1

Automatic Segmentation and Indexing of Specialized Databases

Madirakshi Das R. Manmatha

Multimedia Indexing and Retrieval Group

Center for Intelligent Information Retrieval

Department of Computer Science

University of Massachusetts, Amherst, MA 01003

email : madirakshi.das@kodak.com, manmatha@cs.umass.edu

Abstract

The aim of this work is to index images based on color, in domain specific databases using colors computed from the object of interest only, instead of using the whole image. The main problem in this task is the segmentation of the region of interest from the background. Viewing segmentation as a figure/ground segregation problem leads to a new approach - successful elimination of the background leaves the figure or object of interest. The background elements are eliminated using general observations true for any photograph where there is a single, prominent object of interest. First, we form a hypothesis about possible background colors and eliminate them, using an iterative algorithm which allows for backtracking in the event of erroneous selection of background colors. We then use an edge image at an appropriate scale to eliminate those parts of the image that are not in focus and do not contain contain significant structures. The edge information is combined with

*This work was supported in part by the Center for Intelligent Information Retrieval and in part by the National Science Foundation under grant number IRI-9619117. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsor(s).

the color-based background elimination to produce object (figure) regions. We test our approach on a database of bird images. We show that in 87% of 600 bird images tested, the segmentation is sufficient to determine the colors of the bird correctly for retrieval purposes. These colors are used for indexing in a retrieval system. Retrieval experiments on a database of 1200 birds using known-item-search - where the goal is to retrieve a known target image given a query showing the same bird - show substantial improvements can be achieved by using our object-of-interest finder. The results show good retrieval performance even when the target images were significantly different from the query image in terms of both variations in the pose of the bird and background content. We also show that our framework can be used to incorporate domain-specific knowledge, resulting in correct segmentation of flower regions from a database of 1000 flower images in 86% *of the images.*

Keywords: *Content-based image retrieval, image segmentation, color histogram*

1 Introduction

Most existing image retrieval systems cast the retrieval problem as the users' need to find other images *similar* to a given query from an image database, where similarity is computed using a distance metric in a low-level image feature space. However, similarity is a semantic notion, not necessarily captured by low-level image features computed from the global image. This has lead to a great deal of interest in the problem of *meaningful* retrieval from image databases in recent years.

The basic step towards meaningful retrieval is to ensure that the image descriptions used to index the database are related to the semantic content of the image. This requirement is difficult to meet in the context of content-based image retrieval. Unlike text where the natural unit, the word, has a semantic meaning, the pixel which is the natural unit in an image, has no semantic interpretation by itself. In images, meaning is found in objects and their relationships.

There is a growing number of large image databases which are dedicated to specific types and subjects. Examples include mug shots of human faces, pictures of flowers and birds. These databases are characterized by images which portray a single object which can be clearly identified by a human user. The challenge is to extract the object of interest automatically and index the database based on features gathered from the object only, thus eliminating the effect of the background. This work is motivated by the need for such an object-of-interest finder as a preprocessor to any indexing and retrieval system which will be working on a database of images with clearly defined subjects.

In a database of images with a well-defined subject, in most cases, the intention of a user query is to find other images with the same subject. For example, in a database of images of birds, a query showing a bird flying against a blue sky should be able to retrieve images of the same or similar birds sitting on a branch or flying against a cloudy sky. Instead, current image retrieval methods which are based on low-level image features like color and texture derived from the whole image, would retrieve other images dominated by the blue color. To accomplish meaningful retrieval in this scenario, we need to ensure that the subject is the only part of the image used to generate the database indices, ignoring the background content.

Segmentation of an image into its constituent objects is a very difficult and ill-defined problem. In addition, meaningful indexing also requires the ability to discriminate between foreground object(s) and background elements. In current retrieval systems, the background can play an undesirably important role in the retrieval, even in restricted domains. For example, a picture of a flower against a background of green leaves may not be able to retrieve images of the same flower against a background of soil or in a close-up without any background. This is because the query contains green areas which are given equal importance as the flower regions. The presence of backgrounds is a major problem which needs to be handled intelligently before retrieval can be effective.

Segmentation is a hard problem and it would be very difficult to train classifiers to detect objects which are as varied in color, shape, size and viewpoint as (for example) pictures of birds. The problem may, however be viewed from a new perspective. We observe that photographers often try to ensure that the subject of interest is "prominent" and that the background is less prominent. This is usually done by placing the subject close to the center of the image, by making the subject of interest larger than other objects in the image and by having the subject in sharper focus than the background. Many image databases like pictures of birds (Fig. 1), or flowers, or other animals often have these characteristics.

We use the above characteristics of pictures to propose an approach to automatic segmentation for finding the figure or subject of interest The procedure involves eliminating the background. What is left is assumed to be the figure or object of interest. The algorithm involves three stages. First, some color(s) are hypothesized to be background color(s) based on their probability of occurrence at the borders of the image. The hypothesis is then tested by eliminating those color(s) and evaluating the remaining image. The remaining image after elimination of detected background colors is combined with information from an edge description of the image at an appropriate scale which captures the major structures present in the parts of the image that are in focus. The final result is a segment containing the object (figure) region (see Fig. 8 for examples). We would like to emphasize that significant fractions of the subject (the bird in this picture) may often lie at the borders of the image (Fig. 7), so simply eliminating all the border colors will not work.

Though this work is related to the problem of image segmentation, there are notable differences because of the difference in the final goals. In our case, the primary goal is to identify a region in the image which will produce features derived from the subject only, enabling image indexing and retrieval based on the subject of the image. This does not require perfect segmentation of the subject, as long as the region considered is predominantly covered by the subject. The final segment may have small parts of the object missing or include small areas from the background without much impact on retrieval performance. Where there are multiple copies of the same object in an image, it is sufficient to detect a single, prominent instance of the object. Also, we are not interested in segmenting the background correctly - for example, sky and foliage in the background can all be treated as one "background" mass.

The general framework for background elimination makes it easy to include any domain specific knowledge available for the subject portrayed in the specialized database. In earlier work [3, 4] a solution to the problem of object-of-interest identification in a database of flower images was provided. In that case, domain knowledge about the color of flowers (e.g. flowers are rarely gray, brown, black or green) was used to simplify the problem of segmentation. In fact these additional domain constraints are sufficiently strong that in the case of the flowers the color information is sufficient and the edge based description is not required. We show that in a database of 1000

Figure 1: Some images in the bird database

flower images, a flower region can be segmented in 86% of the images.

The results of our work are illustrated using examples from a database of images of birds. The bird database, unlike the flower database, has no particular domain specific knowledge that can be exploited. The problem is made more difficult by the fact that most birds are designed to merge into their natural backgrounds to avoid detection by predators, unlike flowers which are designed to stand out against their background. These images were downloaded from the world wide web and show wide variations in the type of background (water, sky, ground, man-made surroundings) as well as the size of the object of interest as shown in Fig.1. We show that in 87% of 600 bird images tested, the segmentation is sufficient to determine the colors of the bird correctly for retrieval purposes. These colors are used for indexing in a retrieval system. Retrieval experiments on a database of 1200 birds using known-item-search - where the goal is to retrieve a known target image given a query showing the same bird - show substantial improvements can be achieved by using our object-of-interest finder. The results show good retrieval performance even when the target images were significantly different from the query image in terms of both variations in the pose of the bird and background content.

This paper is organized as follows : section 2 discusses related work, section 3 outlines the detection and elimination of background colors based on a combination of color analysis and edge information. Section 4 discusses experimental results on segmentation and subsequent indexing and retrieval, with section 5 containing concluding remarks.

2 Related Work

Image retrieval has been an active area of research since the early '90s. The initial focus in this area was to develop suitable low-level features to describe the semantic content of images, analogous to words in language. Color [33], texture [17], shape [22] and filter response-based features [28] have been used as attributes for indexing images for content-based retrieval. A recent survey of techniques used in content-based image retrieval [32] provides a good overview of the approaches that have been investigated over the last ten years. Recent papers have focused on building image retrieval systems which use combinations of features and address actual applications like searching for images on the world wide web. A survey of content-based image retrieval systems [30] lists many such end-to-end systems, many of which are available for trial online.

Color is a commonly used low-level feature when the database images are in color. It is useful for indexing objects which have distinctive colors signatures, for example, commercial products, flags, postal stamps, birds, fishes and flowers, or as a first pass for other colored images. Swain and Ballard [33] proposed the use of color histograms to index color images and described an efficient histogram intersection technique for matching. Normalized color histograms along with histogram intersection have been popular for indexing color images because of the fast speed of matching and the fact that they are generally invariant to translation, rotation and scale. However, since color histograms do not incorporate information on the spatial configuration of the color pixels, there are usually many false matches where the image contains similar colors in different configurations.

