CATEGORY PRUNING IN IMAGE DATABASES USING SEGMENTATION AND DISTANCE MAPS

Baris Sumengen, B. S. Manjunath

ECE Department, UC, Santa Barbara
93106, Santa Barbara, CA, USA
email: {sumengen,manj}@ece.ucsb.edu
web: vision.ece.ucsb.edu

ABSTRACT

A novel framework for pruning a category of images is proposed in this paper. We assume no prior information about the contents or semantics of the images. Our framework discovers consistencies and knowledge about the spatial relations of the categories unsupervised using iterative image segmentation and spatial grouping. A measure for deciding how well an image fits to a category is proposed and the effectiveness of this measure is investigated.

1. INTRODUCTION

With the increasing use of digital cameras, growth of the Internet and high importance of periodic acquisition of aerial and satellite images, image databases are becoming more and more important every year. The world wide web (www) itself probably contains a majority of the images in existence and these images can be assigned text captions and keywords, which makes creating image databases from web images an interesting and challenging task. Most of the image databases are or can be divided into semantic subgroups, which we call categories. One example for a category is a directory of photographs in a home user's computer. Another example of a category is the group of images in an internet image database that has the keyword "car" associated to them. Unfortunately most image databases are not generated with the semantic categorization in mind.

In our previous work [1], we show an efficient way to spider and simultaneously categorize web images for content based retrieval. On this database of over 600,000 images, we achieve impressive retrieval results by exploiting the category structure. For example, the "Cars" category consists of images spidered from car related web sites. About 60-80% of images in this category are images of cars. In addition, there are also other pictures that are not of cars on such pages, e.g., picture of the owner of a car dealership. This is a common problem with automatically generated categories. To increase the overall effectiveness of retrievals, further pruning is needed. Pruning will improve both browsing and content-based retrieval quality of these databases. The main objective of the pruning is increasing the precision of the category while preserving the recall. By using our graph partitioning active contour segmentation method to separate background from foreground, we demonstrate that good image segmentation helps in this pruning step and improves the overall retrieval performance.

Image segmentation has been used in image databases as a tool to limit extraction of image features to the segments of interest. These segment-based feature vectors are then used for image or region search in a content-based image retrieval system [2, 3] and for browsing image databases [4]. In this paper we propose a new way of utilizing segmentation for image databases. The objective of our work is to improve the quality of both image segmentation and image database retrieval simultaneously.

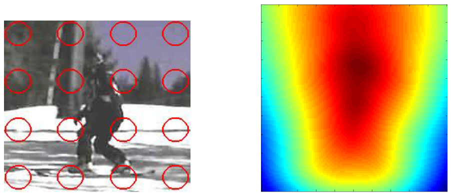
In developing a pruning strategy using segmentation, we make the following assumptions: 1) We do not have any domain specific knowledge, or object models about the category. 2) The category is automatically generated, but has some consistency such that large number (more than 50%) of the images in this category fit to a certain unknown semantic concept. 3) Images can vary in terms of image features or shape of the objects even if they follow the category concept. 4) Images can vary in their sizes. In our case, the images are between 128×128 and 512×512.

We propose a solution to this problem by discovering a loose spatial relationship between the background and foreground of the images. The labels *background* and *foreground* have no importance in our method but are used for consistent association of regions among category images. We first segment all images into foreground and background regions using graph partitioning active contours (GPAC) [5, Chapter 5]. Using these segmentations, a signed distance map is calculated for each image based on the segmentation boundary. Then, these distance maps are averaged to find a distance map for the whole category. A continuous foreground/background spatial relationship within the category is captured by this average distance map. After that, the images are re-segmented, but this time the segmentation cost function also incorporates the spatial relationships discovered from the previous iteration. We show that this step improves both the segmentation of certain images and the distance map itself.

The pruning of the category is achieved by comparing the individual distance maps to the average distance map. We show the distribution of this comparison and interpret the results by showing specific examples. In addition, precision recall curves are drawn for visualization of how well our method works.

