Object Detection Based on Multi-scale Contour Fragments

Xiaofeng Zhang[1,2] , Qiaoyu Sun[1] and Yue Lu [1]

1. Dept. of Computer Science and Technology, East China Normal University, Shanghai, China
2. School of Computer Science and Technology, Nantong University, Nantong, China
ntuzxf@163.com qiaoyusun@163.com ylu@cs.ecnu.edu.cn

Abstract

In this paper, we present a novel object detection scheme using the multi-scale contour fragments. The template fragments are extracted by decomposing the template contour. The multi-scale hinge angle, contour direction and partial Hausdorff distance (PHD) are used to select candidates in the edge image. Then, the matches with different scales and directions are selected by the Multiclass Discriminative Field (MDF) from the candidates. With the matches and their corresponding sample fragments, the contours of the objects can be obtained. The experiments on our postmark dataset and the ETHZ dataset show that the proposed scheme is robust to detect a class of objects with different scales, directions and complex background.

Keywords: *Object detection, contour feature, integral histogram, Hausdorff distance, Multiclass Discriminative Field.*

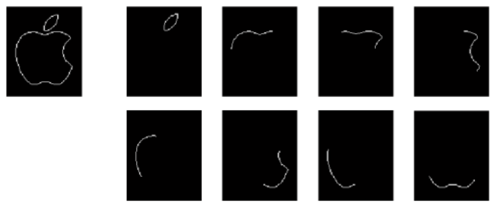
1. Introduction

Contour feature is among the most distinctive object features. As Figure 1 shows, people can easily recognize the apple logo by their shapes. This demonstrates that the contour can be used as an important clue in object detection.

Figure 1. Apple Logos with Different Size, Color and Texture

It is with this intuition that many methods [1-3] have used contour feature for object detection. V. Ferrari et al. [1] presented a family of scale-invariant local shape features (kAS) and used them for object class detection. J. Shotton et al. [2] used a multi-scale orient chamfer distance for matching contour fragments. S. Ravishankar [3] presented an object detection method using a single or few hand drawn examples as models.

Similar to [3], we use a single template contour for object detection. The template contour can be gotten from the silhouette of the object or manually drawn. As shown in Figure 2, the template contour is divided into fragments. A template fragment in our method should be long enough and have a large enough Minimum Enclosing Rectangle (MER). The multi-scale hinge angle, contour direction and partial Hausdorff distance (PHD) [4] are used to select candidates. Then, the matches are selected by the Multiclass Discriminative Field (MDF) [5] from the candidates. With these matches, the contour of the object can be composed. The experiments on our postmark dataset and the ETHZ dataset show the effectiveness of the proposed method.



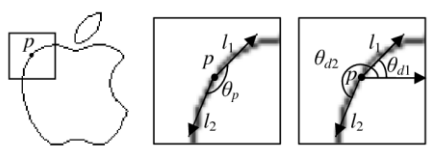
(a) (b)

Figure 2. Sample Fragments of the Apple Logo. (a) Sample Contour. (b) Sample Fragments.

2. Multi-scale Fragment Match

The information used in this section is the edge image E obtained by Canny detector. Three features are used to find the matches of fragment F_k in edge image E: hinge angle, contour direction and PHD. These features are expended to multi-scale by scale parameter s.

The hinge angle and direction are two features used in [6] for writer identification. Our definitions are slightly different from theirs. As shows in Figure 3(b) and (c), pixel p is on the contour. l_1 and l_2 are two rays proceeding from the pixel p and show the directions of contour. The hinge angle of the pixel p is θ_p. The contour directions are θ_{d1} and θ_{d2}.



(a) (b) (c)

Figure 3. The Features of the Contour. (a) The Contour. (b) The Hinge Angle θ_p of p. (c) The Directions θ_{d1} and θ_{d2} of p.

$[0\ \pi]$ is divided into n equal parts. We calculate the multi-scale hinge angle histogram as

$$h_h(s, m_1) = n_{s,m_1} \quad m_1 = 1, 2, \cdots, n, \tag{1}$$

where n_{s,m_1} is the number of the pixels whose angle is in $[(m_1 - 1)\pi/n\ m_1\pi/n]$ and s is scale (The original scale is 1). Similar to the hinge angle histogram, the multi-scale direction histogram is defined as

$$h_d(s, m_2) = n_{s,m_2} \quad m_2 = 1, 2, \cdots, 2n, \tag{2}$$

where n_{s,m_2} is the number of the pixels whose direction is in $[(m_2 - 1)\pi/n\ m_2\pi/n]$.