A few researchers have attempted to include this information in the representation to improve the retrieval results. Zabih et al [12] have proposed the color correlogram which includes information on the spatial correlation of pairs of colors in addition to the color distribution in the image. Matas et al [21] have described a color adjacency graph which can be used to describe multi-colored objects, but the matching phase is too computationally intensive for use in large image databases. An efficient indexing strategy using a hybrid graph representation of color adjacencies in an image is proposed by Park et al [25]. Das et al [6] have proposed a simpler spatial adjacency graph structure which is used in a filtering phase to enforce the spatial properties of the colors required by the query image. Spatial color distribution information has been used for indexing trademark images in [14].

For retrieval systems that work with general databases like generic stock photographs and mixed news photographs, it is not clear a priori which feature (or combination of features) would produce better retrieval performance. This depends on the type of object or scene depicted in the query. Many such systems implement a wide variety of features and let the user choose the important aspects of the query at query time. An example of a system which implements color, texture and shape is *QBIC* [23] which allows queries based on example images, sketches or selected color and texture patterns. The user can select the features to be used as well as the relative importance to be attached to each feature in the final ranking. *Virage* [1] is another general purpose retrieval system which provides an open framework to allow general features like color, shape and texture as well as very domain specific features to be used as plug-ins. The *Photobook* [27] retrieval system uses shape, texture and eigenimages as features in addition to textual annotations. The system can be trained to work on specific classes of images. The *SIMPLIcity* system described in [35] uses a wavelet-based approach for feature extraction, combined with integrated region matching. The regions in the image are characterized by their color, texture, shape and location. Other examples of existing systems using multiple features and multiple query modes are *Candid* [16] and *Chabot* [24].

There is a need for automatic retrieval solutions in a number of specialized domains which are currently indexed by manual annotations and specialized codes which involve extensive, tedious human involvement. In many of these specialized domains, features specific to the domain need to be formulated to produce good retrieval results. For example, Pentland et al [26] describe the eigenimage representation which measures the similarity in appearance of faces which is used to search for similar faces in the Photobook system. Even when the domain has a wide variety of images (for example trademarks), the application may be specialized. For example, for trademark retrieval, Ravela and Manmatha [29] have used a global similarity measure for images based on curvature and phase to produce superior results on a database of trademark images when compared to general-purpose shape-based approaches. Eakins et al [7] have developed a trademark retrieval system (named ARTISAN) which uses Gestalt theory to group low-level elements like lines and curves into perceptual units which describe the trademark. In addition to developing appropriate features for specialized databases, one may be able to segment and describe the objects depicted in the image using knowledge about the objects to simplify the segmentation process. Forsyth and Fleck [10] describe a representation for animals as an assembly of almost cylindrical parts. On a database of images of animals, their representation can retrieve images of horses, for example, in a variety of poses. Fleck et al [8] use knowledge about the positions of attachment of limbs and head to the human body to detect the presence of naked people in the database images. Forsyth et al illustrate some specialized applications of image retrieval in [9].

There has been a lot of work in the area of image segmentation. Recent work has focused on combination of different cues like color, texture and edges for segmentation [2, 20, 19]. Relational graph matching has been used for segmenting natural images in [31] However, these techniques produce segments which may not necessarily correspond to single objects in the scene and also, there is no way of discriminating foreground and background elements. Automatic foreground/background disambiguation based on multiple features like color, intensity and edge information has been studied in [13], but these techniques work well on relatively smooth backgrounds and objects with sufficient contrast. Recently proposed techniques for detecting natural shapes in real images [18] also work best with simple backgrounds.

3 Detection and elimination of background

Our approach to elimination of background based on color entails the generation of a hypothesis identifying the background color(s), elimination of those colors and checking the remaining image for the presence of a valid segment. The check provides a feedback mechanism for background elimination which indicates whether the hypothesis was correct or a new one needs to be formulated. The image remaining after elimination of detected background colors is combined with information from an edge description of the image, which captures the major structures present in the parts of the image that are in focus. The final result is a segment containing the object (figure) region.

3.1 Observations about photographs

The specific observations we exploit are derived from general rules-of-thumb followed when photographing a subject. Since no domain-specific assumptions are made, these observations are true of most images with clearly defined subjects. The subject is usually centered in the middle three-quarters of the image (defined as the "central region" in Figure 4) and occupies a reasonable portion of the image. When photographing a specific subject, there is usually an attempt to keep other competing foci-of-interest out of the picture. For example, the subject is often in sharper focus than the background. Further, a picture of a parrot and a sparrow has two subjects, unless one is clearly larger and more in focus than the other. In such cases, we assume that the larger region is more significant and ignore smaller regions.

Based on these observations, we know *a priori* that we are looking for a segment in the image which is large enough, is centered somewhere in the central region of the image and has prominent edges, since it is in focus. Conversely, the background regions surround the main subject and thus, are more likely to be visible along the periphery of the image. If the background is out-of-focus, there may not be significant edge information detected in that region. However, none of these observations are true in *all* cases. In such cases, it may not be possible to discriminate between the foreground and background of the image in the absence of additional constraints. The design of our algorithm takes this possibility into account, and produces no segmentation where good subject extraction is not possible based on the color and edge information gathered from the image. In the context of image retrieval, this would mean that the whole image is used for indexing, which is the starting point we are trying to improve on.

3.2 Segmentation strategy

The outline of the algorithm used to produce a segment from which the color of the bird can be estimated is shown in Figure 2. The elimination of background color is described in this sub-section and the incorporation of edge information is discussed in the next sub-section.

The first step in producing a list of possible background colors is to select a suitable color space to label the image pixels. The RGB space in which the original image is described, has too many colors to be useful. Instead, we use the colors defined by the X Window system which has only 359 colors and is also perceptually grouped into visually distinct colors. Since the mapping from the RGB space to X Color names is sparse, for points with no exact map the nearest color name (by city block distance) is used to map the point to a color defined in X. This mapping both reduces the number of colors and also ensures that small variations in the color of an object are classified as the same perceptual color.

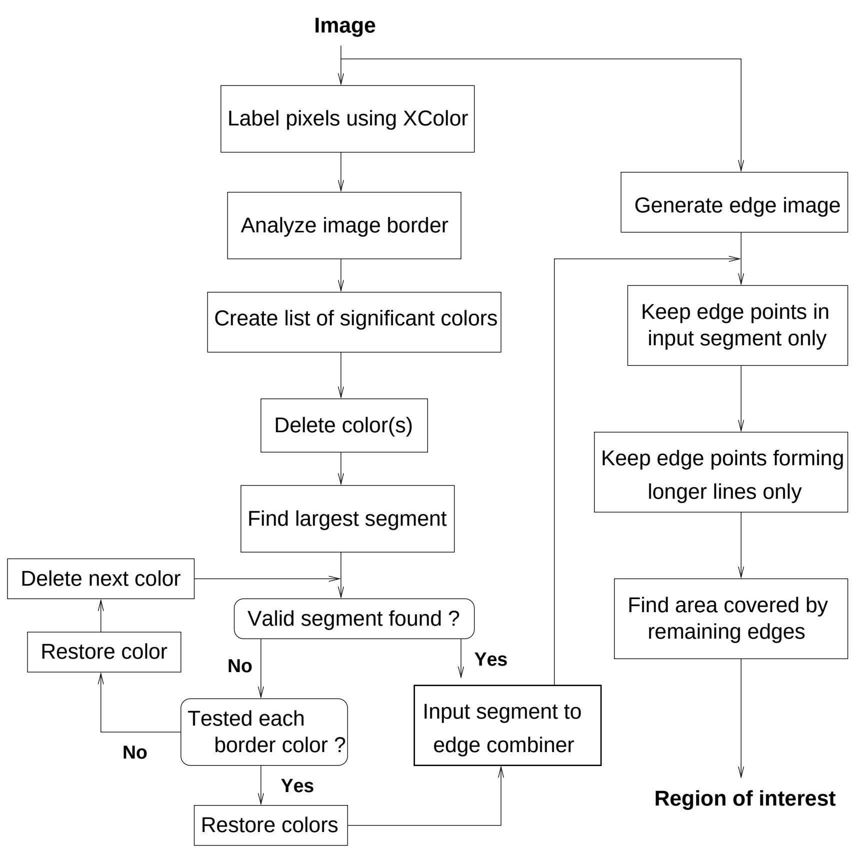


Figure 2: Overview of segmentation strategy

color	no. of blocks
SteelBlue3	**14**
olive green	**10**
cornflower blue	**8**
SteelBlue2	**4**
DarkSeaGreen4	**4**

(a) (b) (c)

