

Automatic Segmentation and Indexing in a Database of Bird Images *

Madirakshi Das

R. Manmatha

Multimedia Indexing and Retrieval Group

Department of Computer Science

University of Massachusetts

Amherst, MA 01003

e-mail : mdas,manmatha@cs.umass.edu

Abstract

The aim of this work is to index images in domain specific databases using colors computed from the object of interest only, instead of 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 - eliminating the background leaves the figure or object of interest. To find possible object colors, we first find background colors and eliminate them. We then use an edge image at an appropriate scale to eliminate those parts of the image which are not in focus and do not contain significant structures. The edge information is combined with 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 450 bird images tested, the segmentation is sufficient to determine the colors of the bird correctly for retrieval purposes. We also show that our approach provides improved retrieval performance.

1 Introduction

The problem of meaningful retrieval from image databases has generated a great deal of interest in recent years. Most retrieval algorithms have targeted a general image database which may contain diverse types of images [2, 10]. However, 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.

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, flying against a cloudy sky etc. 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 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 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

*This material is based on work supported in part by the National Science Foundation, Library of Congress and Department of Commerce under cooperative agreement number EEC-9209623, in part by the National Science Foundation under grant number IRI-9619117 and in part by NSF Multimedia CDA-9502639. 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).

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 bird may often lie at the borders of the image (Fig. 7), so simply eliminating all the border colors will not work.



Figure 1: Some images in the bird database

The results of our work are illustrated using examples from a database of images of birds. 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. In earlier work [4, 5] 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. The bird database, on the other hand, 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.

Since the primary purpose is to find object colors accurately enough for retrieval, perfect segmentation of the subject is not necessary. The final segment may have small parts of the bird missing or include small areas from the back-

ground without much impact on retrieval performance.

This paper is organized as follows : the next section surveys related work; section 3 discusses the detection and elimination of background colors. The combination of the resulting image with edge information is described in section 4. Section 5 discusses experimental results on segmentation. Section 6 compares retrieval based on the segmentation scheme suggested here versus retrieval based on the entire image. Finally section 7 concludes the paper.

2 Related work

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 [3, 9, 8]. Relational graph matching has been used for segmenting natural images in [12]. 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 [6], but these techniques work well on relatively smooth backgrounds and objects with sufficient contrast. Recently proposed techniques for detecting natural shapes in real images [7] also work best with simple backgrounds.

In the area of color-based indexing, most systems primarily rely on low-level features like color histograms [13, 11]. Since these are computed without extracting the object of interest, the background plays a significant part in the retrieved results. The QBIC image retrieval system [10] uses some automatic and semi-automatic segmentation of object of interest [1].

3 Detection and elimination of background

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 Fig.3) and occupies a reasonable portion of the image. The camera is focused on the subject and distant background is usually out-of-focus as a result. When photographing a specific subject, there is usually an attempt to keep other competing foci-of-interest out of the picture. For example, 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.

3.1 Segmentation strategy

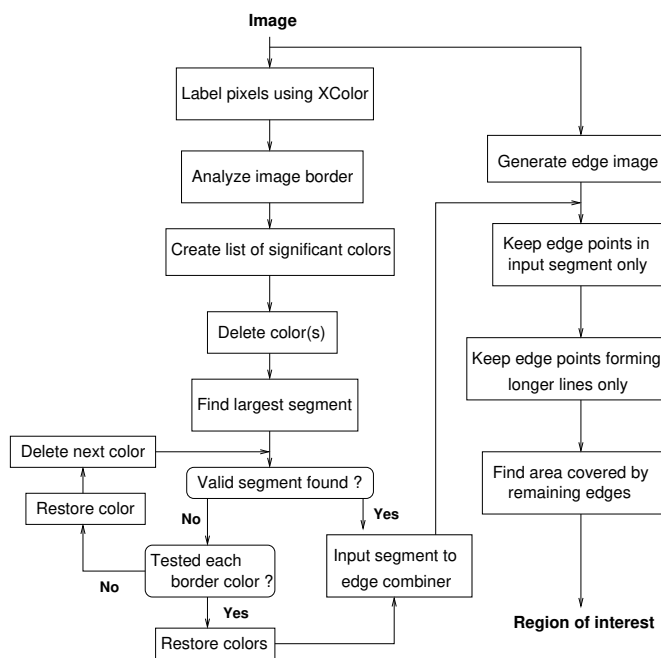


Figure 2: Overview of segmentation strategy

The first step in producing a list of possible background colors is to select a suitable colorspace to label the image pixels. The RGB space in which the original image is de-

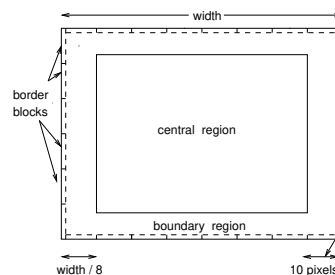


Figure 3: Definitions of image regions

scribed, has too many colors to be useful. 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. The color definitions in X also provide commonly used names for colors e.g. “khaki”, “aquamarine” etc. 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.