2. SEGMENTATION FRAMEWORK

We use segmentation as two-region (background and foreground) partitioning of an image. We will show in Section 3 situations where multi-region (3 or 4 regions) segmentation might be necessary. Our segmentation approach, GPAC, is based on grouping similar points as foreground and background while increasing the dissimilarity between these regions. One important characteristic of GPAC is its flexibility in defining the segmentation cost function. By controlling



(a) (b)

Figure 1: a) Initialization of the curve evolution for image segmentation. b) Average distance map.

the similarity measure, we can change and adapt the segmentation behavior to the problem at hand. We calculate the dissimilarities as:

$$w(p_i, p_j) = \|\vec{F}(p_i) - \vec{F}(p_j)\| + \alpha \|p_i - p_j\| \tag{1}$$

where $\vec{F}(p_i)$ is some low level image feature at point p_i (we use color) and the second term measures the spatial distance between two points. Based on this dissimilarity metric, the segmentation functional that we maximize is given by:

$$E = \iint_{R_i(C)} \iint_{R_o(C)} w(p_1, p_2) dp_1 dp_2 \tag{2}$$

where C is a closed curve on the image domain, R_i is the interior and R_o is the exterior of C. Conceptually this means finding the curve (partitioning of the image) for which the dissimilarity between its interior (foreground) and exterior (background) is maximized. This problem is solved by starting with an initial curve and maximizing (2) using steepest descent. After adding geometric constraints such as area normalization and boundary length minimization, the corresponding curve evolution equation becomes:

$$\frac{\partial C}{\partial t} = \left(\frac{1}{|R_o|} \iint_{R_o(C)} w(c,p) dp - \frac{1}{|R_i|} \iint_{R_i(C)} w(c,p) dp \right) \vec{N} - \gamma \kappa \vec{N} \tag{3}$$

This result can be visualized as the competition of foreground and background to push the curve towards the optimum boundary. Fig. 1a shows the curve initialization used in this paper. We initialize one multi-part curve with 16 (4×4) sub-parts, so that the curve can cover most of the image area.

3. PRUNING A CATEGORY

The pruning strategy is based on first segmenting the images in a category. The signed distance of the pixels from the segmentation boundary is used as a measure for spatial relations within the images. By averaging image level spatial relations, we discover category-wide spatial relations and propose a measure for calculating the association of images to their corresponding categories.

The images in the categories are of various sizes between 128×128 and 512×512. Before segmenting these images, we resize them proportionally such that the smaller dimension of the image is mapped to 64. This small size is selected for computational reasons so that large number of images can be segmented quickly. After segmenting and extracting the boundaries, we generate a signed distance map. This map is calculated as the distance of each pixel to the segmentation boundary. The value is positive if the pixel belongs to the foreground and negative if the pixel belongs to the background. Before calculating this map, we need to decide which region is background and which region is the foreground. This is important for consistency among the images. So, we calculate the average distance from the image boundaries (not segmentation boundary) for each region. We define the foreground as the region with higher distance from the image boundaries. After calculating the signed distance map, we linearly scale the values between 0 to 255. Then using standard image resizing algorithms the distance maps are resized to the size $N \times N$. We select $N = 100$ in our experiments. The distance maps of images are averaged pixel-wise to create the average distance map of the category. We use this average map as the representation of the category.

After calculating the average map, we re-segment all images by using a new dissimilarity measure that incorporates spatial relations discovered in the previous step.

$$w(p_i, p_j) = \|\vec{F}(p_i) - \vec{F}(p_j)\| + \alpha \|p_i - p_j\| + \beta \|m(p_i) - m(p_j)\| \tag{4}$$

where m is the average distance map. The constant β is selected such that segmentation is not completely biased to this new term. After a second iteration of the segmentations, a new average distance map is generated. We now select this new distance map as the representation of the category's spatial relations.