Figure 3: Background elimination : (a) original image (b) significant colors detected along image periphery (c) image left after deleting colors in *(b)* found along the image periphery

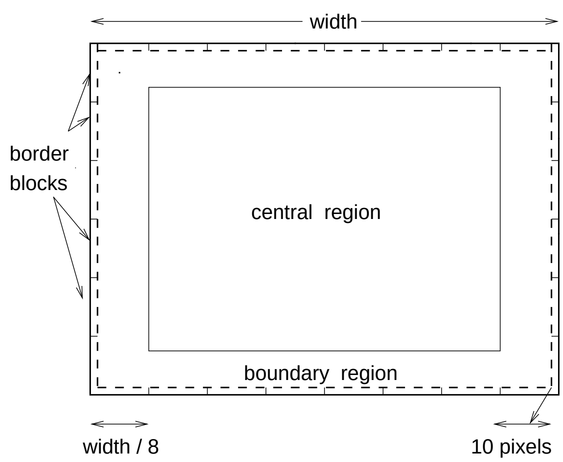


Figure 4: Definitions of image regions : Border blocks (shown in alternating color), central region and boundary region

The presence of background colors is detected by analyzing the color composition of the image margins. The margins of the image are divided into border blocks which are narrow rectangles as shown in Figure 4. We use the complete image periphery (all four sides) and divide the periphery into 24 equal border blocks. The distribution of X colors in these blocks is computed, and colors present in significant proportion in more than one border block are marked as possible background colors. Eliminating pixels belonging to the hypothesized background color(s) significantly reduces the background elements in most images, and can produce a residual image containing the foreground object only, in the simpler cases. Figure 3 shows an example of the elimination of background by deleting pixels of colors detected along the periphery.

After eliminating all the pixels of the hypothesized background color(s), the largest segment in the remaining image is computed. We use the *connected components* algorithm for identifying segments in the image, where each segment is a connected component. The connected components algorithm is run after *binarizing* the image, where the only two classes are pixels which have been eliminated and those that remain. Figure 5 shows examples of the largest segment obtained when the colors detected along the periphery are deleted after being identified as background colors. The largest segment obtained closely matches the bird region of the image in these images, (which are relatively simple).

We use two criteria for evaluating whether the segment produced is valid; its size and the location of its centroid. As discussed in the previous sub-section, the segment cannot be a possible candidate for the subject of the image if it is too small or if its centroid falls in the boundary region of the image (as defined in Figure 4). Examples of segments that are correctly flagged as invalid are shown in Figure 6. A lack of valid segments after elimination of the hypothesized background colors, is an indicator that the background color selection was wrong.

When there is feedback that the background color chosen was incorrect, the color(s) is restored and each color present in the image periphery is tested separately as a potential background color. If no valid segments are found when any of the colors present in the border are eliminated, we can conclude that the bird and the background cannot be differentiated based on color, and the whole image is output as the segment of interest. This happens when the background color and the color of the bird match. Figure 12 show two examples of this case, which is not uncommon in this database because many birds depend on camouflage to remain undetected.

In some images, background color deletion is sufficient to produce a good segmentation of the bird from the background as shown in Figure 5. In most images, however, the output can be further improved by additional processing as described in the next section.

3.3 Using edge information

It is not always possible to extract a segment containing only the bird on the basis of differentiation of background and bird colors. In addition to images where the color of the bird closely matches the background colors, there are images where background colors remain because they were not

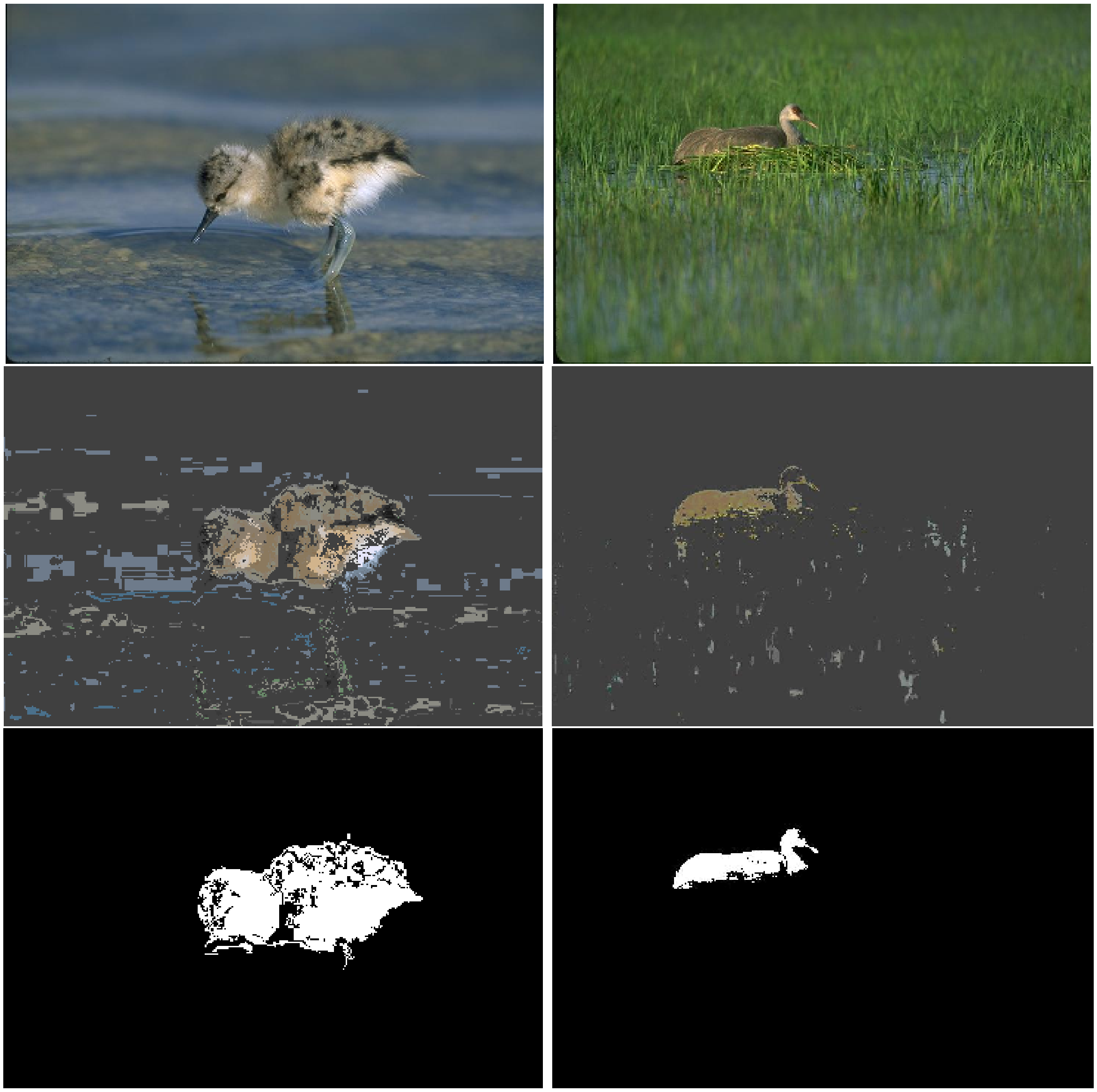


Figure 5: Examples showing extraction of bird segment obtained where the background color elimination step is very effective : (row 1) original images (row 2) image after deleting background colors (row 3) largest segment produced

present along the image periphery, and therefore, were not detected by the background elimination process. Edge information can be used in many cases to refine the segmentation.

