

Iterative Figure-Ground Discrimination

Liang Zhao and Larry S. Davis
UMIACS
University of Maryland
College Park, MD, USA

Abstract

Figure-ground discrimination is an important problem in computer vision. Previous work usually assumes that the color distribution of the figure can be described by a low dimensional parametric model such as a mixture of Gaussians. However, such approach has difficulty selecting the number of mixture components and is sensitive to the initialization of the model parameters. In this paper, we employ non-parametric kernel estimation for color distributions of both the figure and background. We derive an iterative Sampling-Expectation (SE) algorithm for estimating the color distribution and segmentation. There are several advantages of kernel-density estimation. First, it enables automatic selection of weights of different cues based on the bandwidth calculation from the image itself. Second, it does not require model parameter initialization and estimation. The experimental results on images of cluttered scenes demonstrate the effectiveness of the proposed algorithm.

1. Introduction

Figure-ground discrimination is an important low-level computer vision process. It separates the foreground figure from the background. In this way, the amount of data to be processed is reduced and the signal to noise ratio is improved.

Previous work on figure-ground discrimination can be classified into two categories. One is pairwise grouping approach; the other is central grouping approach. The pairwise grouping approach represents the perceptual information and grouping evidence by graphs, in which the vertices are the image features (edges, pixels, etc) and the arcs carry the grouping information. The grouping is based on measures of similarity between pairs of image features. Early work in this category [1, 2] focused on performing figure-ground discrimination on sparse features such as edges or line segments and did not make good use of region and color information. Later, Shi and Malik [3] developed

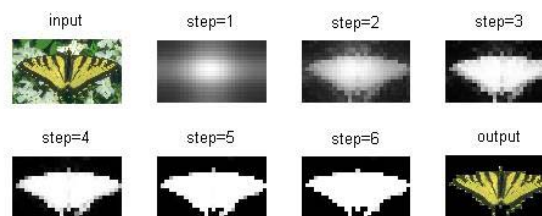


Figure 1. An example of iterative figure-ground discrimination

a normalized-cut mechanism for partitioning a graph into groups. They perform grouping directly at the pixel-level, which is computationally expensive for dense graphs. If the number of pairs is restricted, then a foreground object may be segmented into multiple pieces.

In contrast, the central grouping approach works by comparing all pixels to a small number of cluster centers; examples include k-means [6], and EM clustering [4, 5] with a mixture of Gaussians. Central grouping methods tend to be computationally more efficient, but are sensitive to initialization and require appropriate selection of the number of mixture components. The Minimum Description Length Principle [4] is a common way for selecting the number of mixture components. Our experiments however have shown that finding a good initialization for the traditional EM algorithm remains a difficult problem.

To avoid the drawback of traditional EM algorithm, we propose a non-parametric kernel estimation method which represents the feature distribution of each cluster with a set of sampled pixels. An iterative sampling-expectation (SE) algorithm is developed for segmenting the figure from background. The input of the algorithm is the region of interest of an image provided by the user or another algorithm. We assume the figure is in the center of the region and initialize its shape with a Gaussian distribution. Then the segmentation is refined through the competition between the figure and ground. This is an adaptive soft labeling procedure (see

Fig. 1 for an example).

Our algorithm overcomes the disadvantages of both pairwise and central grouping methods. Since both the figure and background are encoded with non-parametric models, our algorithm avoids the parameter initialization and updating steps in the EM algorithm. And since all pixels are compared to a small number of sampled pixels instead of pairwise comparison, our algorithm is computationally more efficient than pairwise grouping methods. The experiments on real images of natural scenes demonstrate that the SE algorithm is robust to initialization and gives better results than the traditional EM algorithm.

2. Kernel Density Estimation of Color Distribution

2.1. Parametric vs. Non-parametric Estimation

There are two ways for estimating the probability density function of a stochastic process — the parametric and non-parametric approaches. The parametric density estimator assumes a parametric model for the underlying process, specifying a particular form for the underlying density such as Gaussian or a mixture of Gaussians. This approach is efficient but makes strong assumption about the structure of the data. On the other hand, the purely non-parametric approach such as histogram estimators, makes no assumptions about the underlying process, and imposes no formal structure on the data. Although the non-parametric approach is informative and flexible, it suffers from two serious drawbacks: it is not smooth, and it does not provide an adequate description of local properties of a density function. Both of these problems can be solved using a kernel density estimator.