Then, hinge angle difference $D_h(F_k(s), E_i)$ and direction difference $D_d(F_k(\theta), E_i)$ between the fragment F_k and the edge E_i (a part of E) are calculated by using their histogram differences. The differences between two histograms can be computed in an efficient manner with their integral images [7].

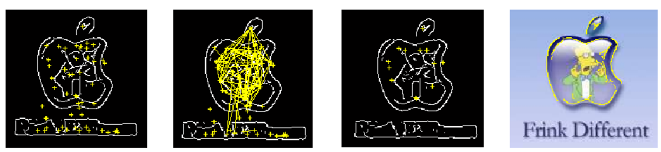
The PHD can give an accurate description of the similarity of two point sets. To detect whether E_i contains F_k, we use the PHD

$$h(F_k(s,\theta), E_i) = \frac{1}{N_{F_k(s,\theta)}} \sum_{p \in F_k(s,\theta)} d(p, E_i), \tag{3}$$

where $F_k(s,\theta)$ is the F_k with scale s and direction θ, $d(p, E_i) = \min_{p_e \in E_i} \|p - p_e\|$ and $\|\cdot\|$ is some underlying norm (In this paper, it is Euclidean distance.).

Suppose fragment F_k with scale s and direction θ is expected to match the edge E_i. The edge E_i should satisfy the following rules: (1) $D_h(F_k(s), E_i) < T_1$, (2) $D_d(F_k(\theta), E_i) < T_2$, (3) $h(F_k(s,\theta), E_i) < T_3$, (4) $h(F_k(s,\theta), E_i)$ should be a local minimum, where T_1, T_2 and T_3 are thresholds. These rules are cascade used.

Figure 4(a) shows the matches detected by these features. The crosses are the centers of E_i.



(a) (b) (c) (d)

Figure 4. An apple logo detection process. (a) Candidate matches. The crosses are centers of the matches. (b) An undirected graph for MDF. The crosses are the nodes of the graph and the straight lines are edges of the graph. (c) Matches selected by MDF. (d) An apple logo contour obtained by connecting the matches.

MDF is one of Discriminative Random Field (DRF). It is usually used in multi-class labeling problems [4]. In MDF, A and I, which called as association potential and interaction potential, are the unary and pairwise potentials, respectively. The association potential A is

$$A(x_i, y) = \sum_{k=1}^{C} \delta(x_i = k) \log P'(x_i = k|y), \tag{4}$$

where C is the number of classes, $\delta(x_i = k)$ is 1 if $x_i = k$ and 0 otherwise, $P'(x_i = k|y)$ is the probability of x_i belonging to class k under the observation y. We describe P' as

$$\log P'(x_i = k|y) = \begin{cases} 1 & \text{if } h \le T_4 \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

where h is a PHD, T_4 is a threshold. The interaction potential I is

$$I(x_i, x_j, y) = \sum_{k=1}^{C}\sum_{l=1}^{C} \nu_{kl}^T \mu_{ij}(y)\delta(x_i = k)\delta(x_j = l), \tag{6}$$

where ν_{kl} are the model parameters and $\mu_{ij}(y)$ encodes the pairwise features. We set $\mu_{ij}(y)$ as

$$\mu_{ij}(y) = \begin{cases} (|s_i - s_j| \quad |\theta_i - \theta_j|)' & \text{if } |D-1| \le T_5 \\ (0 \quad 0)' & \text{otherwise} \end{cases}, \tag{7}$$

where s is the scale, θ is the direction, $D = |D_{ij}/D_{kl}|$ denotes the difference between D_{ij} and D_{kl}, D_{ij} is the difference between edge E_i and E_j, D_{kl} is the difference between fragments F_k and F_l, T_5 is a threshold.