There are differences between edges associated with the outline of the bird and edges contributed by other parts of the image. The edges associated with the background are usually present only at smaller scales. This is due to several reasons:

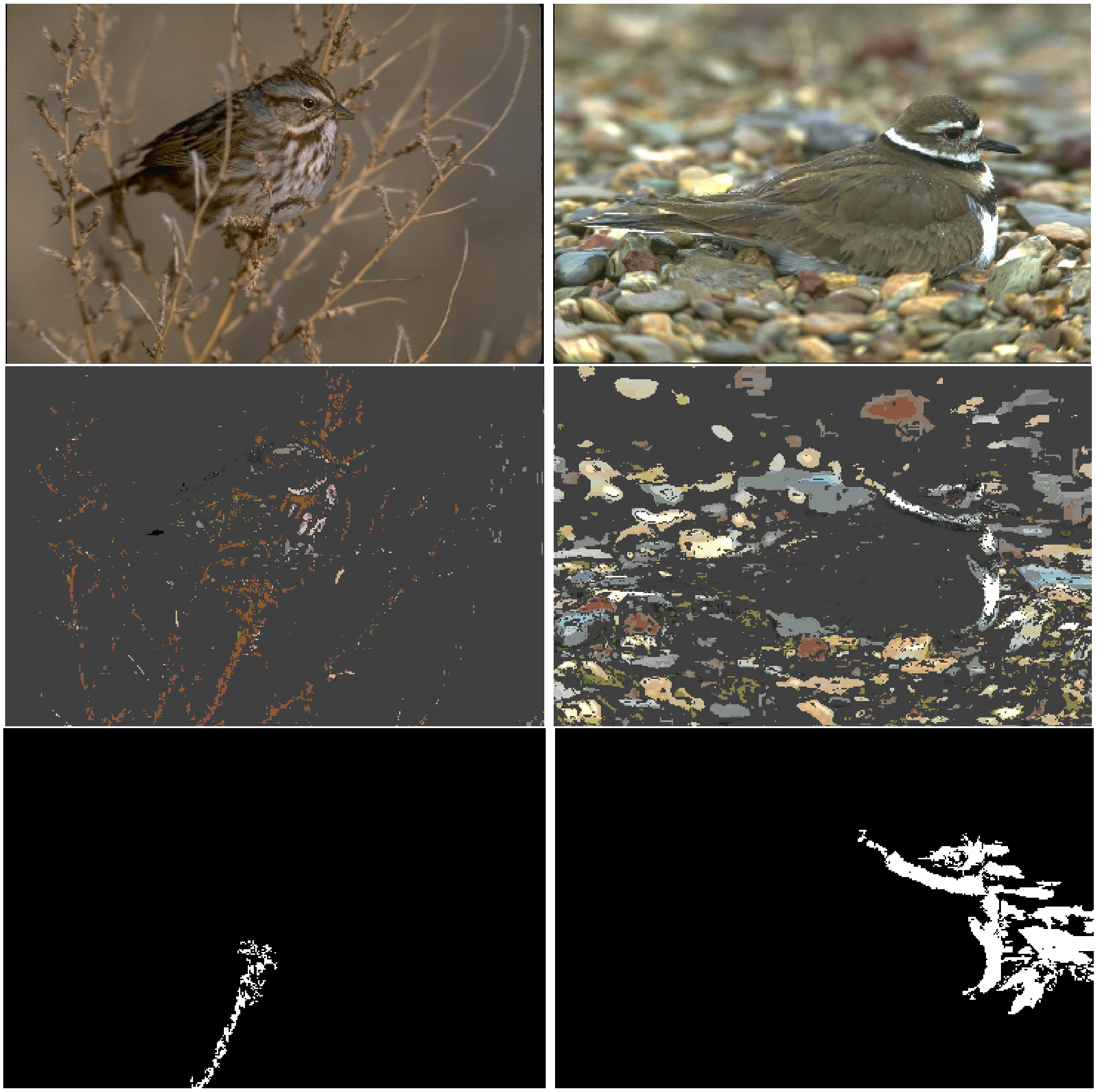


Figure 6: Examples showing detection of invalid segments : (top) original images (mid) after deletion of hypothesized background colors (bottom) largest segments produced (invalid since too small (left) or centroid is in the image boundary region (right))

- The background often consists of uniform regions such as sky in which edges, if any, appear only at the smallest (finest) scales.

- The background may often be blurred (for example, the top left images in Figure 7 and Figure 8) because of the limited depth of focus of cameras, an effect that is often accentuated by the photographer. In this case too, there are no strong edges at larger scales.

- Many backgrounds associated with bird images consist of textured surfaces such as grass, mud, water or trees. The scale of such textures is usually much smaller than that of the bird.

In contrast, the bird is usually large and distinctive in the image. Thus, the edges associated with it are usually present at a wide range of scales. It is to be noted that the edge structure of the internal feathers of the bird is often present only at small scales. However, this does not matter for our purposes since we are only interested in the external contour of the bird.

These effects can be taken advantage of in eliminating background regions by using a relatively larger scale for detecting edges. Thus, only edges present in the bird's contour would be detected. The main steps in computing an edge image are listed below :

- The image is convolved with the two first derivatives of a Gaussian [11] to include both vertical and horizontal edge directions. The derivatives of Gaussians are energy normalized (by dividing by the scale).

- The derivative outputs are combined to produce the gradient magnitudes. The energy normalization ensures that the range of the gradient magnitude images is roughly the same at all scales.

- The output of the image is then thresholded to find edges. We have found that a scale of $\sigma = 2$ and a threshold of 15 works for all our images.

The third row in Figure 7 shows the output of the edge detector on the bird images in the first row. Note that large portions of the background do not have any edges present while the edges on the bird are still present. It is clear from the image on the right side that the edge image alone is insufficient for eliminating the entire background and the combination of edge and color information provides improved background elimination.

3.4 Generation of final region of interest

At this point in processing, the edge image and the segment of interest output by the color-based background elimination process are available (images shown in the second and third rows of Figure

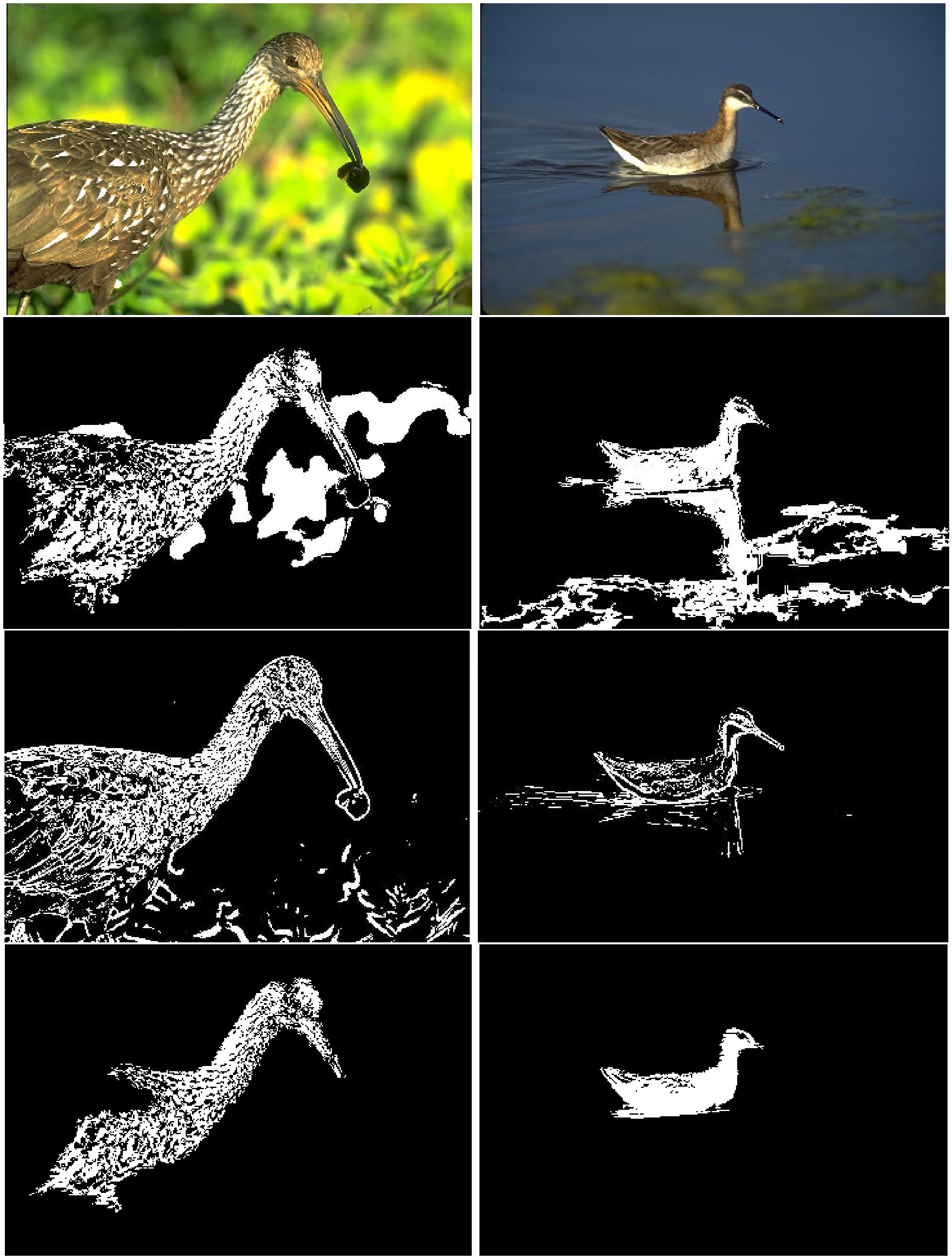


Figure 7: Examples showing improvements in the bird segment extracted when edge information is incorporated : (row 1) original images (row 2) largest segment after background color deletion (row 3) edge image (row 4) final output