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

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 Fig 3. The distribution of X colors in these blocks is computed and colors having a high probability of being in more than one block are marked as possible background colors.

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. Fig.4 shows an example of the largest segment obtained when the colors detected along the periph-

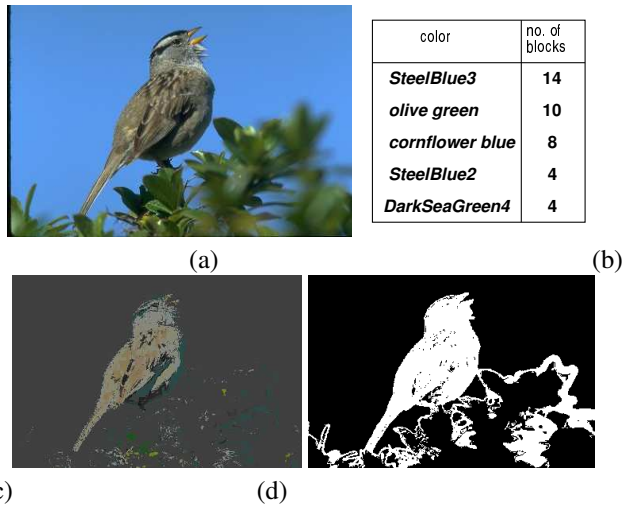


Figure 4: Background elimination : (a) original image (b) significant colors detected along image periphery (c) image left after deleting colors in (b) found along the image periphery (d) largest segment obtained from (c)

ery are deleted after being identified as background colors. Some examples where the largest segment obtained closely matches the bird region of the image are shown in Fig.5.

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 (Fig.3). 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. Fig.6 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 Fig.5. In most images, however, the output can be further improved by additional processing as described in the next section.



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

4 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. There are images where background colors remain because they were not present along the image periphery, and therefore, were not detected by the background elimination process.

Edges associated with the outline of the bird tend to be present at every scale of the image. However, edges associated with the background are usually present only at smaller scales. This is due to several reasons:

The background sometimes consists of uniform regions such as sky. Edges (in the background) if any appear only at the smallest (finest) scales. Many backgrounds associated with bird images consist of textured surfaces such as grass, mud water or trees. The scale of such texture images is usually much smaller than that of the image of the bird. The background may often be blurred (see the top left image in Fig.7) or even if it is not blurred, it is often not as sharp (top left image in Fig. 8) because of the limited depth of field of cameras. This effect is often accentuated by the photographer. Notice that edges associated with the internal structure of the feathers of the bird are 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.

Conveniently, all these effects reinforce each other and

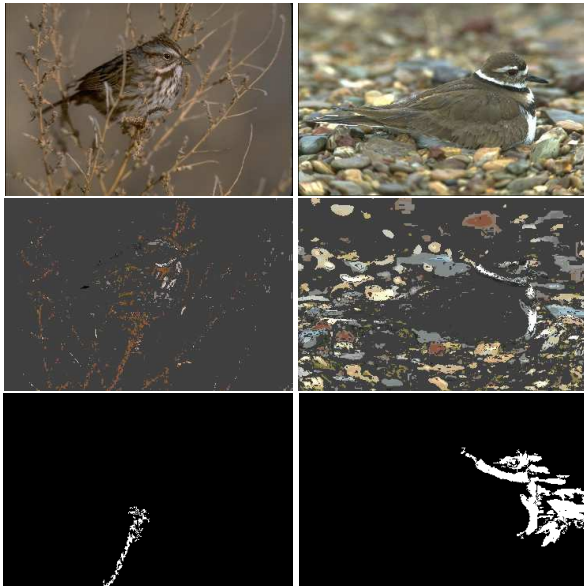


Figure 6: Examples showing detection of invalid segments : (top) original images (mid) after deletion of hypothesized background colors (bottom) largest segments produced (invalid since too small (left) or centroid is in the image boundary region (right))

can be taken advantage of by using a relatively larger scale for detecting edges. At such scales only edges present in the bird will be detected. We do this as follows. The image is convolved with the two first derivatives of a Gaussian. The derivative outputs are then combined to produce the gradient magnitude. The derivatives of Gaussians are energy normalized (by dividing by the scale). This 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 Fig. 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 to eliminate the entire background and that it is only the combination of the edge and color information which allows background elimination.

