

# An Intelligent Image Retrieval System Based on the Synergy of Color and Artificial Ant Colonies

Konstantinos Konstantinidis, Georgios Ch. Sirakoulis, and Ioannis Andreadis

Laboratory of Electronics, Dept. of Electrical and Computer Engineering, Democritus University of Thrace, 12 V. Sofias Str., 67100 Xanthi, Greece  
{konkonst, gsirak, iandread}@ee.duth.gr

**Abstract.** In this paper a new image retrieval algorithm is proposed which aims to discard irrelevant images and increase the amount of relevant ones in a large database. This method utilizes a two-stage ant colony algorithm employing in parallel color, texture and spatial information. In the first stage, the synergy of the low-level descriptors is considered to be a group of ants seeking the optimal path to the “food” which is the most similar image to the query, whilst settling pheromone on each of the images that they confront in the high similarity zone. In the second stage additional queries are made by using the highest ranked images as new queries, resulting in an aggregate deposition of pheromone through which the final retrieval is performed. The results prove the system to be satisfactorily efficient as well as fast.

**Keywords:** Ant Colony Algorithm, Image Retrieval.

## 1 Introduction

Since multimedia technology is daily enhanced, vast image, video and audio databases are rapidly growing. These digital libraries are produced by various applications such as medicine, military, entertainment or education and belong to either private or public (World Wide Web) databases. The need for these databases to be indexed has led to an increasing interest on the research and development of automatic content-based image retrieval systems. Effective retrieval of image data is important for general multimedia information management. For an image to be retrievable, it has to be indexed by its content. Color can provide significant information about the content of an image. Among the methods that use color as a retrieval feature, the most popular one is probably that of color histograms [1,2]. The histogram is a global statistical feature which describes the intensity distribution for a given image [1]. Its main advantage is that low computational cost is required for its manipulation, storage and comparison. Moreover, it is insensitive to rotation and scale of the image scene and to any displacement of objects in the image. On the other hand it is also somewhat unreliable as it is sensitive even to small changes in the context of the image. Swain and Ballard [2] proposed a simple but nonetheless effective method of matching

images through the intersection of their color histograms. Other low-level features widely used by researchers for indexing and retrieval of images, except color [1,2], are texture [1,3] and shape [1,4]. In order to exploit the strong aspects of each of these features while constructing an optimum and robust CBIR system, a plethora of methods, introduced over time, have been based on combinations of these features [4,5].

In this paper the synergy of such features, specifically color and texture, is performed by use of an artificial ant colony. This specific type of insect was selected since studies showed that when in groups, the ants show self-organization as well as adaptation which are desirable attributes in image retrieval. Artificial ant colonies have previously been used in text based site retrieval (i.e. search engines) [6] but also in texture classification [7]. To the best of our knowledge this is the first time that ant colony behavior is applied on image retrieval in very large databases which contain images of general interest, as is the case with the LabelMe database (80000 images and increasing) [8]. The main thrust of the proposed method is a two stage ant colony algorithm employing in parallel color, texture and spatial information which are extracted from the images themselves. The study of ant colony behavior and of their self-organizing abilities inspired the algorithm. In the first stage, the synergy of the low-level descriptors is considered to be a swarm of ants, seeking for the optimal path to the “food” which is actually the most similar image to the query, whilst settling pheromone on each of the images that it confronts in the high similarity zone represented by 200 images. The terrain on which the ants move is predefined through the low-level features of the query image. The features involved are a newly proposed spatially-biased color histogram, a color-texture histogram and a color histogram inspired by attributes of the Human Visual System (HVS). Although these particular descriptors are proposed for use in this paper, the ant colony algorithm can employ a variety of features depending on the implementer or user. In the second stage the terrain changes as additional queries are made by using the highest ranked images as additional query images. The results prove the system to be satisfactorily efficient as well as fast since the features have already been extracted.

Following the histogram creation procedures resembling the “food” in the proposed method which are illustrated in the next section, the Matusita distance metric is used in order to measure the similarity between the features of the query image and of the ones stored in the database. In this manner the pheromone is dispensed to the corresponding images which form the path to the image of highest similarity and through descending order to the less similar ones.

## 2 Feature Extraction

The three features which represent the ants seeking the closest “food” (images) are a spatially biased histogram, a color-texture histogram and a color histogram inspired by the attributes of the HVS.

