Discriminative Generative Contour Detection

Chao Zhang[1]
bjzchao@cn.ibm.com
Xiong Li[2]
lixiong@sjtu.edu.cn
Xiang Ruan[3]
gen@omm.ncl.omron.co.jp
Yuming Zhao[2]
arola_zym@sjtu.edu.cn
Ming-Hsuan Yang[4]
mhyang@ucmerced.edu

[1] IBM China Research, China
[2] Shanghai Jiao Tong University, China
[3] Omron Corporation, Japan
[4] University of California Merced, United States

Object contour is of prime importance as it contains essential visual information, such as shape and identity that finds numerous applications.Contour detection is a fundamental problem in computer vision which is closely related to other tasks, e.g., segmentation, shape analysis, pose estimation, visual saliency, and object recognition, to name a few.

In this work, we propose a learning algorithm for contour detection based on multi-level visual cues. We extract pixel-level features that integrate both local and global visual information. Basic point (pixel) features have been widely used for representing edge information of grayscale and color images such as image gradients, texture inhibition, brightness and color gradients as well as compass operators.As these features capture different visual information, we extract and combine these features as local visual cues for contour detection at different scales. Global information extracted from visual saliency is then incorporated to provide cues of salient objects in the scenes. These features are integrated to form effective pixel-level features to represent object contours.

In addition, segment-level features are extracted to exploit structural information of contours. Structural cues such as segments contain important information more than pixelwise evidence.Toward this goal, we compute superpixels to extract structural segments with the SLIC algorithm [1], which has been verified to perform well terms of efficiency and effectiveness. Point features are used to describe edge pixels on the line fragment, and segment-level features are then extracted by computing their mean value, variance and differences in this work.

To extract discriminative information from features, we propose a mapping method based on the posterior divergence (PD) measure over Gaussian mixture model (GMM). We transform the vectors formed by point-level and segment-level features based this mapping to obtain more discriminative information for contour detection. Let $\mathbf{x} \in \mathrm{R}^D$ be the observed random variable, i.e., the combination of multi-scale features. Let $\mathbf{z} = \{z_1, \cdots, z_K\}$ be the hidden variable, where $z_k = 1$ if the k-th mixture center is selected to generate a sample and $z_k = 0$ otherwise; $\mathbf{a} = (a_1, \cdots, a_K)^\top$ be the mixture prior satisfying; $\mathbf{u}_k$ and Σ_k respectively be the mean and variance matrix of the k-th mixture center. For any observed sample $\mathbf{x}^t$, similar to [3], we assume that the posterior distribution of $\mathbf{z}$ takes the same from with its prior $P(\mathbf{z})$ but with different parameter $\mathbf{g}^t = (g_1^t, \cdots, g_K^t)^\top$, $Q^t(\mathbf{z}) = \prod_{k=1}^K g_k^{t\, z_k}$. Let θ be the model estimated from a set of $N-1$ training samples $\mathcal{X} = \{\mathbf{x}^i\}_{i=1}^{N-1}$, and θ_{+t} be the model estimated from a set of N samples $\mathcal{X} \cup \{\mathbf{x}^t\}$. PD derives three measures, where the posterior divergence Φ^{pd} measures how much $\mathbf{x}$ affects the model, the fitness function Φ^{fit} measures how well $\mathbf{x}$ fits the model, the entropy function Φ^{ent} measures how uncertain the fitting is. The mappings according to the variable $\mathbf{x}$ can be derived as follows,

$$\Phi_x^{pd} = \sum_{i\neq t}^{N} \sum_{k=1}^{K} \sum_{d=1}^{D} g_k^i\left(-\frac{(x_d^t-u_{d,+t})^2}{2\delta_{d,+t}^2} + \frac{(x_d^t-u_d)^2}{2\delta_d^2} + \delta_{d,+t}^{D/2} - \delta_d^{D/2}\right) = \sum_{d=1}^{D} \Phi_{x_d}^{pd},$$

$$\Phi_x^{fit} = \sum_{k,d=1}^{K,D} g_k^t\left(-\frac{(x_d^t-u_{d,+t})^2}{2\delta_{d,+t}^2} + \delta_{d,+t}^{D/2}\log\sqrt{2\pi}\right) = \sum_{d=1}^{D} \Phi_{x_d}^{fit},$$

The feature mappings according to the hidden variable $\mathbf{z}$ are,

$$\Phi_z^{pd} = \sum_{i\neq t}^{N} \sum_{k=1}^{K} g_k^i \log\frac{a_{k,+t}}{a_k} = \sum_{k=1}^{K} \Phi_{z_k}^{pd},$$

$$\Phi_z^{fit} = \sum_{k=1}^{K} g_k^t \log a_{k,+t} = \sum_{k=1}^{K} \Phi_{z_k}^{fit}, \quad \Phi_z^{ent} = \sum_{k=1}^{K} g_k^t \log g_k^t = \sum_{k=1}^{K} \Phi_{z_k}^{ent}.$$

Therefore for input $\mathbf{x}^t$, we obtain a set of feature mapping:

$$\Phi^t = \mathrm{vec}(\{\Phi_{x_d}^{pd}, \Phi_{x_d}^{fit}, \Phi_{z_k}^{pd}, \Phi_{z_k}^{fit}, \Phi_{z_k}^{ent}\}_{d,k}). \tag{1}$$

With the features mapping, various classifiers within the random forest [2] learning framework can be constructed based on different representations. A random forest is an ensemble classifier consisting of numerous decision trees where the class label is determined based on the mode of the outputs by individual trees.

We evaluate the proposed algorithm on the Berkeley segmentation data set. Figure 1 shows the precision-recall curves and the F-measures with different thresholds.

Figure 2 shows one contour detection results using features with and without using the proposed feature extraction method via posterior divergence.

Experimental results bear out feature selection from multi-scale visual cues via posterior divergence with a random forest classifier facilitates effective contour detection in natural images.

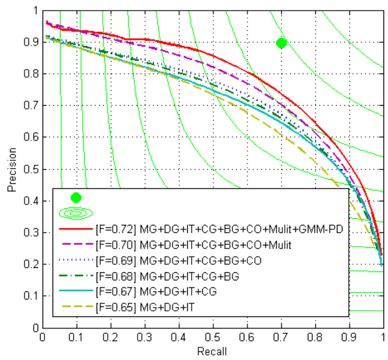

Figure 1: Precision-recall curves on BSDS500: different feature combinations.

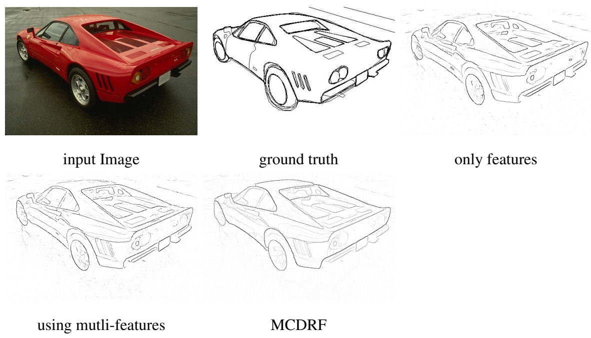

Figure 2: Contour detection results with and without feature extraction.

[1] Radhakrishna Achanta, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine SÍźsstrunk. SLIC Superpixels. Technical Report 149300, EPFL, 2010.

[2] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[3] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Sail. An introduction to variational methods for graphical models. In *Machine Learning*, volume 37, pages 183–233, 1999.