13-9

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Extracting Object Regions Using Locally Estimated Probability Density Functions

Hidenori Takeshima† Takashi Ida† Toshimitsu Kaneko†

†Corporate Research & Development Center, Toshiba Corporation

Abstract

In this paper, a novel method for estimating a precise object region using a given rough object region is proposed. For determining whether each pixel belongs to an object or not, the proposed method estimates a joint probability density function (joint p.d.f.) of position, color, and class (object or background). For each pixel, the class with a higher joint p.d.f. is selected. The joint p.d.f. is estimated using a kernel density estimator. Since only local distributions of colors are used for estimating the joint p.d.f., the proposed method can classify pixels correctly even if the object and the background have the same color, which is the case where the conventional methods fail. In the experiments, the number of misclassified pixels estimated by the proposed method was from 12% to 76% of those extracted by conventional methods.

1. Introduction

The need arises to extract a desired region of an object from an image for various applications such as video compression, video editing and image retrieval. Regarding the extraction of objects from images, there are many reports on refining an object region given roughly to a precise region[1, 2, 3, 4]. A rough region of an object can be extracted using various detectors such as face detectors [5, 6] or inter-frame subtractors. By refining the outputs of such detectors, a desired object can be extracted without any interactions.

The purpose of this paper is to estimate a precise region of an object region using an image and a rough object region. Conventional methods are classified into the following two categories.

- To optimize some energy functions that represent parametric lines or regions [1, 2].
- To select an appropriate region for each pixel [3, 4].

The former methods cannot extract an object region precisely when its model (parametric lines or regions) doesn't match its actual form. For example, it is hard to represent sharp corners of objects using parametric lines used in Snakes[1]. Therefore, Snakes cannot extract such objects precisely.

The latter methods don't have such problem. However, conventional methods[3, 4, 7] assume that the distributions of colors are globally constant. Therefore, some pixels are misclassified when the object and the background have similar colors.

The method proposed in this paper estimates a joint probability density function (joint p.d.f.) of position, color, and class (object or background) for determining whether each pixel belongs to an object or not. Since only local distributions of colors are used for estimating the joint p.d.f., the proposed method can classify pixels correctly even when there are sharp corners in the contour of the object or when the object and the background have the same color. To estimate local distributions, Kernel Density Estimator [8, 9] is used. The proposed method can also be used to refine the multi-class labels such as labels assigned by image segmentation algorithms (e.g. [10]).

A method to estimate local distributions using Kernel Density Estimator is explained in section 2. The proposed method is explained in section 3. Experimental results are shown in section 4 and the conclusions are presented in section 5.

2. Estimating Probability Density Function in Images

Kernel Density Estimator [8, 9] (KDE) is a non-parametric estimator of the probability density function (p.d.f.). Let $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}$ be d-dimensional row vectors that represent samples generated from an i.i.d. variable $\mathbf{x} \in R^d$. In KDE, the p.d.f. is estimated as

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K((\mathbf{x} - \mathbf{x}^{(i)})/h). \tag{1}$$

The smoothness of $\hat{f}(\mathbf{x})$ is controlled by the bandwidth h. $K(\mathbf{x})$ is a kernel function that satisfies $\int K(\mathbf{x})d\mathbf{x} = 1$. For example, the uniform kernel $K_U(\mathbf{x})$ and the Gaussian kernel $K_N(\mathbf{x})$ are defined as

$$K_U(\mathbf{x}) = \begin{cases} c_U & if\ \mathbf{x}\mathbf{x}^T < 1 \\ 0 & otherwise \end{cases} \tag{2}$$

and

$$K_N(\mathbf{x}) = c_N \exp\left(-\frac{1}{2}\mathbf{x}\mathbf{x}^T\right), \tag{3}$$

respectively. The constants c_U and c_N are determined to satisfy $\int K(\mathbf{x})d\mathbf{x} = 1$.