## 2.1 Spatially-Biased Histogram

The use of global histograms for image retrieval has proven to be an efficient and robust retrieval method [1,2], as it describes the overall statistics of the color in the images and is insensitive to rotation and scaling of the images themselves. However, such techniques lack in cases in which the images have similar colors, but are spatially distributed differently. This led to the need of the adoption of global histograms with embedded local characteristics, such as the newly proposed spatially-biased histogram. The suggested histogram creation method has a two stage straightforward algorithm and only the hue component is enriched with spatial information so as to maintain the original histogram speed.

In the first stage a histogram is created with the hue component being divided into 16 regions, whereas saturation and value into 4 each. This unbalance is due to the fact that the hue component carries the majority of color information from the three in the HSV color space and is hence considered more important in this method. Finally, the three color components are interlinked, thus creating a (16x4x4) histogram of 256 bins. In the second and most important part of this method the spatial information is inserted into the final histogram via the use of the mask  $M$  illustrated below. This mask is used to collect the color information from a 5 pixel "radius" neighborhood for each separate pixel in the manner of a shattered cross so as to increase the speed of the system and to decrease the chance of taking into consideration random noise; although this is also covered by use of a threshold when checking the concentration.

$$\begin{array}{cccccccc}
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 M = & 1 & 0 & -1 & 0 & 1 & 1 & 1 & 0 & -1 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
 \end{array}$$

The negative values in the cross are used to counteract the fact that significantly different adjacent pixel values could have the same sum as identical pixel values. In this way the possibility of false positive is somewhat decreased, though maintaining the extraction speed. Thus the whole image is convolved with the  $M$  matrix, resulting in a new hue component which contains the color information for the neighborhood of each pixel. If the pixels which are included in the vicinity of the full length of the cross possess a color similar to the one of the central pixel, then an additional hue value is added to the extension of the final histogram resulting in 272 bins (256+16). The similarity of the surrounding pixels indicating the concentration is revealed through the absolute difference between the energy of the mask multiplied by the central pixel value, in respect with the convolved one.

## 2.2 Color-Texture Histogram

The second color-texture-based system is a newly proposed histogram creation method which uses the  $L^*a^*b^*$  color space. The color-texture feature extracted from the  $a^*$  and  $b^*$  components is expressed through a single, 64-bin histogram.

In the first part of the method, all the images are transformed into the  $L^*a^*b^*$  color space and only the  $a^*$  and  $b^*$  components are reserved. The color-texture feature is extracted from the image by means of convolution with Laws' energy masks [3] which were selected due to their speed as they possess low computational complexity. The  $a^*$  and  $b^*$  components are convolved with the  $5 \times 5$  masks, hence producing the respective  $ENa^*$  and  $ENb^*$  components which represent the chromaticity texture energy of the image. The reason why having  $25 \times 2$  color-texture components for each image does not pose a problem is that only the components with the greatest energy are reserved. The energy volume for each image is computed by Eq. 1.

$$E_{Vol} = \sum \left( \sum_{x=1, y=1}^{m, n} ENa^*(x, y), \sum_{x=1, y=1}^{m, n} ENb^*(x, y) \right) \quad (1)$$

When the components with the greatest energy are found, the discretization process is activated, during which the two components ( $ENa^*$ ,  $ENb^*$ ) are divided into sections. The color-texture components are split up into 8 regions each, and by interlinking them a  $(8 \times 8)$  64 bin histogram is created.

## 2.3 Center-Surround Histogram

This method is based on the retinal signal processing of the Human Visual System. A center-surround operator  $C$  similar to the receptive fields of the ganglion cells of the retina is employed to create a Center-Surround Histogram (CSH) [9]. The contribution of the CSH to the proposed system is that it incorporates a degree of spatial information into the histogram in a dual manner that only the pixels around the edges are considered, and furthermore that these pixels are weighted in respect to whether they belong to a strong or weak edge. Moreover, it reduces the processed visual information by using only the colored area surrounding the zero-crossings of an image. The proposed histogram contains only the chromatic information of these areas and consists of 256 bins.

## 3 Ant Colonies

In the approach discussed in this paper we distribute the search activities over the so-called "ants," that is, agents with very simple basic capabilities which, to some extent, mimic the behavior of real ants, in order to advance research in image retrieval. The Ant Colony Optimization (ACO) heuristic has been used successfully to solve a wide variety of problems such as the traveling salesman problem [10]. The simple question arising from the usage of the ant colony optimization in the above indiscipline applications is how the ant algorithms work?

The ant algorithms are basically a colony of cooperative agents, designed to solve a particular problem. These algorithms are probabilistic in nature because they avoid the local minima entrapment and provide very good solutions close to the natural solution.