To make decisions for eliminating images from the category, we need to check if an image follows the concept of the category or not. To create a measure of how well an image fits to a category, we calculate the distance between the individual distance maps and the average distance map. Distances are calculated using:

$$\|m_{avg} - m_i\| = \sqrt{\sum (m_{avg}(p) - m_i(p))^2 / N^2}$$

After calculating the distances, a cutoff point d_{cut} needs to be estimated such that images with higher distance values are discarded from this category. One way is to analyze the shape of the histogram and make decisions. Simple decisions such as eliminating the tail of the histogram improves the quality of the category but this type of ad hoc decisions might be suboptimal. A better approach would be using a learning framework [6] to discover a methodology for deciding on d_{cut}. This may require generating ground truth for large number of categories.

4. EXPERIMENTAL RESULTS

The category we experiment with is *Sports: Winter Sports: Skiing: Guides: North America: United States*. The main theme for the images in this category (79%–110 out of 139) is that a skier is either posing for a photograph or skiing or snowboarding downhill when the picture is taken. These images vary significantly in terms of foreground and background color, environment and pose. The rest of the images consist of scenic views of snowy mountains, maps, picture of jackets and backpacks, and a set of unrelated images. Fig 2 shows examples of characteristic images and Fig. 3 shows examples of uncharacteristic images for the category.

First, images are resized such that the smaller dimension of the image is mapped to 64. After that, all images are segmented using GPAC. Segmentation results for some images

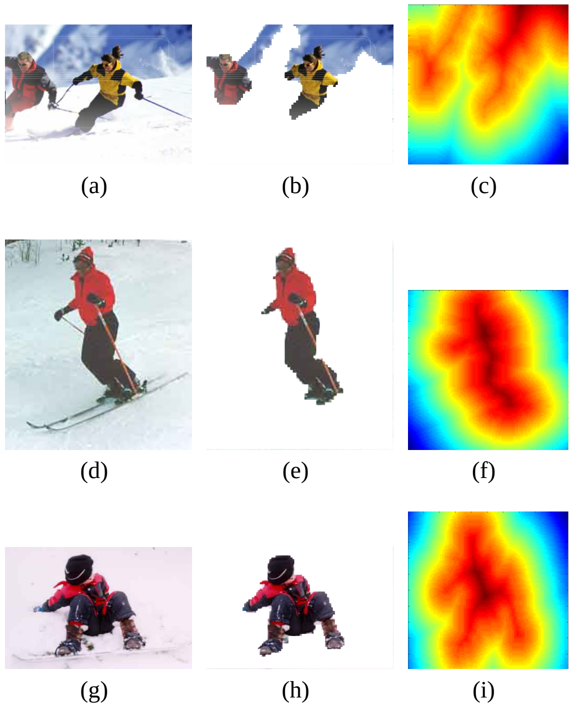


Figure 2: First column shows the original image. Foreground is shown in second column. Third column show the signed distance map.

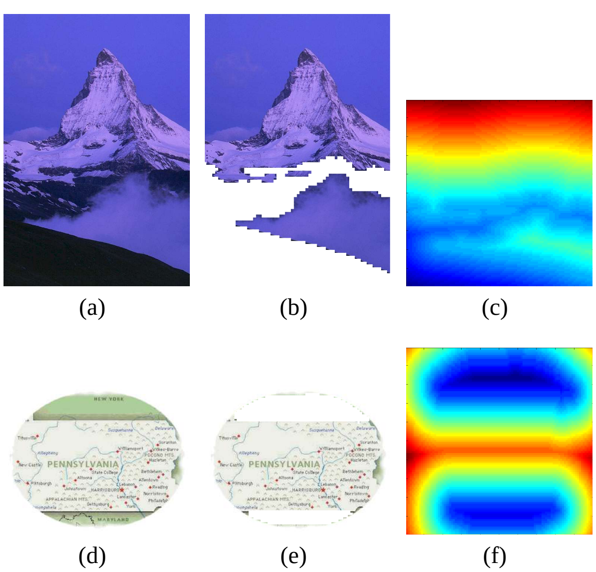


Figure 3: First column shows the original image. Foreground is shown in second column. Third column show the signed distance map.