The edge image is then combined with the foreground segment output by the color-based background elimination process. The first step in the combination process is to eliminate edge points that are not in the foreground segment. This should eliminate most of the edges from the background. The next step finds connected components in the remaining edge points and eliminates small and isolated edge segments.

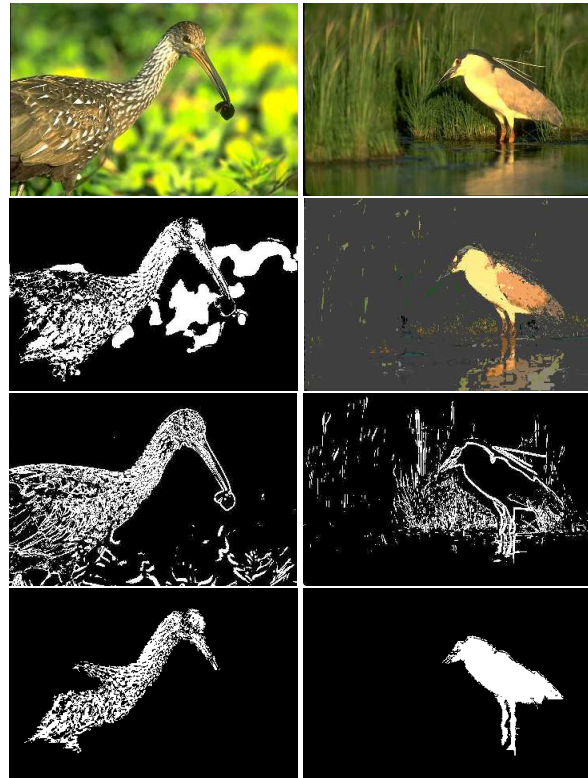


Figure 7: Examples showing improvements in the bird segment extracted when edge information is incorporated : (row 1) original images (row 2) segment after background color deletion (row 3) edge image (row 4) final output

Edge segments smaller than 20% of the total number of edge points are considered to be too small. This leaves the longer edge segments only. 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 inside/outside relationship (it was also suggested for use in object recognition by Ullman [14]). 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. 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. Some examples where the edge information is able to improve the segmentation produced by color-based foreground-background discrimination are shown in Fig.7.

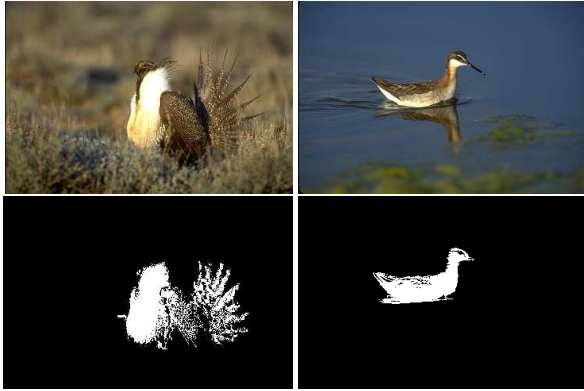


Figure 8: Examples showing elimination of background and final segment obtained

5 Experimental results

The automatic segmentation results were manually checked for 450 images in the bird database. Table 1 shows the breakup of the results. In 57% of the images, the background is totally eliminated, some examples of this case is shown in Fig.8. In 13% of the images, the greater part of the background is eliminated and the remaining background does not alter the color distribution of the final segment significantly. Some examples of this case are shown in Fig.9. When significant amount of background remains in the final segment, the color distribution computed for the bird is not accurate. Examples of such images is shown in Fig.10 and these constitute 11% of the images. In images where the bird is well camouflaged, it is not possible to extract the foreground on the basis of color. In such cases, the whole image is used for indexing. The earlier figure, Fig.6 shows some examples of this case, Fig.11s show some additional examples. However, since the background color and the color of the bird are the same or very similar in this case, indexing is not adversely affected. Fig.12 shows two cases where the segmentation algorithm failed; the bird was eliminated altogether and the output consists of parts of the background. This happened 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. However, the 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.

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. Fig.13 show an example of other subjects ex-

No background left	57%	Good for indexing
Insignificant background left	13%	Good for indexing
Image unchanged	17%	OK for indexing
Significant background left	11%	Indexing affected
Incorrect segmentation	2%	Indexing affected

Table 1: Automatic segmentation results.

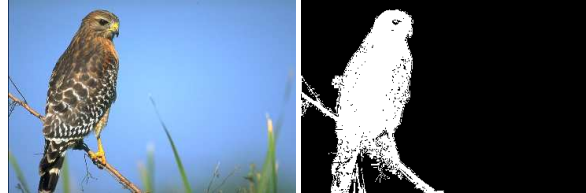


Figure 9: Example showing partial elimination of background where the included background does not affect the color distribution of the final segment significantly.