More specifically, one of the problems studied by ethnologists was to understand how almost blind animals like ants could manage to establish shortest route paths from their colony to feeding sources and back. It was found that the medium used to communicate information among individuals regarding paths, and used to decide where to go, consists of pheromone trails. A moving ant lays some pheromone (in varying quantities) on the ground, thus marking the path by a trail of this substance. While an isolated ant moves essentially at random, an ant encountering a previously laid trail can detect it and decide with high probability to follow it, thus reinforcing the trail with its own pheromone. The collective behavior that emerges is a form of autocatalytic behavior where the more the ants following a trail, the more attractive that trail becomes for being followed [10]. The process is thus characterized by a positive feedback loop, where the probability with which an ant chooses a path increases with the number of ants that previously chose the same path.

Consider for example the experimental setting shown in Fig. 1 [10]. There is a path along which ants are walking (for example from food source A to the nest E, and vice versa, see Fig. 1a). Suddenly an obstacle appears and the path is cut off. So at position B the ants walking from A to E (or at position D those walking in the opposite direction) have to decide whether to turn right or left (Fig. 1b). The choice is influenced by the intensity of the pheromone trails left by preceding ants. A higher level of pheromone on the right path gives an ant a stronger stimulus and thus a higher probability to turn right. The first ant reaching point B (or D) has the same probability to turn right or left (as there was no previous pheromone on the two alternative paths). Because path BCD is shorter than BHD, the first ant following it will reach D before the first ant following path BHD (Fig. 1c). The result is that an ant returning from E to D will find a stronger trail on path DCB, caused by the half of all the ants that by chance decided to approach the obstacle via DCBA and by the already arrived ones coming via BCD: they will therefore prefer (in probability) path DCB to path DHB. As a consequence, the number of ants following path BCD per unit of time will be higher than the number of ants following BHD. This causes the quantity of pheromone on the shorter path to grow faster than on the longer one, and therefore the probability with which any single ant chooses the path to follow is quickly biased towards the shorter one. The final result is that very quickly all ants will choose the shorter path.

Inspired by this probabilistic behavior of the real ants, ant algorithms are the software agents that coordinate by updating the information of a common memory similar to the pheromone trail of the real ants. When a number of these simple artificial agents coordinate based on the memory updating they are able to build good solutions to hard combinatorial optimization problems.

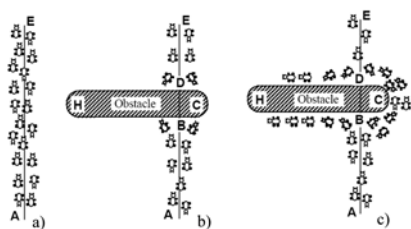


Fig. 1. An example with real ants [10]

As mentioned before, the artificial ant agents have many properties that differentiate them from the real ants and thus involve various ant algorithms based systems [10]. Along with these unique features that enhance the capabilities of the artificial agents there are other governing parameters such as the optimum number of ants, the pheromone decay rate, and the constants that make the solution to converge to the experimental results. As we are not interested in the simulation of ant colonies, but in the use of artificial ant colonies as an optimization tool in the field of image retrieval, the proposed system will have some major differences with a real (natural) one that will be discussed next. More specifically, Fig. 2 represents a generalized block diagram of the proposed ant algorithm. The problem is defined in the form of a network. All possible links between the components of the network and limiting criteria’s are identified.

### 3.1 First Stage

Following the extraction of the three descriptors described beforehand from the images in the database and considering each histogram bin to be a virtual ant, a query is posed by mobilizing a sum of 592 (272+64+256) ants. The terrain of the "ground" where the ants "walk" depends strictly on the query in a way that, through comparison with the other images, it provides the information about the relative position and distance of the surrounding "food". On this terrain, the ants move in a straight line from image to image. In this first stage of the algorithm a comparison is performed using all three features, in other words the whole population of the ants, and an initial ranking of the images takes place.

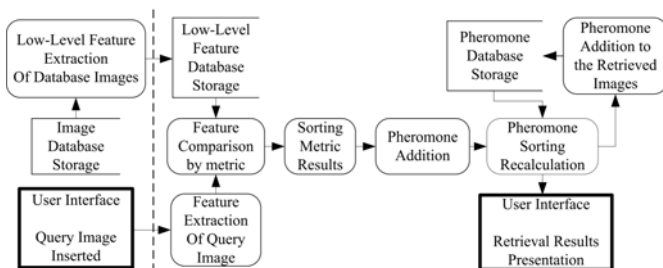


Fig. 2. Block diagram of the ant colony image retrieval system

Following numerous tests and simulations, it was concluded that the best similarity metric in order to compare the features was the Matusita distance expressed by the equation shown below:

$$M(H_Q, H_C) = \sqrt{\sum_i (\sqrt{H_Q(i)} - \sqrt{H_C(i)})^2} \quad (2)$$

where  $H_Q$  is the query histogram,  $H_C$  is the histogram to be compared and  $(i)$  is the number of bins. This distance is a separability measure which provides a reliable criterion presumably because as a function of class separability it behaves much more like probability of correct classification.

