, beach, fish, etc. for the Corel test set. For the Virus database example targets were Rubeola, Influenza, Rubella, etc. 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. Results are shown in Figures 5 and 6.

In the video summarization experiments, 28 users were asked to find "excellent" representative thumbnails for 10 videos. For the content similarity features, we used the features in the image retrieval section appended with inter-frame motion vectors of a 9x9 grid found from correlation. Another question we had was whether the learned blocks from the video summarization tests could be used to improve the thumbnail selection for new videos. In another test we found that the relevance accuracy of automatically found "excellent thumbnails" improved from 4.9 to 26 percent. Results are displayed in Figure 7.

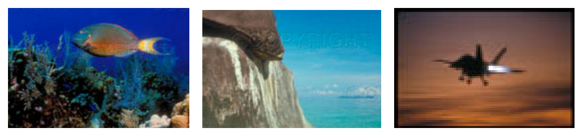

Figure 2. Examples of images in the Corel Set

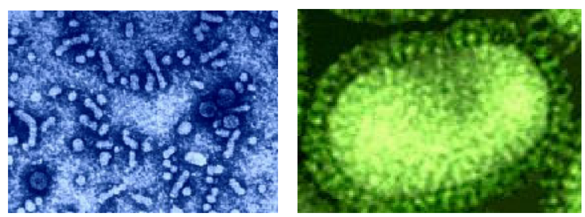

Figure 3. Examples of images in the Virus database.

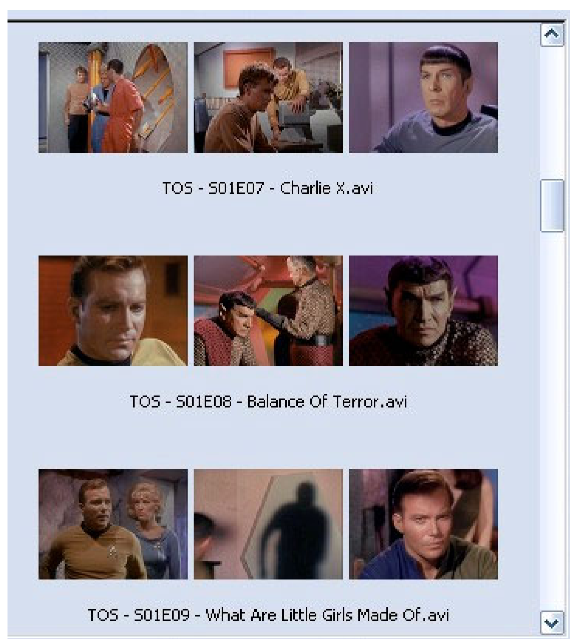


Figure 4. Video Summarization Example

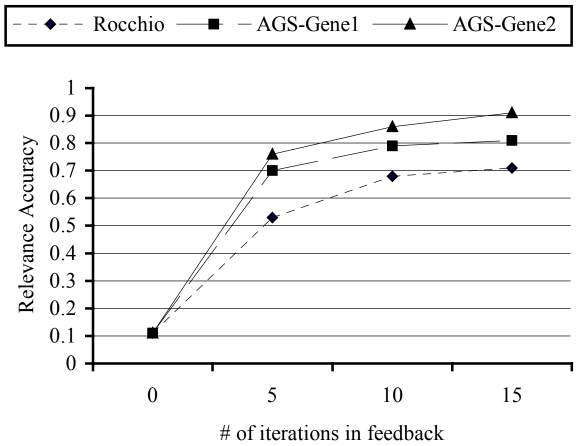


Figure 5. Corel test set relevance accuracy

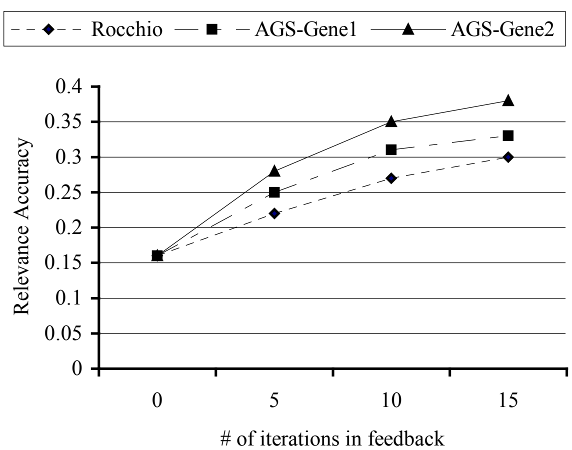


Figure 6. Virus database relevance accuracy

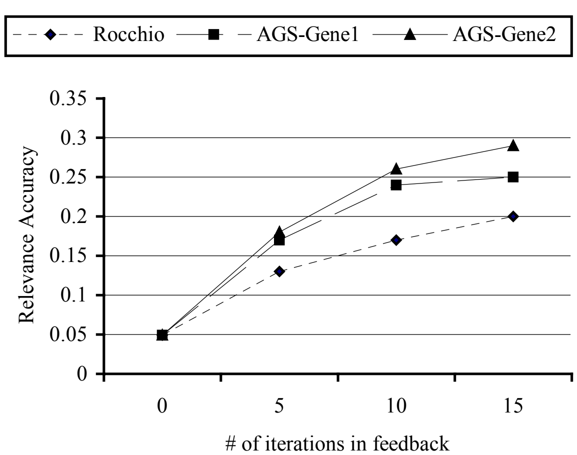


Figure 7. Video summarization relevance accuracy

## 6. DISCUSSION AND CONCLUSIONS

In this paper we described a novel system for using relevance feedback in content based retrieval and video summarization. We addressed the subimage problem by selecting wavelet-based salient points and then using the Stable-Block information for the indexing. The search process uses the AGS genetic algorithm for finding the most relevant images to the interactive user query. The system was tested on several databases and showed consistent improvement in terms of relevance accuracy as the amount of user interaction increased. We also found that using the gradient of the relevance factor is an important factor in achieving better relevance accuracy. Future work will focus on improving the AGS modeling toward improved perceptual similarity.

## ACKNOWLEDGEMENTS


We are grateful for collaboration with Harvard University and partial support by the Bsik sponsored BioComputing Search Project of Cyttron.


## REFERENCES


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

[2] 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.

[3] Q. Tian, N. Sebe, M.S. Lew, E. Loupias, T.S. Huang, Image Retrieval using Wavelet-based Salient Points, *Journal of Electronic Imaging*, vol. 10, no. 4, pp. 835-849, 2001.

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

[5] 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.

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

[7] 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.

[8] H. Müller, W. Müller, S. Marchand--Maillet, D. McG Squire. Strategies for positive and negative relevance feedback in image retrieval. *In Proceedings of the International Conference on Pattern Recognition* (ICPR'2000), Barcelona, Spain, September 3--8, 2000.

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