In [10], an image pixel is treated as a sample in $(s + r)$ dimension where s is the dimension of image positions and

r is the dimension of colors. In this paper, each space is assumed to be isotropic. Let n be the number of image pixels. Let $\mathbf{x}^{(i)} = (\mathbf{x}_s^{(i)}, \mathbf{x}_r^{(i)})$ be the value of the i-th pixel where $\mathbf{x}_s^{(i)}$ represents the s-dimensional position and $\mathbf{x}_r^{(i)}$ represents the r-dimensional color. By using KDE, the p.d.f. of a variable $\mathbf{x} = (\mathbf{x}_s, \mathbf{x}_r)$ in the $(s+r)$-dimensional space is estimated as

$$\hat{f}(\mathbf{x}_s, \mathbf{x}_r) = \frac{1}{n(h_s)^s(h_r)^r} \sum_{i=1}^{n} K_s\left(\frac{\mathbf{x}_s - \mathbf{x}_s^{(i)}}{h_s}\right) \times K_r\left(\frac{\mathbf{x}_r - \mathbf{x}_r^{(i)}}{h_r}\right) \quad (4)$$

where $K_s(\mathbf{x_s})$ and $K_r(\mathbf{x_r})$ are kernel functions, and h_s and h_r are bandwidths.

3. Selecting Appropriate Class for Each Pixel

In this section, a novel method for estimating a precise object region using a given rough object region is proposed. For determining whether each pixel belongs to an object region or not, the proposed method estimates a joint p.d.f. of position, color, and class.

Let $L(\mathbf{x})$ be the class label of the pixel $\mathbf{x}$. In 2-class problems, the class label $L(\mathbf{x})$ is 1 if the pixel belongs to an object region and 0 otherwise. However, the number of objects is not limited to one in the proposed method. The number of class labels can be greater than 2.

By using KDE, a joint p.d.f. of position, color, and class is estimated as

$$\hat{f}(\mathbf{x}, L(\mathbf{x})) = \frac{1}{n(h_s)^s(h_r)^r} \sum_{i=1}^{n} \delta(L(\mathbf{x}), \hat{L}(\mathbf{x}^{(i)})) \times K_s\left(\frac{\mathbf{x}_s - \mathbf{x}_s^{(i)}}{h_s}\right) K_r\left(\frac{\mathbf{x}_r - \mathbf{x}_r^{(i)}}{h_r}\right) \quad (5)$$

where

$$\delta(\alpha, \beta) = \begin{cases} 1 & if\ \alpha = \beta \\ 0 & otherwise. \end{cases} \quad (6)$$

$\hat{f}(\mathbf{x}, L(\mathbf{x}) = \alpha)$ represents the joint p.d.f. where the pixel $\mathbf{x}$ is generated from the region whose class is α. $\hat{L}(\mathbf{x}^{(i)})$ represents the current estimate of the class label of the pixel $\mathbf{x}^{(i)}$. Each $\hat{L}(\mathbf{x}^{(i)})$ is initialized by the given rough regions.

The proposed method assigns the class label with the highest joint p.d.f. to each pixel. The proposed algorithm is described as follows.

(inputs) The kernel functions (K_s,K_r) and the bandwidths (h_s,h_r) are given as inputs.

(step 1) Set $i = 1$.

(step 2) Select class label for each pixel $\mathbf{x}^{(i)}$ by using

$$\hat{L}(\mathbf{x}^{(i)}) = \underset{j}{\operatorname{argmax}}\, \hat{f}(\mathbf{x}^{(i)}, j). \quad (7)$$

(step 3) Terminate if $i = n$. Otherwise, add 1 to i and jump to step 2.

Table 1. Numbers of erroneous pixels. KDE: the proposed method. KDE-10: the proposed method with 10 iterations.

	fish	leaf	personA	personB
Snakes	2546	3199	2751	1143
global GMM	1601	7475	9909	15359
KDE	1684	1916	3077	4466
KDE-10	1213	645	336	323

The proposed method estimates the joint p.d.f. locally for each pixel. The proposed method can assign correct class labels even if the same color is included in two or more classes as long as the pixels with the color do not locate near pixels in the different classes with the same color.

When the numbers of the erroneous pixels are reduced against the given rough regions by applying the proposed method, they are expected to be reduced further by applying it iteratively. The effectiveness of the iterative operation is confirmed experimentally in section 4.