Following the initial ranking of all the images in the database, pheromone is attached to the 200 first most similar images in a descending manner according to their ranking, meaning that the highest ranked image acquires the most pheromone. The aggregate pheromone deposition from all the ants results in the creation of a pool consisting of only 200 images, thus creating a new much smaller sub-terrain, in the vicinity of which the second stage takes place.

### 3.2 Second Stage

Having completed the pre-classification of the 80000 image database and concluded to a pool of just 200, the final retrieval takes place. Taking into consideration the possibility that the initial ant search can result in false positive results, the terrain is slightly altered by substituting the initial query image with the second highest ranked image from the pool and a new query takes place. In the new query, a new group of 592 ants is mobilized and the process from the first stage is repeated, although instead of having the whole of the database as a terrain, the ants are constrained strictly in the vicinity of the pool, thus changing the amount of pheromone attached to each of the previously ranked images. In order to restrain the overall time cost and to avoid false terrain alterations caused by false positives in the first stage, this new process is repeated for four iterations meaning that the sub-terrain of the pool is altered four times and that the first four images of the initial query are used as queries themselves, thus continuously altering the pheromone deposition in accordance to the new ranking of each repetition.

The final retrieval is not based upon the value produced by the metric stating the distance between the features of the images, but on the final amount of pheromone that has accumulated on each image at the conclusion of the two stages.

## 4 Performance Evaluation

We evaluate the total performance of our system in terms of retrieval accuracy and image recall versus precision. The database used to measure the system's effectiveness and efficiency was the LabelMe database [8], which consists

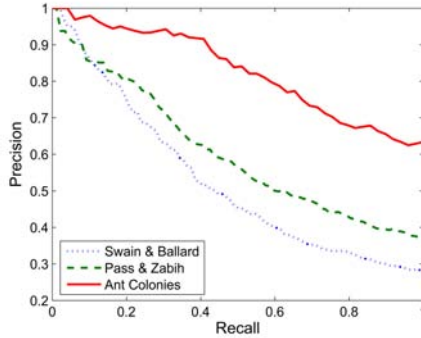
of 80000 images (and increasing); is one of the largest databases available freely on the internet and due to its immense volume is adequate for CBIR testing (<http://labelme.csail.mit.edu/>).

The retrieval outcome is presented through a query session which produces images ranked in similarity according to the pheromone laid by the ants. The larger the amount of pheromone aggregation, the higher the similarity of that specific image. The measurement used in order to evaluate the system is the retrieval performance percentage, which is the percentage of actual similar images produced in the 20 to 50 first images retrieved by the system (the percentage displayed depends on the number of existing similar images, e.g. for the first image set there are only 20 relevant images in the database; hence the accuracy presentation reaches only the first 20 images). In Table 1 one can see a synopsis of the comparisons of the method's performance for eight different image sets and notice the high percentage in accuracy, which means that most of the existing similar images were retrieved in the first 20 to 30. These image sets are characteristic of the image database and also show the worst and best precision and recall performances of the proposed system. The nature of the images varies from natural scenes (e.g. sunsets, fields, people, animals, flowers), to indoor and outdoor images of buildings as well as city images (streets, cars, parks). It should be stated here that the whole of the database is used in each query and that no preclassification is performed to enhance the performance. The time cost requested to perform a single query, in the supposition that the descriptors have been extracted a priori, is 28 seconds, which is fairly reasonable considering that the database consists of 80000 images. In addition to the precision perspective, another aspect of retrieval performance is presented by the graph in Fig. 3: precision versus recall. Precision is the proportion of relevant images retrieved (similar to the query image) in respect to the total retrieved, whereas recall is the proportion of similar images retrieved in respect to the similar images that exist. Generally, precision and recall are used together in order to point out the change of the precision in respect to recall [11]. In most typical systems the precision drops as recall increases, hence, in order for an image retrieval system to be considered effective the precision values must be higher than the same recall ones, which is mostly the case in the current system. In order to provide a sense