Figure 8: Examples of region of interest segmentation : (row 1) original images (row 2) remaining image after background colors are eliminated (row 3) edge image (row 4) final segment obtained

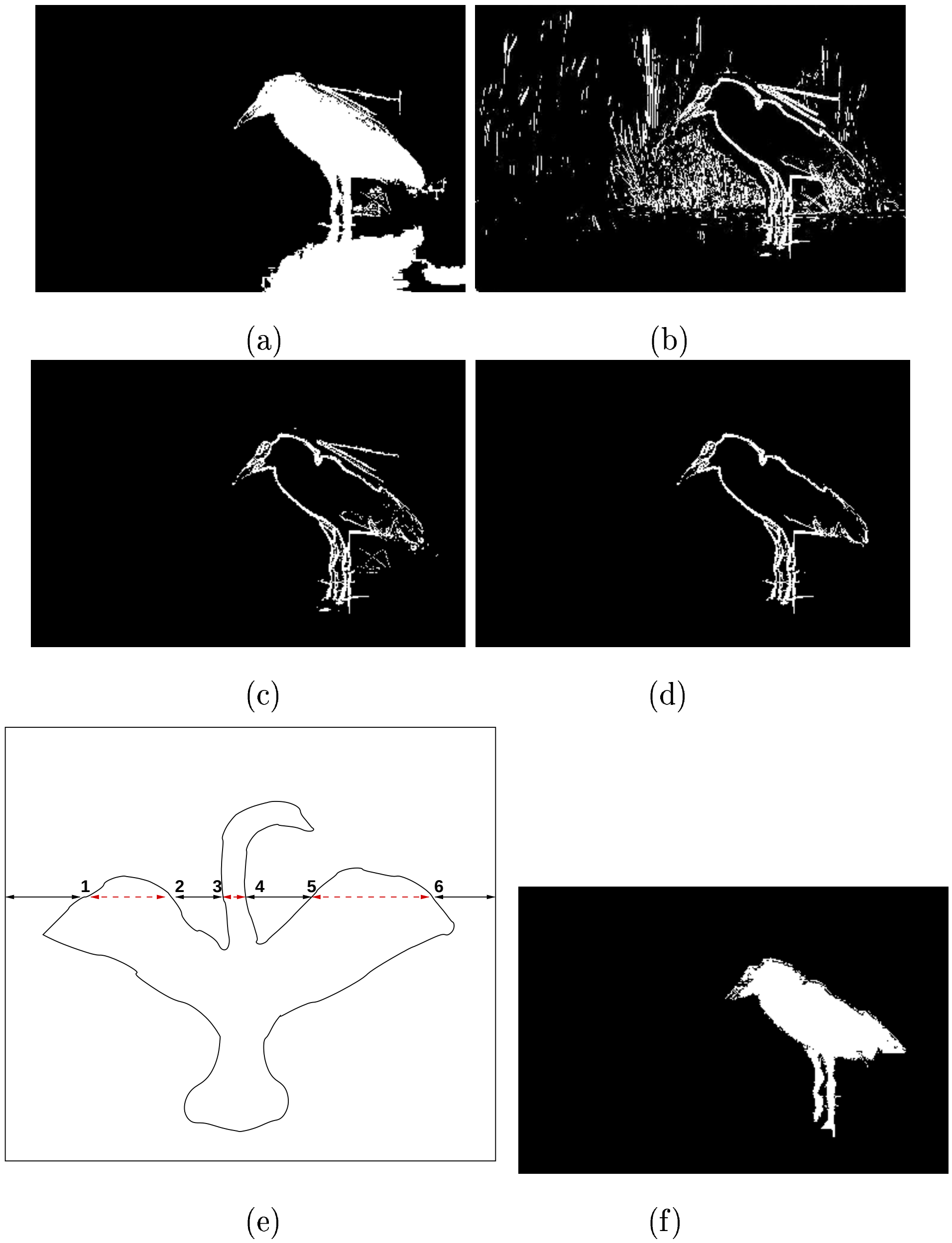


Figure 9: Combination of color-based segment of interest with edge information : (a) region of interest output from color-based segmentation (b) edge image (c) edge image left after deleting pixels which do not overlap with *(a)* (d) remaining edge image after small edge segments have been removed (e) filling the contour using edge crossings (numbered 1 to 6) on a scan line. Note that the parts of the line between an odd and even crossing are within the object, and the segments between an even and odd crossing are outside the object (f) final region of interest obtained

7). The segments of interest based on color and the edge information present in the edge image need to be combined to arrive at a final region of interest. The combination process places higher confidence on the color-based background removal, since the edge-based background elimination is effective only when the conditions described in the previous sub-section are satisfied. Often, edges from large structures in the background or smaller structures close to the bird (and therefore, in sharp focus) are present in the edge image generated. The main steps involved in this process are listed below and illustrated using the bird image to the right in Figure 8. Figure 9(a) shows the largest segment obtained from the remaining image after background color elimination (the same picture as in the second row of Figure 8).

- The edge pixels that are not included in the color-based segment of interest are eliminated. For example, when the edge image (Figure 9(b)) is filtered using the color-based segment of interest (Figure 9(a)), edge pixels shown in Figure 9(c) remain. This should eliminate most of the edges from the background.

- The next step finds connected components linking edge pixels into edge segments in the remaining edge image. Small and isolated edge segments are eliminated (edge segments containing less than 20% of the total number of edge pixels remaining are considered to be too small). This process leaves the longer edge segments only as shown in Figure 9 (d).

To estimate the area covered by these remaining edge lines, a closed contour is assumed and a commonly used technique from computer graphics is used to determine the inside/outside relationship [34]. The image is processed one scanline at a time and the region between the odd and even edge crossings on each scanline is included in the final output segment which represents the object of interest (bird) in the image. Figure 9(e) shows that the segment between the odd and even crossings represents the part of the scanline inside the object. The scanlines containing only one edge crossing are ignored, these occur when there are pieces of the background remaining or when the bird contour is incomplete. A reasonable bird region will be obtained even when some scanlines are missed if the contour of the bird is mostly detected correctly. This process results in the final bird segment shown in Figure 9(f). Some examples where the edge information is able to improve the segmentation produced by color-based foreground-background discrimination are shown in Figure 7 and Figure 8.

4 Experimental results

The test database used for this work consists of 1200 images of birds downloaded from the world wide web. These images vary widely in quality – the resolution of the images varies from barely acceptable to very high and the photographs themselves range from professionally taken to clearly flawed. The image sizes range from 12Kb to 40Kb. There is a wide variation in the type of background (water, sky, ground, man-made surroundings) as well as the size of the object of interest. Some examples of the database images can be seen in Figure 1.

4.1 Results of automatic segmentation of region of interest

The automatic segmentation results were manually verified [1] and divided into five classes as follows.

1. *No background remaining* : This class consists of images where the background is totally eliminated and the region of interest includes the bird region only. Some examples of this case are shown in Figure 8. In most of these images, the color of the background is different from the bird or the background is sufficiently blurred.

2. *Insignificant background remaining* : In these images, the greater part of the background is eliminated and the remaining background is not large enough to alter the color distribution of the final segment significantly. An example of this case is shown in Figure 10. In most cases, the background included consists of the object (branch, rock) on which the bird is resting, since this object is in as sharp a focus as the bird, and is often of the same color (to provide the bird with camouflage).

3. *Image unchanged* : In images where the bird is well camouflaged, it is not possible to extract the foreground on the basis of color or edge information. In such cases, the whole image is used for indexing. Figure 12 shows some examples of this case. Note that the background contains the same colors as the bird and is very cluttered (so the edge image provides no useful discrimination).

[1] The manual verification was done on half (600) of the total number of images because of the labor intensive nature of the checking process.

4. *Significant background remaining* : In this class of images, a significant amount of background remains in the final segment so that the color distribution computed for the bird is not accurate. Examples of such images are shown in Figure 11. The region of interest produced includes some large and prominent object(s) from the background in addition to the bird region.

5. *Incorrect segmentation* : Figure 13 shows two cases where the segmentation algorithm failed; the bird was eliminated altogether and the output consists of parts of the background. This happens when the main background color matched that of the bird, but there were other background colors in the central region of the image occupying a significant area.

Figure 10: Example showing partial elimination of background where the included background does not affect the color distribution of the final segment significantly.

No background remaining	57%
Insignificant background remaining	13%
Image unchanged	17%
Significant background remaining	11%
Incorrect segmentation	2%

Table 1: Results of automatic segmentation on bird images

The percentages of images falling under each of the above classes is listed in Table 1. Since segmentation acts as a pre-processing step to indexing and retrieval, the success of the segmenta-

Figure 11: Examples showing partial elimination of background where the included background does affect the color distribution of the final segment.