2.2. Kernel Density Estimation

Given a set of samples $S = \{x_i\}$ where $i = 1 \dots N$ and x_i is a d -dimensional vector, kernel density estimation [8] can be used to estimate the probability that a data $y = (y_1, \dots, y_d)$ is from the same distribution as S

$$\hat{p}(y) = \frac{1}{N\sigma_1 \dots \sigma_d} \sum_{i=1}^N \prod_{j=1}^d k\left(\frac{y_j - x_{ij}}{\sigma_j}\right), \quad (1)$$

where the same kernel function is used in each dimension with different bandwidth σ_j . The additive form of Eq. (1) implies that the estimate retains the continuity and differentiability properties of the kernel function. Theoretically, kernel density estimators can converge to any density shape with enough samples [8].

The kernel density estimator provides a bridge between making no assumptions on formal structure (a purely non-parametric approach) and making very strong assumptions (a parametric approach). By making the relative weak assumption about what the true density of the data might be, it is possible to extract more information from the data than is possible using purely non-parametric methods, and it avoids rigid distributional assumptions, thereby providing density estimation that is both flexible and robust.

2.3. Color Density Estimation

In the figure-ground discrimination case, we represent the color of each pixel as a three-dimensional vector (r, g, s) where $r = R/(R + G + B)$, $g = G/(R + G + B)$ are two chromaticity variables and $s = (R + G + B)/3$ is a brightness variable. The separation of chromaticity from brightness in the rgs space allows the use of a much wider kernel with the s variable to cope with the variability in this variable due to shading effects. On the other hand, the chromaticity variables r, g are invariant to shading effects and therefore much narrower kernels can be used in these dimensions, which enables more powerful chromaticity discrimination.

Given a sample of pixels $S = \{x_i = (r_i, g_i, s_i)\}$ from figure F , an estimate $\hat{P}(x_i \in F)$ for the color density can be calculated as

$$\hat{P}(x_i \in F) = \frac{1}{N} \sum_{i=1}^N K_{\sigma_r}(r - r_i) K_{\sigma_g}(g - g_i) K_{\sigma_s}(s - s_i), \quad (2)$$

where $K_\sigma(t) = 1/\sigma K(t/\sigma)$. We use Gaussian kernels, i.e., $K_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2(\frac{t}{\sigma})^2}$ with different bandwidth in each dimension. Theoretically, for Gaussian case the bandwidth can be estimated as $h \approx 1.06\hat{\sigma}n^{-1/5}$ [8] where $\hat{\sigma}$ is the estimated standard deviation and n is the sample size. The use of kernel density estimation for color distribution enables automatic selection of weights of different cues based on the bandwidth calculation from the image itself.

2.4. Normalized KL-Divergence

We evaluate the significance of a figure by comparing the color distributions between the figure and ground. The more dissimilar they are and the larger the figure is, the more significant the figure is. Accordingly, we design the following normalized KL-divergence measure for evaluating the significance of a figure at a certain scale, and use this measure to select the proper scale for figure-ground discrimination.

$$nKL(\hat{P}_{fg} || \hat{P}_{bg}) = \frac{-\sum_{i=1}^N \hat{P}_{fg}(x_i) \log \frac{\hat{P}_{bg}(x_i)}{\hat{P}_{fg}(x_i)}}{\sum_{i=1}^N \hat{P}_{fg}(x_i)}, \quad (3)$$

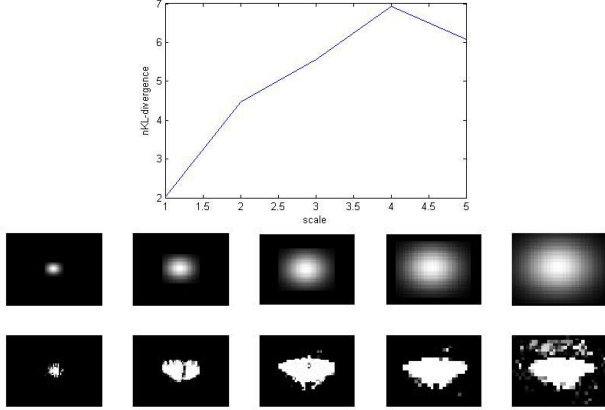


Figure 2. Segmentation results at different scales

where \hat{P}_{fg} and \hat{P}_{bg} are the PDFs of the figure and ground respectively, and x_i is a sampled pixel in $S = \{x_i\}$.