4. Experiments

In the 1st experiments, 2-class label data that represent object and background regions were evaluated. The proposed method, Snakes, and the classifier using a global Gaussian mixture model to represent colors (referred to as global GMM in this section) instead of (5) were applied to four pairs of image and rough object region shown in Fig. 1. A fast energy minimizer[11] was used for Snakes. In global GMM, object and background regions had different Gaussian mixture models[1]. The numbers of Gaussians assigned to object regions were same as those of background regions and were selected from $\{6, 8, 10, 12, 15, 20, 40, 60\}$. The kernels used in the proposed method were selected experimentally. In these experiments, the spatial and color kernels used in the proposed method were $K_U(\mathbf{x})$ and $K_N(\mathbf{x})$, respectively. The spatial and color dimensions were $s = 2$ and $r = 3$, respectively. The spatial and color bandwidths were selected from $h_s = \{24, 32\}$ and $h_r = \{16, 24\}$, respectively.

The inputs and results are shown in Fig. 1. The numbers of erroneous pixels are shown in Table 1. The sizes of images are 320 x 240 (fish, leaf) and 320 x 288 (person). The parameters of algorithms are selected by trial and error.

In the results of Snakes, complex contours such as a tail fin of the fish image and contours of the leaf image were not extracted precisely. The models used in Snakes cannot represent such contours precisely.

In the result of the personB image of global GMM, some pixels of hair and clothing were classified to the background region. The assumption of global GMM is violated when

[1] In the experiments, no prior knowledge of smoothness is used in both global GMM and the proposed method. Some trivial errors are expected to be removed by using prior knowledge[3, 4]. Combining the proposed method with such prior knowledge is remained as future work.