Figure 12: Examples showing cases where a valid bird segment could not be extracted based on color

tion can be judged by the impact it has on the image retrieval problem. The following discussion lists the implication of the segmentation obtained for image retrieval.

1. The first class of images (no background remaining) is the ideal case, where only the colors

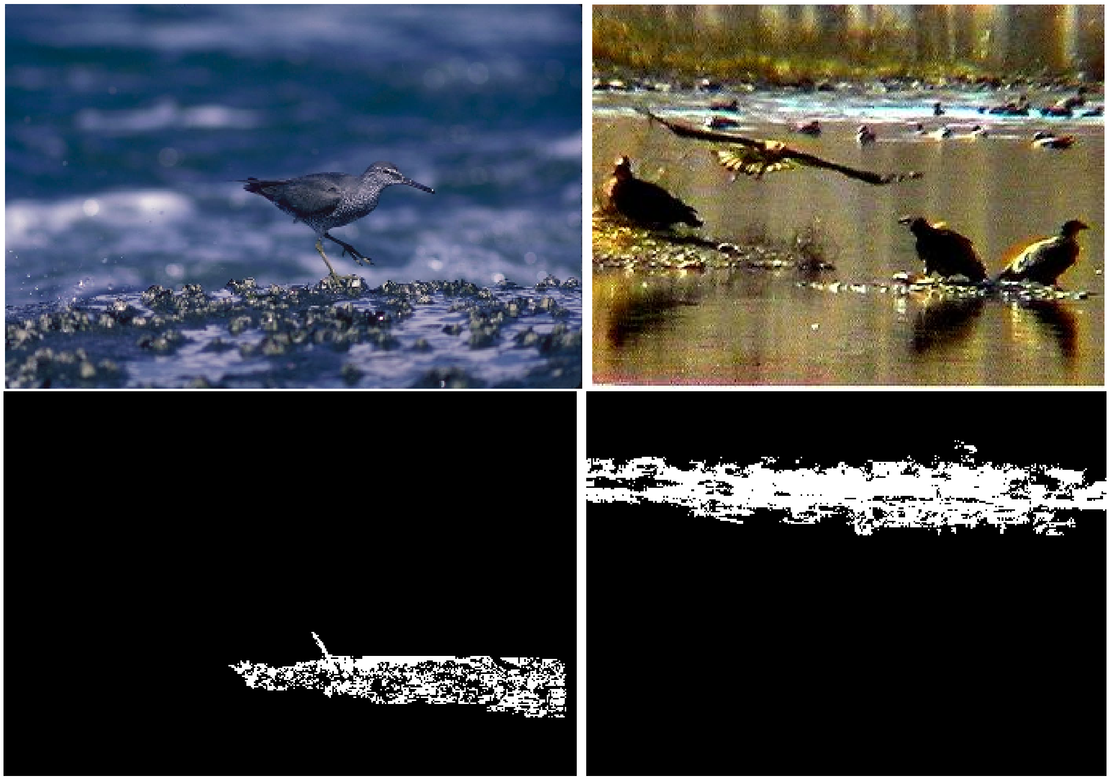


Figure 13: Examples showing failure cases where the bird segment was deleted : (top) original images (bottom) final segment obtained

of the bird are used for indexing. In this case, the retrieval results are based on the color of the bird alone, and should closely match the user's expectations.

2. The second class of images (insignificant background remaining) is indistinguishable from the first class in terms of image retrieval, since the colors indexed are predominantly from the bird regions. The small amount of pixels contributed by the background does not play a significant part in the retrieval.

3. Though it seems obvious that the cases where no segmentation could be achieved should result in poor performance, just the opposite is true. The retrieval performance in this case is very close to the ideal case where no background is present. This is due to the fact that in these images (Figure 12) the background colors match the bird very closely, and including these colors does not make a significant difference to the indexing process, though they may change the proportion of colors to some degree,

4. When significant background remains (as in class 4), the images are indexed by colors from some background elements in addition to the bird regions, and this results in degraded retrieval performance (though still an improvement on using the whole image. For example, in Figure 10, with the elimination of the blue sky which dominates the image, the emphasis placed on the colors from the bird is increased (since they now occupy a larger portion of the image).

5. The last class (incorrect segmentation) represents the true failures of our approach. The user would have better results using whole image indexing in these cases. In fact, since the bird region is excluded from the final segment, it is ensured that the user will not find any birds with matching colors in the retrieval results. and the output consists of parts of the background. However, this problem was encountered in a very small proportion of the images. In most cases where the bird was indistinguishable from the background, the segmentation algorithm was able to detect this situation, and output the image without segmenting it.

The overall results suggests that indexing based on the color of the bird is achieved in 87% of the images. In 11% of the images, the indexing includes some background colors, and in 2% of the images the colors of the bird were absent in the index.

Since the proposed foreground segment detection method does not use information specific to birds, it can be used without alteration on other images with single subjects with good results. Figure 14 show an example of other subjects (snake and butterfly) extracted correctly (in the case of the snake the segment extracted is sufficient to determine its color) [2].

4.2 Segmentation using additional domain knowledge

The framework for figure/ground segmentation suggested above can also be used to exploit additional domain information. For a specialized domain such as flowers, where there is color-based domain knowledge available about the subject, this domain knowledge can be used in the automatic extraction of the flower region from the rest of the image. In this case, the use of domain

[2] These images were accidentally introduced in the database by the web crawler designed to collect bird images

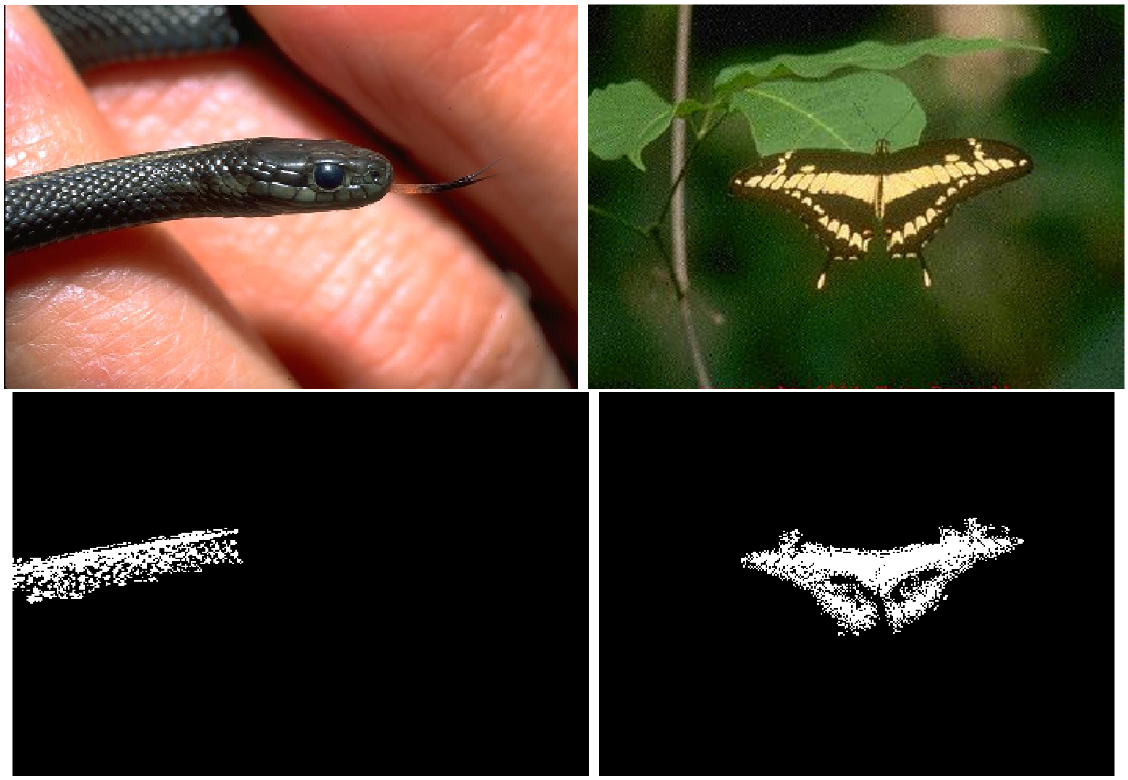


Figure 14: Examples showing correct detection of subject in other domains (top) original images (bottom) final segment obtained

knowledge eliminates background sufficiently, making the edge-based background elimination step unnecessary. This work is described in [3, 4].