**Table 1.** Retrieval efficiency of the proposed system

	Total images retrieved				
	1-10	11-20	21-30	31-40	41-50
Image set 1	100%	90%			
Image Set 2	70%	60%	63%	65%	62%
Image Set 3	90%	90%	77%	65%	58%
Image Set 4	100%	90%	63%	50%	
Image Set 5	90%	55%			
Image Set 6	90%	50%			
Image Set 7	100%	95%	90%	85%	80%
Image Set 8	100%	100%	93%	90%	88%





**Fig. 3.** Average precision versus recall graph for Swain and Ballard’s Method, Pass and Zabih’s Method and the proposed method

**Table 2.** Computational cost of the three features.  $n$  is the number of pixels,  $N$  is the number of nonzero elements of each respective mask ( $M$  for spatial,  $L$  for Law’s masks and  $C$  for center-surround) and  $E$  is the number of pixels surrounding the edges.

Extraction Method	Operation	Additions	Multiplications
Spatially-Biased Histogram	1 <sup>st</sup> Histogram	$n$	0
	2 <sup>nd</sup> Histogram	$n$	0
	Convolution	$nN_M$	$nN_M$
Texture Histogram	Histogram	$2n$	0
	Convolution	$2nN_L$	$nN_L$
	Energy Volume	$2n$	0
Center-Surround Histogram	Histogram	$E$	0
	Convolution	$nN_C$	$nN_C$
Total		$(N_M + 2N_L + N_C + 6)n + E$	$(N_M + N_L + N_C)n$

of the proposed method’s performance, the average precision versus recall factor of all the queries, is compared to the respective one of Swain and Ballard’s method [2], as well as to Pass and Zabih’s joint histogram method [5]. The last aspect considered is the computational cost of the three descriptors as shortly described in Table 2. We believe that the total cost of the operations performed during the extraction of the features is not a heavy price to pay, taking into account the significant increase in accuracy produced by the presented method.

## 5 Conclusions

In this paper we have presented a new two-stage content-based image retrieval system based on the behavior of ant colonies. The entities of these ants are represented through the basic elements of three image descriptors: a newly proposed spatially-biased histogram, a second histogram acquired through utilization of the chromaticity-texture-energy of the image and last a histogram which results

through incorporation of certain attributes of the human visual system. Despite the bulk of the LabelMe database which consists of 80000 images, a precision versus recall graph for eight different image sets proves that the system exhibits an effective performance as the precision values are significantly higher compared to the respective recall ones.

**Acknowledgements.** This work is funded by the EU-EPEAEK Archimedes-II Project (2.2.3.z,subprogram 10).

## References

1. del Bimbo, A.: Visual Information Retrieval. Academic Press, London (1999)
2. Swain, M.J., Ballard, D.H.: Color indexing. *International Journal of Computer Vision* 7(1), 11–32 (1991)
3. Laws, K.: Rapid texture identification. In: *Image processing for missile guidance; Proceedings of the Seminar, San Diego, CA, July 29-August 1, 1980.* (A81-39326 18-04) Bellingham, WA, Society of Photo-Optical Instrumentation Engineers, pp. 376–380 (1980)
4. Jain, A.K., Vailaya, A.: Image retrieval using color and shape. *Pattern Recognition* 29(8), 1233–1244 (1996)
5. Pass, G., Zabih, R.: Comparing images using joint histograms. *Multimedia Systems* 7(3), 234–240 (1999)
6. Kouzas, G., Kayafas, E., Loumos, V.: Ant seeker: An algorithm for enhanced web search. In: *AIAI*. pp. 649–656 (2006)
7. Ramos, V., Muge, F., Pina, P.: Self-organized data and image retrieval as a consequence of inter-dynamic synergistic relationships in artificial ant colonies. In: *HIS*. pp. 500–512 (2002)
8. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. *MIT AI Lab Memo AIM-2005-025* 1, 1–10 (2005)
9. Panitsidis, G., Konstantinidis, K., Vonikakis, V., Andreadis, I., Gasteratos, A.: Fast image retrieval based on attributes of the human visual system. In: *7th Nordic Signal Processing Symposium (NORSIG 2006)*, Reykjavik, Iceland, Reykjavik, Iceland pp. 206–209 (2006)
10. Dorigo, M., Maniezzo, V., Colorni, A.: The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics* 26(1), 29–41 (1996)
11. Muller, H., Muller, W., Squire, D.M., Marchand-Maillet, S., Pun, T.: Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recogn. Lett.* 22(5), 593–601 (2001)