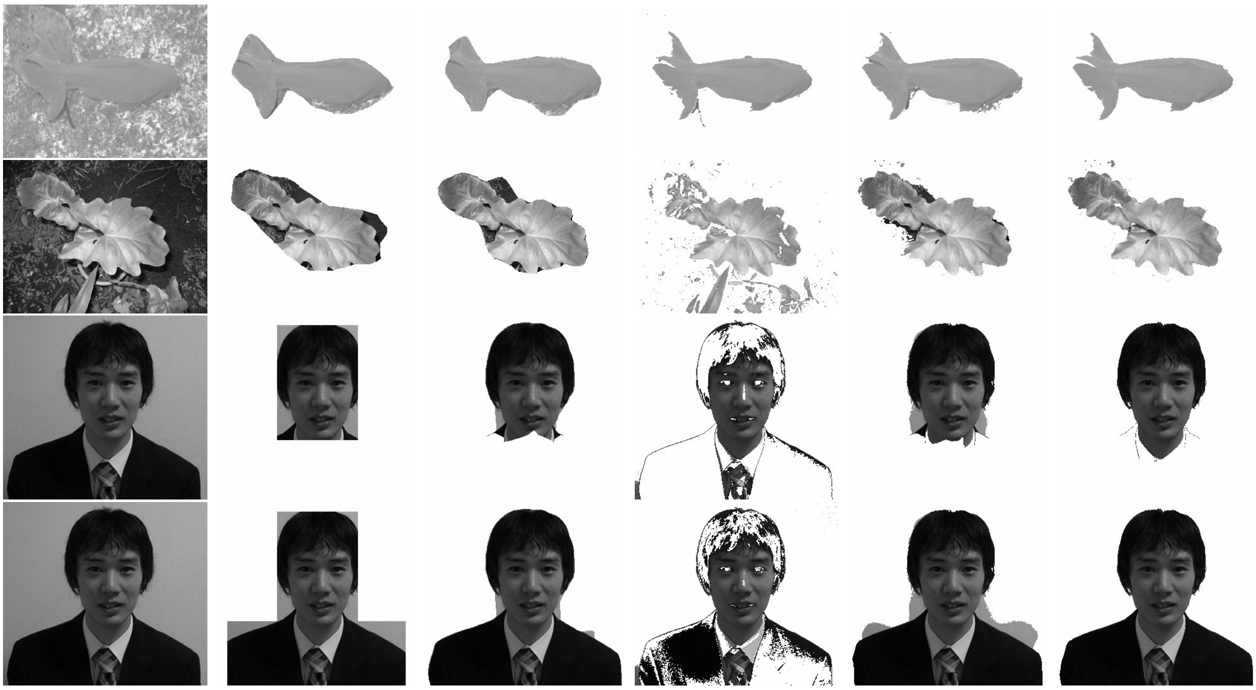

Figure 1. A comparison of different algorithms. From top to bottom: fish, leaf, personA, personB. From left to right: images, given rough regions, results of Snakes, global GMM, the proposed method, and the proposed method with 10 iterations. The background regions are filled with white.

some pixels of a desired object are marked as background in the given rough regions.

The numbers of erroneous pixels were reduced by applying the proposed method once. By applying it 10 times iteratively, the most erroneous pixels were changed to the desired values. The numbers of erroneous pixels using the proposed method were reduced to 12% - 76% of those in the case of using conventional methods (see Table 1).

4.1. Refining multi-class label data

As explained in section 3, the proposed method can be applied to multi-class label data. An example of such data is label data assigned by image segmentation algorithms. Many image segmentation algorithms (e.g. [10]) suffer from generating small regions with the same label. The proposed method can be used for merging such small regions into other regions. Unlike removing regions where numbers of pixels are less than a threshold, the proposed method does not remove small regions whose color is locally different from other regions.

In the 2nd experiments, the proposed method with 10 iterations was applied to 3 pairs of image and label data. The images were the same as those in the previous experiments and label data were generated using the Mean Shift segmentation[10].

Input labels and results are shown in Fig. 2. The numbers of their regions are shown in Table 2. As shown in the results, boundaries were preserved and small regions such as the body of the fish were removed. These results show that the proposed method can be used as post-processing of a segmentation algorithm.

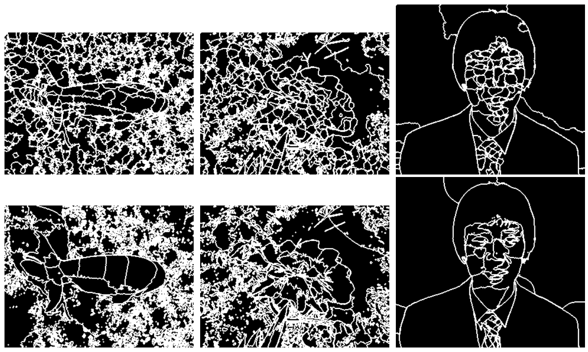

Figure 2. From left to right: fish, leaf, person. From top to bottom: given labels and results. Boundaries between different labels are drawn with white.

5. Conclusions

In this paper, a novel method for estimating a precise object region using a given rough object region is proposed. To determine the region of each pixel, a joint p.d.f. of position, color, and class (object or background). is estimated for each pixel separately. The number of misclassified pixels estimated by the proposed method was from 12% to 76% of those extracted by conventional methods. The proposed method can be also used to refine segmented images as shown in the 2nd experiments.

Table 2. Numbers of different labels.

	fish	leaf	person
inputs	3977	3163	745
results	577	586	94

In this study, the bandwidths of kernel functions were assigned experimentally. Finding a good method for estimating appropriate bandwidths automatically is a subject for future work.

References

[1] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[2] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-22(3):266–280, Mar. 2000.

[3] Y. Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proceedings of International Conference on Computer Vision*, pages 105–112, 2001.

[4] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. In *Proceedings of the 8th Europian Conference on Computer Vision*, pages 428–441, 2004.

[5] M. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-24(1):34–58, 2002.

[6] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, pages 511–518, 2001.

[7] P. Kohli and P. H. S. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *Proceedings of IEEE International Conference on Computer Vision*, pages 922–929, 2005.

[8] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman & Hall. 1994. ISBN:0-412-55270-1.

[9] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall. 1986. ISBN:0-412-24620-1.

[10] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1197–1203, 1999.

[11] D. J. Williams and M. Shah. A fast algorithm for active contours. In *Proceedings of International Conference on Computer Vision*, pages 592–595, 1990.