Color-based domain knowledge takes the form of natural language facts about the color of the object of interest. In this case, we know that flowers are rarely *green*, *black*, *gray* or *brown*. This domain knowledge is true of most insect-pollinated flowers (most of the decorative flowers of value fall under this category). These flowers depend on insects for pollination and therefore, have evolved to be attractive to insects by producing nectar or perfumes. The flowers are also designed to be easily distinguishable from the background so that insects can locate them. Since the background usually contains green (leaves), brown (soil), gray and black (shadows) color regions, these colors are avoided to make the flower stand out from the background.

Since color-based domain knowledge is available in terms of natural language color descriptions, the color space needs to be mapped to commonly used color names. We use the ISCC-NBS color system [15] which produces a dense map from the Munsell color space to names. The naming system uses a standard set of 27 base hues (e.g. red, green, bluish green etc) and generates 267 color names using hue modifiers (e.g. light, very dark etc). These color names can be easily decomposed into a hierarchy of colors where we may use the full color name, partial names, base hues or coarser classes comprising groups of base hues (e.g. bluish green, yellowish green and green could all be merged into the class "green"). More details are available in [3, 4, 5].

The results of the proposed automatic segmentation algorithm were evaluated by viewing the output segments produced on images from a test database of 1000 flower images downloaded from the world-wide-web, and comparing them against the original image which provides the ground truth.

Correct segmentation	86%
Some background	4%
Wrong segmentation	10%

Table 2: Results of automatic segmentation on 1000 flower images

Table 2 shows the tabulated results for the test database. The most common cause for the erroneous segmentation, accounting for 60% of the errors, was that the flowers present in the images were too small. Figure 15 shows examples of automatically segmented flower regions from images. The final segments obtained (shown on the right) were solely from flower regions, and no background was included. It should be pointed out that a complete flower need not be present in the final segment for the segmentation to produce an accurate description of flower color. The examples illustrate the wide variety of backgrounds and the large variations in the area covered by flower regions that the automatic segmentation algorithm can handle.

A retrieval system was built by indexing the colors of the flowers. The reader is referred to [3, 4, 5] for details.

4.3 Results of indexing and retrieval

A system for retrieving birds similar in color to a query image of a bird was built. The system indexes the colors of the bird as determined by the region-of-interest algorithm. The database of bird images is indexed using color histograms [33] (in the X color name space) generated from

Figure 15: Some examples of images where a correct flower segment was obtained by the iterative segmentation algorithm. Note that it is not necessary to detect *all* the flower regions present in an image to produce an accurate description of the flower color

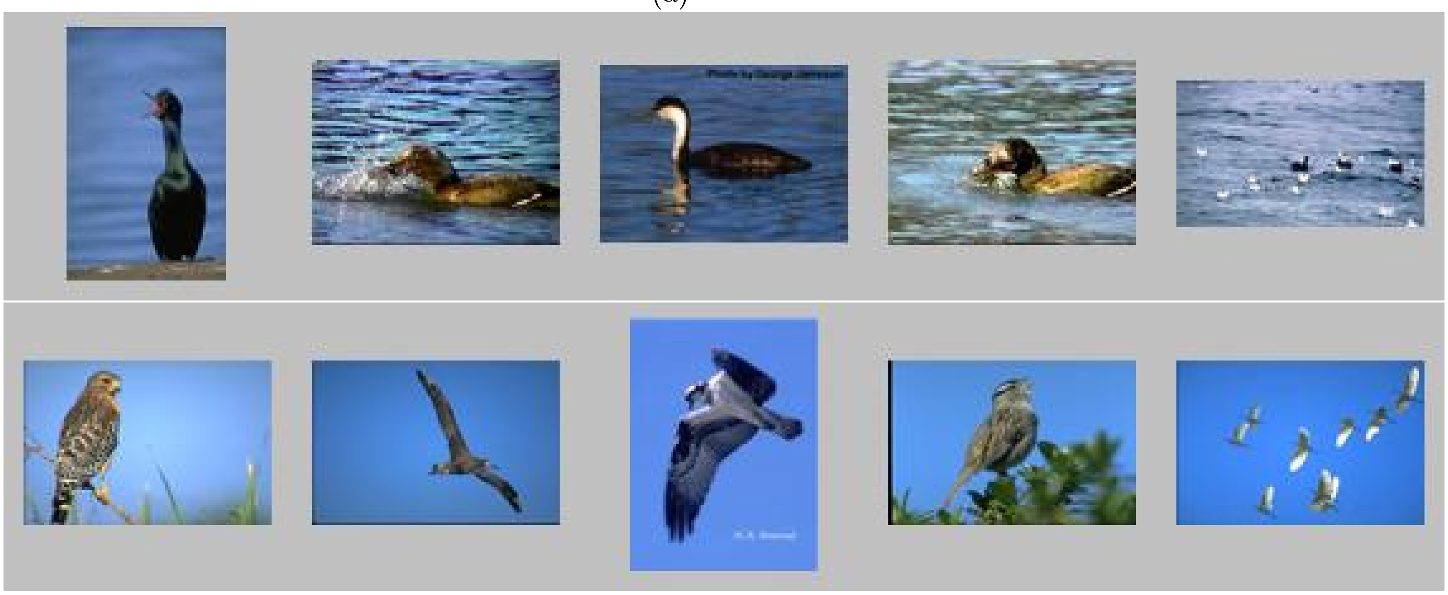
the region of interest determined by the color and edge-based background elimination process described here, and using the whole image.

The retrieval performance of this system is compared with color-based whole image indexing which is very popular and forms the baseline we are proposing to improve upon. Some examples of improvement in retrieval after using our region-of-interest pre-processing are shown in Figure 16. The retrieval results show that birds with colors similar to the query are retrieved in a variety of backgrounds. In some cases, the top-ranked images contain the same species of bird as the query, for example, in the first row of retrieved images in Figure 16(a), the top four images contain the same bird (cormorant). In the second example, all the retrieved birds are predominantly brown, with black and white specks. Except the second bird, all the other birds retrieved in this case are birds of prey, which are relevant to the query image of a kite. As noted in the introduction to this chapter, color alone is insufficient to guarantee that the same species of birds are retrieved. Species with unusual and distinctive colorations are more likely to produce retrieval results where the species of the bird matches that of the query. In other cases where the bird has no distinctive colors, a color-based system can be expected to find other birds of similar color, at best. However, even in this case, it is likely that other birds of the same species would be ranked high in the retrieved list because of the similarity in the colors present and their relative proportions.

For comparison, Figure 16(b) shows the retrieval obtained using the same queries but using the whole image for color-based indexing. The examples clearly demonstrate that the background elements dominate the retrieval in this case. The first query produces other images with water as the background where none of the retrieved birds match the query bird's colors. The query features a black cormorant, while the retrieved images show brown ducks. The second query produces other birds against a blue sky where the third and fifth images show black and white birds, unlike the query which is brown. The same queries when posed on the database after background elimination retrieve images of other birds of similar color which are relevant to the query, without being affected by the type of background they are viewed against.

Unlike some other databases (databases of commercial products, for example), it is very difficult to judge the retrieval results in the bird database without having the knowledge-base of a

(a)



(b)

Figure 16: (a) Region-of-interest-based retrieval : retrieved images in response to the query (the first retrieved image) when the database images were indexed by color from the automatically detected segment of interest only (b) Whole image-based retrieval : images retrieved in response to the same queries as (a), but when the database was indexed by color from whole images, with no segmentation

birdwatcher [3]. The difficulty lies in determining which birds in the database can be considered "similar" to the query. There are hundreds of images in the database in which the bird can be categorized by a layperson to be "brown" or "black" (the most common colors encountered). Therefore, recall-precision scores cannot be computed without a user study among a group of birdwatchers.