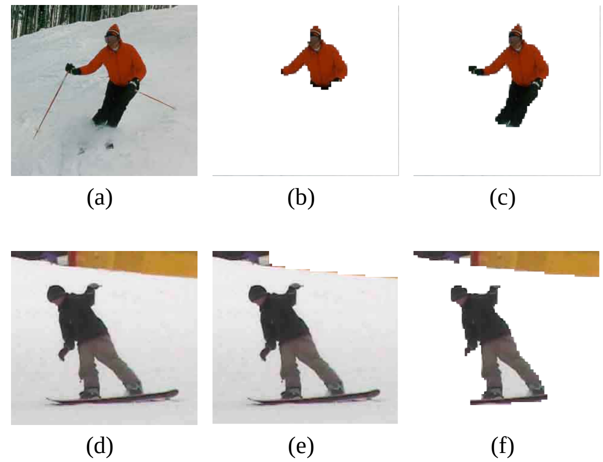


Figure 4: First column shows the original image. Initial foreground is shown in second column. Foreground after using spatial relations is shown in third column.

are shown in the second column of Fig. 2 and Fig. 3. After extracting the boundaries, the signed distance maps are generated. Distance maps for some images are shown in the third column of Fig. 2 and Fig. 3. The average distance map for the category is shown in Fig. 1b. As can be seen, the average distance map follows this category's spatial relations such that the foreground objects are in the middle part of the images in an upright position.

Using the average map, images in this category are re-segmented as discussed in Section 3. Several examples are shown in Fig. 4. New foregrounds in the third column show that the segmentation process has incorporated the spatial relations discovered from the category. The observation regarding the new average distance map is that this average distance map is not much different than the previous one. The reason for this is that high percentage of the images follow the category characteristics (average distance map) and segmentations for these images did not change. This and our experiments also show that more iterations of segmentations are not necessary.

After the second iterations of the segmentations, distances of individual distance maps to the average distance map are calculated. In Fig. 5, the first row shows 4 images with the smallest distances and second row shows the images with the largest distance values. Fig. 6 (a) shows the histogram of distance values for this category. It can be seen from the figure that there are three peaks that are separated at 0.55 and 0.7. After the third peak there are outliers starting from 0.78 and up. We observe that the images within the first peak are following the category theme very well. Within the second peak, it is a combination of good and bad images for the category. Fig. 7 shows examples of images that fall into the second and third peak. Images in Fig. 7 (a-d) show that the pose of the skiers are not vertical unlike most other images. Adding some level of rotation invariance to our method would be helpful for these images. Fig. 7 (e-f) show two skiers using the lift. The problem with this image is that the skiers are not located in the middle part of the image while most category images have skiers in the middle parts. To handle this kind of images, our method needs to incorporate

Figure 5: a-d) Images with smallest distances. e-h) Images with the largest distances.

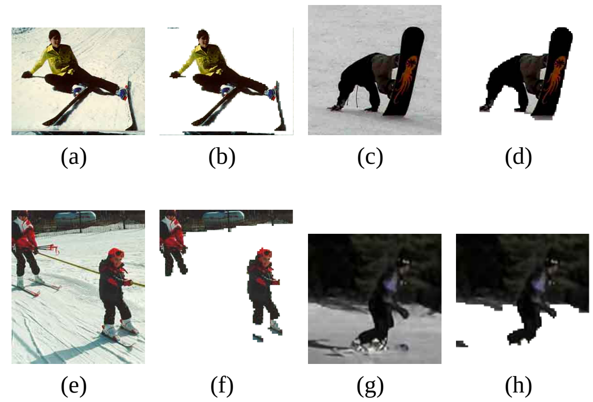


Figure 7: Images that could be misclassified.