3. The Iterative Sampling - Expectation algorithm

We assume that the pixels in an image were generated by two processes — the figure and ground processes. Then the figure-ground discrimination problem involves assigning each pixel to the process that generated it. If we knew the probability density functions of the two processes, then we could assign each pixel to the process with the maximum likelihood. Likewise, if we knew the assignment of each pixel, then we could estimate the probability density functions of the two processes. This chicken-and-egg problem suggests an iterative framework for computing the segmentation. Unlike the traditional EM algorithm which assumes mixed Gaussian models, we employ the kernel density estimation to approximate the color density distribution of each process. A set of pixels are sampled from the image for kernel density estimation. Thus, the maximization step in the EM algorithm is replaced with the sampling step. This gives the basic structure of an iterative sampling- expectation (SE) algorithm:

1. Start with a Gaussian spatial distribution for all pixels in the image. We select the scale of the initial Gaussian distribution which maximizes the normalized KL-divergence given in Eq. (3). Fig. 2 demonstrates that we can obtain the correct segmentation at the right scale.
2. S step: uniformly sample a set of pixels from the image for kernel density estimation.

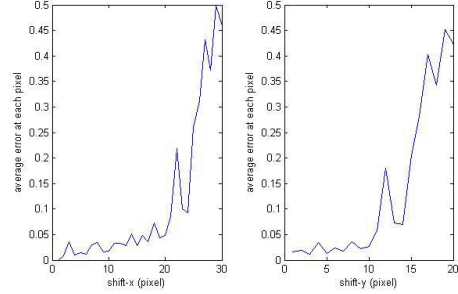


Figure 3. Sensitivity to the initial location of the center of the distribution

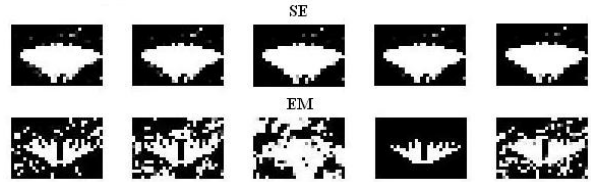


Figure 4. Compare the sensitivity of the SE and EM methods to initialization

3. E step: re-assign pixels to the two processes based on maximum likelihood estimation.
4. Repeat steps 2 and 3 until the segmentation becomes stable.

Since the assignments of pixels are soft, we cannot use the kernel density estimation in Eq. (1) directly. Instead we design a weighted kernel density estimation and make use of the samples from both the foreground and background. Given a set of samples $S = \{x_i\}$ from the whole image, we estimate the probabilities of a pixel belonging to the foreground and background by first calculating the following two values

$$w_{fg}(y) = \sum_{i=1}^N \hat{P}_{fg}(x_i) \prod_{j=1}^d k\left(\frac{y_j - x_{ij}}{\sigma_j}\right), \quad (4)$$

$$w_{bg}(y) = \sum_{i=1}^N \hat{P}_{bg}(x_i) \prod_{j=1}^d k\left(\frac{y_j - x_{ij}}{\sigma_j}\right), \quad (5)$$

where N is the number of samples (we use 6% of the pixels in the image) and d is the dimension of the feature space. Through normalization we get the soft assignments of each pixel to the foreground and background:

$$\hat{P}_{fg} = w_{fg} / (w_{fg} + w_{bg}), \quad (6)$$

$$\hat{P}_{bg} = w_{bg} / (w_{fg} + w_{bg}). \quad (7)$$

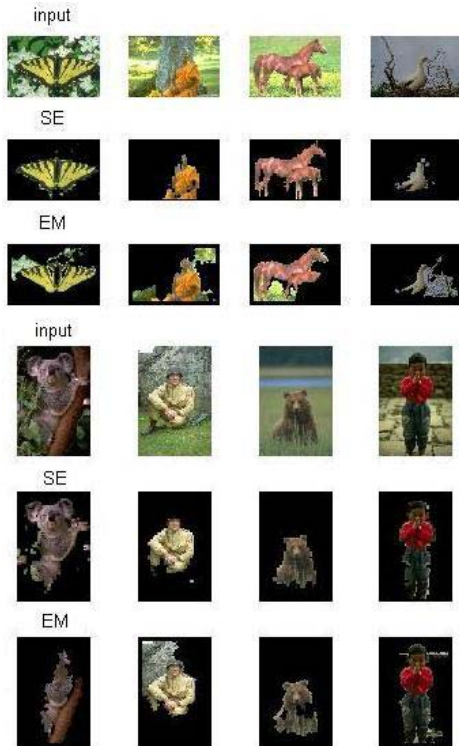


Figure 5. Examples of figure-ground discrimination results: SE vs. EM

The above equations update the assignments of a pixel by integrating evidence from the samples in its neighborhood. To obtain the final segmentation, we assign a pixel to the foreground if $\hat{P}_{fg} > \hat{P}_{bg}$, otherwise to the background. In this way, we let the figure and ground processes “compete” to explain the pixel data.

4. Experiments

We did experiments to test the sensitivity of the SE algorithm to the initial distribution. We shift the center of the initial Gaussian distribution off the center of the image and compare the results with the one obtained by locating the distribution at the image center. Fig. 3 indicates that the average assignment error at each pixel is less than 0.05 when the shift is less than 10 pixels or 12% of the figure size.

To test how sensitive the SE algorithm is to the initial sampling, we ran the SE algorithm over the same image several times. Fig. 4 illustrates that the results are very stable. In comparison, we ran the traditional EM algorithm over the same image with three mixture Gaussians for the figure and three for the background. When initializing the cluster centers at different places, we obtained very different segmentation results as shown in Fig. 4.

We further compared the SE algorithm with a more sophisticated EM algorithm proposed by Carson *et. al.* [4]. In [4], the Minimum Description Length principle is used to select the number of mixture models. Fig. 5 demonstrates that the EM algorithm tends to merge the figure with part of the background, while our algorithm gives better segmentation results.

5. Conclusion

In this paper, we present a novel segmentation method for figure-ground discrimination. The use of kernel density estimation for color distribution enables automatic selection of weights of different cues based on the bandwidth calculation from the image itself. The size and shape of the figure are determined adaptively by the competition between the figure and background using the iterative sampling-expectation algorithm. Consequently, the combination of kernel density estimation for color distribution and the iterative sampling-expectation algorithm have resulted in encouraging segmentation results.

References

- [1] L. Hérault, R. Horaud: Figure-Ground Discrimination: A combinatorial Optimization Approach. *IEEE Trans. Pattern Recognition and Machine Intelligence*, 15(9):899-914, 1993.
- [2] A. Amir, M. Lindenbaum: Ground from figure Discrimination. *Computer Vision and Image Understanding*, 1(76):7-18, 1999.
- [3] J. Shi and J. Malik: Normalized Cuts and Image Segmentation. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June, 1997.
- [4] C. Carson, S. Belongie, H. Greenspan, J. Malik, “Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying,” *IEEE Trans. Pattern Recognition and Machine Intelligence*, 24(8):1026-1038, 2002.
- [5] S. Ayer, H.S. Sawhney: Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding. *Proc. Int. Conf. Computer Vision*, 777-784, 1995.
- [6] A. Jain, R. Dubes. *Algorithms for Clustering Data*. Englewood Cliffs, NJ, 1988.
- [7] J. Puzicha, T. Hofmann, J.M. Buhmann: Histogram Clustering for Unsupervised Segmentation and Image Retrieval. *Pattern Recognition Letters*, 20:899-909, 1999.
- [8] D.W. Scott. *Multivariate Density Estimation*. Wiley Interscience, 1992.