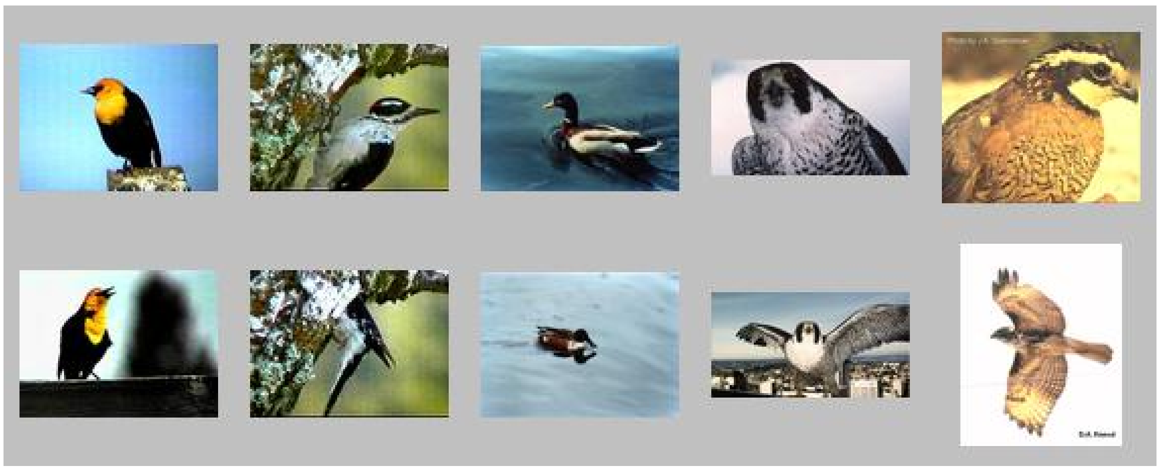


Figure 17: Examples of image pairs used to test retrieval results showing wide variations in size, pose and background

Retrieval method	Average rank of pair
Whole image indexing	6 (for a sub-set of 18) (≥40 for remaining 12)
Indexing using region of interest	3 (on the set of 30)

Table 3: Comparative retrieval results using whole image indexing and indexing in a region of interest

Instead, we have adopted an objective measurement criteria which effectively compares our system with whole image retrieval, without the need for judging each image in the database. This technique known-item-search is well known in information retrieval. A set of 30 pairs of birds were selected where each pair of pictures of the same bird (either by obvious similarity or using

[3]In many cases the identity of the birds in the database is not known to us

cues from the image name given by the original photographer e.g. cormorant1 and cormorant2 obviously refer to two images of the same bird species). Some examples of image pairs used are shown in Figure 17. The pairs show wide variations in the appearance of the birds due to their non-rigid structure, in addition to differences in backgrounds. Using one image of the pair as a query, the rank of the other image of the pair was noted. It is expected that the corresponding image, being the same bird as the query, should appear near the top of the list of retrieved images. Table 3 summarizes the observed results from this test. Using the segmentation technique suggested here, the average rank of the second member of the pair is 3 for all 30 pairs of queries. On the other hand, using whole image retrieval, the average rank in 18 cases is 6 and in the remaining 12 image pairs the rank was beyond 40 (effectively, the pair was not retrieved) using whole image indexing. So we can conclude that using the computed region-of-interest for indexing significantly improves the effectiveness of retrieval. It is worth noting that in many cases this performance is achieved even when the two paired pictures of a bird are taken from widely separated viewpoints and with different backgrounds (see Figure 17).

5 Conclusion

We have proposed a solution to the problem of region of interest extraction while making very general assumptions about the images in the database, which are true of a broad class of images. Our approach to foreground segment detection is based on the elimination of the background. This is accomplished by combining a color-based background detection step with refinement of the segmentation using edge information. We also show that domain specific knowledge can be incorporated into this framework, where available. A more detailed description of this work can be found in [5].

Color histograms from the automatically detected foreground segment are used to index a database of bird images. The retrieval results on this database show that the color of the bird is used for retrieval, without being affected by the colors present in the background. This is a very important improvement in a database of images with single subjects where the query is usually on the subject, and the background is incidental.

A possible extension of this work involves the incorporation of a region of interest selector in the user interface. This would ensure that the color of the bird in the query image would be correctly assessed. When combined with the region-of-interest pre-processor on database images, it could be ensured that a failure to segment the query image correctly (which would lead to very poor results) is avoided. There is some robustness in the database images to erroneous segmentation, since at worst, it would result in the non-retrieval or false retrieval of a few images. In the case of the query, which is a single image, incorrect segmentation would guarantee poor results.

In a domain such as birds, the success of retrieval based on color alone is limited, since color cannot be used to distinguish between birds of different species. Other information such as shape, texture and rules formulated by expert birdwatchers need to be incorporated to ensure better discrimination between different types of birds. However, our method still provides the starting point for computing additional information by segmenting the region of interest from which such information should be gathered.

References

[1] J. R. Bach, C. Fuller, A. Gupta, and et al. The Virage image search engine: An open framework for image management. *Proceedings of SPIE*, 2670:76–87, 1996.

[2] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based segmentation using em and its application to content-based image retrieval. *Sixth International Conference on Computer Vision*, pages 675–682, Jan 1998.

[3] M. Das, , R. Manmatha, and E. M. Riseman. Indexing flowers by color names using domain knowledge-driven segmentation. *IEEE Workshop on Applications of Computer Vision*, pages 94–99, Oct 1998.

[4] M. Das, , R. Manmatha, and E. M. Riseman. Indexing flower patent images using domain knowledge. *IEEE Intelligent Systems*, 14(5):24–33, Oct 1999.

[5] M. Das. Studies in the use of color for image indexing and retrieval in specialized databases. *Ph.D thesis and CIIR Multimedia Technical Report, Dept. of Computer Science, University of Massachusetts*, MM-35, Sept 2001.

[6] M. Das, E. M. Riseman, and B. A. Draper. Focus: Searching for multi-colored objects in a diverse image database. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 756–761, Jun 1997.

[7] J. P. Eakins, K. Shields, and J. Boardman. Artisan - a shape retrieval system based on boundary family indexing. *SPIE Proc. on Storage and Retrieval for Image and Video Databases*, 2670:17–28, Feb 1996.

[8] M. M. Fleck, D. A. Forsyth, and C. Bregler. Finding naked people. *Fourth European Conference on Computer Vision*, pages 72–77, 1996.

[9] D. Forsyth, J. Malik, and R. Wilensky. Searching for digital pictures. *Scientific American*, 276(6):72–77, Jun 1997.

[10] D. A. Forsyth and M. M. Fleck. Body plans. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 678–683, 1997.

[11] B. K. P. Horn. *Robot Vision.* The MIT Press, 1993.

[12] J. Huang, S. Ravi Kumar, M. Mitra, W. J. Zhu, and R. Zabih. Image indexing using color correlograms. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 762–768, Jun 1997.

[13] Q. Huang, B. Dom, D. Steele, J. Ashley, and W. Niblack. Foreground/background segmentation of color images by integration of multiple cues. *International Conference on Image Processing*, 1:246–249, 1995.

[14] M. Kankanhalli, B. Mehtre, and H. Huang. Color and spatial feature for content-based image retrieval. *Pattern Recognition Letters*, 20(1):109–118, January 1999.

[15] K. L. Kelly and D. B. Judd. The iscc-nbs method of designating colors and a dictionary of color names. *National Bureau of Standards Circular*, (553), Nov 1 1955.

[16] P. Kelly, M. Cannon, and D. Hush. Query by image example : the candid approach. *SPIE Proc. on Storage and Retrieval for Image and Video Databases*, pages 238–248, Feb 1995.

[17] F. Liu and R. W. Picard. Periodicity, directionality and randomness : Wold features for image modeling and retrieval. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 18(7):722–733, Jul 1997.

[18] S. Mahamud, K. K. Thornber, and L. R. Williams. Segmentation of salient closed contours from real images. *Seventh International Conference on Computer Vision*, Jun 1999.

[19] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions : Cue integration in image segmentation. *Seventh International Conference on Computer Vision*, Sept 1999.

[20] R. Manduchi. Bayesian fusion of color and texture segmentations. *Seventh International Conference on Computer Vision*, Sept 1999.

[21] J. Matas, R. Marik, and J. Kittler. On representation and matching of multi-coloured objects. *Fifth International Conference on Computer Vision*, pages 726–732, Jun 1995.

[22] B. Mehtre, M. Kankanhalli, and W. Lee. Shape measures for content-based image retrieval : A comparison. *Information Processing and Management*, 33, 1997.

[23] W. Niblack, 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. *SPIE Proc. on Storage and Retrieval for Image and Video Databases*, 1908:173–187, Feb 1993.

[24] V. Ogle and M. Stonebraker. Chabot : Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, Sep 1995.

[25] I. Park, I. Yun, and S. Lee. Color image retrieval using hybrid graph representation. *Image and Vision Computing*, 17(7):465–474, May 1999.

[26] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 84–90, Jun 1994.

[27] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.

[28] S. Ravela and R. Manmatha. Image retrieval by appearance. *SIGIR*, pages 278–285, Jul 1997.

[29] S. Ravela and R. Manmatha. On computing global similarity in images. *Workshop on Applications of Computer Vision*, pages 82–87, Oct 1998.

[30] Y. Rui, T. Huang, and S. Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39–62, March 1999.

[31] J. Shi and J. Malik. Self inducing relational distance and its application to image segmentation. *Fifth European Conference on Computer Vision*, Jun 1998.

[32] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[33] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[34] S. Ullman. *High level vision : Object recognition and visual cognition.* MIT Press, 1996.

[35] J. Wang. *Integrated Region-based Image Retrieval.* Kluwer, 2001.

List of Figures