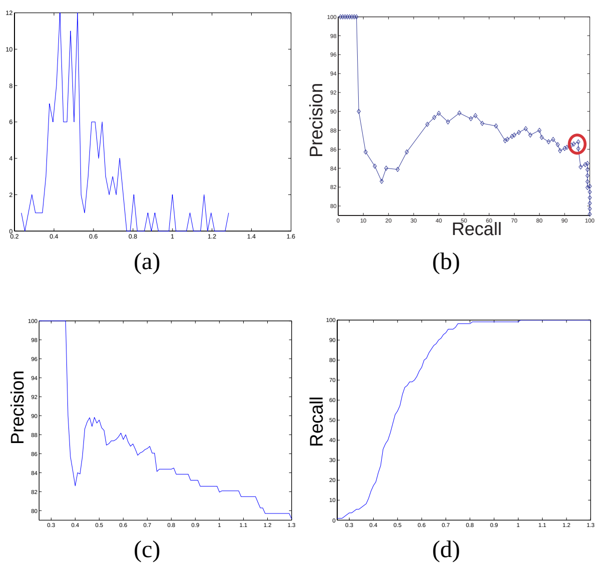


Figure 6: a) Histogram of the distances. b) Precision-Recall curve. c) Precision vs. cutoff point. d) Recall vs. cutoff point.

shift invariance. Last example is in Fig. 7 (g-h). In this example, the background consists of two different regions, the snow and the darker background corresponding to the trees. Two region segmentation is not able to handle this situation and part of the background is labeled as foreground. A 3 region or 4 region segmentation (can be achieved by recursive bi-partitioning) would be helpful when handling more complex images.

Now suppose we decide on a cutoff point d_{cut} for the distances, such that images with higher distance values are discarded from this category. Fig. 6(b) shows the precision-recall[1] graph for cutoff points starting with 0.25 and with increments of 0.01 up to 1.3. At 1.3, nothing is discarded, so we have 100% recall and 79% precision. We see from the graph that cutoff point 0.71 (red circle in the graph) is optimal with 96% recall and 87% precision. At this point, 5 out of 110 good images (4.5%) and 13 out of 29 bad images (45%) are discarded. We see that the precision of the category improves by 8% from 79% to 87%.

[1] After discarding images with distances bigger than a cutoff point, precision measures the percentage of good images within the remaining images in the category, and recall measures the percentage of the good images that are still left in the category.

5. DISCUSSION

We have proposed a simple method for category pruning and discovering spatial relations in image databases. The signed distance maps created from each image can be thought of a new feature for the image. Well known image features such as color and texture can also be used together with the distance map when capturing a concept for a category. It is possible that in some categories there is no coherence within images in terms of color whereas the spatial relations are consistent. For other categories the opposite may be true. Because of the variety of images in different categories, more complex analysis of images might be necessary. Moreover, rotation invariant, scale invariant and shift invariant spatial relations or a weighted combination of these might improve pruning results for certain categories.

Acknowledgements: This work was supported by the National Science Foundation award NSF ITR-0331697.

REFERENCES

[1] S Newsam, B Sumengen, and B S Manjunath. Category-based image retrieval. In *ICIP*, pages 596–9, 2001.

[2] W. Y. Ma and B. S. Manjunath. Netra: a toolbox for navigating large image databases. In *ICIP*, pages 568–71, 1997.

[3] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE PAMI*, pages 1026–38, Aug 2002.

[4] B. S. Manjunath and W. Y. Ma. Browsing large satellite and aerial photographs. In *ICIP*, pages 765–8, 1996.

[5] Baris Sumengen. *Variational image segmentation and curve evolution on natural images*. PhD thesis, UC, Santa Barbara, Sep 2004.

[6] Vladimir Cherkassky and Filip Mulier. *Learning from Data: Concepts, Theory, and Methods*. Wiley-Interscience, March 1998.