Figure 10: Example showing partial elimination of background where the included background affects the color distribution of the final segment.



Figure 11: Examples showing cases where a valid bird segment could not be extracted based on color



Figure 12: Example showing failure cases where the bird segment was deleted.

tracted correctly (in the case of the snake the segment extracted is sufficient to determine its color). We believe that this technique will be applicable to image databases of many

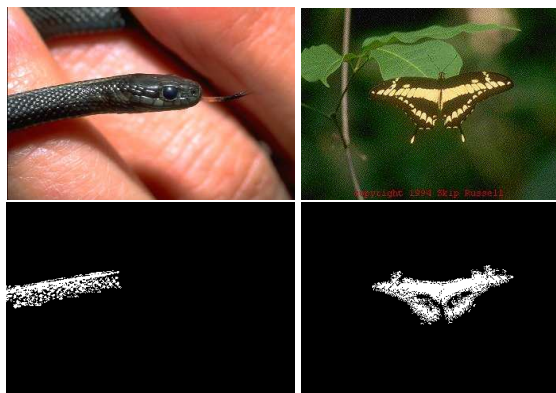


Figure 13: Examples showing correct detection of subject in other domains (top) original images (bottom) final segments obtained

other objects including snakes, butterflies, fish and mammals.

6 Indexing and retrieval

The database of bird images is indexed using color histograms [13] generated from the region of interest determined by the color and edge-based background elimination process described earlier. Our database has about 700 images downloaded from the world wide web¹. The images vary widely in quality, with sizes ranging from 12Kb to 40Kb, with birds occupying a variable part of the image.

Some examples of the retrieval obtained are shown in Fig.14. Since the retrieval is on the basis of color, all that we can ensure is that the colors of the query birds and the retrieved birds are the same. The first retrieval actually seems to produce retrievals of birds of the same kind while the second retrieval interestingly seems to retrieve other birds of prey (perhaps many birds of prey share the same coloring scheme). Note that the retrievals are not affected by background color and are independent of viewpoint and scale.

For comparison, Fig.15 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, the second query produces other birds against a blue sky and the third query generates birds against green backgrounds, with the color of the bird playing a secondary role in the retrieval. The same queries when posed on the database after background elimination retrieve images of other birds of

¹We ran the automatic segmentation on all 700 images even though the segmentation was only manually checked on 450 images because of the labor intensive nature of the checking process.

similar color, without being affected by the type of background they are featured against.

Other retrievals are similar in nature. We hope to do a more quantitative evaluation when time permits.

7 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 elimination of background. This is accomplished by combining a color-based background detection step with refinement of the segmentation using edge information.

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.

References

- [1] J. Ashley and et al. Automatic and semi-automatic methods for image annotation and retrieval in qbic. *SPIE Conference on Storage and Retrieval for Image and Video Databases*, pages 24–35, 1995.
- [2] J. R. Bach, C. Fuller, A. Gupta, and et al. The virage image search engine: An open framework for image management. *SPIE Conference on Storage and Retrieval for Still Image and Video Databases*, 2670(IV):76–87, Feb 1996.
- [3] 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.
- [4] 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.
- [5] M. Das, , R. Manmatha, and E. M. Riseman. Indexing flower patent images using domain knowledge. *IEEE Intelligent Systems*, 14(5):24–33, Oct 1999.
- [6] 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.
- [7] S. Mahamud, K. K. Thornber, and L. R. Williams. Segmentation of salient closed contours from real images. *Seventh International Conference on Computer Vision*, pages 891–897, Sept 1999.



Figure 14: Examples showing images retrieved in response to the query (seen as the first retrieved image) when the database images were indexed by color from the automatically detected foreground segment only



Figure 15: Examples showing images retrieved in response to the same queries as the previous figure, but when the database was indexed by color from whole images, without segmentation

- [8] 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.
- [9] R. Manduchi. Bayesian fusion of color and texture segmentations. *Seventh International Conference on Computer Vision*, pages 956–962, Sept 1999.
- [10] W. Niblack, R. Barber, and et al. The qbic project: Querying images by content using color, texture and shape. *SPIE Conference on Storage and Retrieval for Image and Video Databases*, 1908:173–187, Feb 1993.
- [11] G. Pass and R. Zabih. Histogram refinement for content-based image retrieval. *Workshop on Applications of Computer Vision*, pages 96–102, 1996.
- [12] J. Shi and J. Malik. Self inducing relational distance and its application to image segmentation. *Fifth European Conference on Computer Vision*, Jun 1998.
- [13] M. J. Swain and D. H. Ballard. Indexing via color histograms. *Third International Conference on Computer Vision*, pages 390–393, 1990.
- [14] S. Ullman. *High level vision : Object recognition and visual cognition*. MIT Press, 1996.