Using our candidate fragments as nodes, an undirected graph for MDF is generated (Figure 4 (b)). A node x_i contains the information of its location, scale s_i and direction θ_i. Each sample fragment is a category. Figure 4(c) shows the matches selected by MDF. The contour of the object in Figure 4 (d) is composed of the matches corresponding with the sample fragments.

3. Experiments

To evaluate the valid of the proposed method, we apply it to detect the template on our postmark dataset. This dataset has 80 envelope images which are categorized into two groups according to the complexities of their backgrounds. Each group has 40 images. Figure 5 shows some examples. The postmarks in the group (I) have similar scales and clean backgrounds. The postmarks in the group (II), comparatively, have various scales and complex backgrounds.

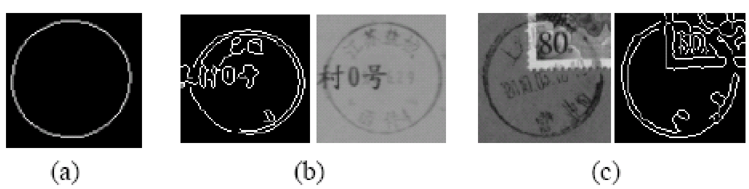


Figure 5. The template and sample postmarks. (a) The template. (b) and (c) belong to different groups according to their background complexities.

We measure the performance by $F-measure$:

$$F-measure(F) = \frac{2 \times P \times R}{P+R}, \tag{8}$$

where $R(Recall) = \frac{\text{The number of correct detected objects}}{\text{The number of true objects}}$ and $P(Precision) = \frac{\text{The number of correct detected objects}}{\text{The number of detected objects}}$.

The statistical results are shown in Figure 6. We compare our method with the method using PHD which is often used in object detection [2]. We use two template: t1 and t2. t1 is a silhouette of one postmark in group (I) and has the similar scale with the postmarks in this group, while t2 is from group (II). The parameters of MDF are trained by ten labeled images. Both of the two methods work well on group (I) with t1. In group (II), the method using PHD can not tolerate the variance of scale. Its performance degrades drastically with the template t1 and t2. Meanwhile, our method has relative high performances with both t1 and t2.

We also applied our method on ETHZ dataset [8]. Figure 7 shows some results of apple logo and bottle. Despite their deformation, backgrounds changing, color and texture variation, the rectangles cover the most parts of the objects.

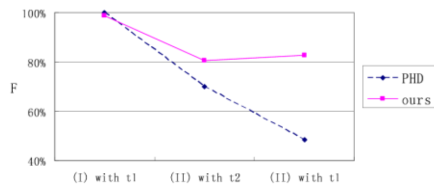


Figure 6. Postmark Detection Performance

Figure 7. Some results of ETHZ database. The first column is the sample contours provided by ETHZ database and the rest are the results of the images.

4. Conclusion

In this paper, a multi-scale object detection method is proposed. We decompose the contour into sample fragments. With the application of the multi-scale hinge angle, contour direction and partial Hausdorff distance, candidate matches corresponding to sample fragments are detected in the edge map of the image. Then, MDF selects the matches which constitute the contour of the object. Finally, contours are obtained by connecting these matches. By using multi-scale fragments, our model, which requires only a sample contour, tolerates scale invariant, rotation and complex backgrounds. The experiments show that our method is a practicable scheme.

References

[1] V. Ferrari, L. Fevrier and F. Jurie, "Groups of adjacent contour segments for object detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, **(2008)**, vol. 30, pp. 36-51.

[2] J. Shotton, A. Blake and R. Cipolla, "Multiscale categorical object recognition using contour fragments", Ieee Transactions on Pattern Analysis and Machine Intelligence, **(2008)**, vol. 30, no. 7, pp. 1270-1281.

[3] S. Ravishankar, A. Jain and A. Mittal, "Multi-stage contour based detection of deformable objects", Computer Vision. Proceedings 10th European Conference on Computer Vision, **(2008)**, pp. 483-496.

[4] B. Takacs, "Comparing face images using the modified Hausdorff distance", Pattern Recognition, **(1998)**, vol. 31, no. 12, pp. 1873-1881.

[5] S. Kumar and M. Hebert, "Multiclass discriminative fields for parts-based object detection", Snowbird Learning Workshop, **(2004)**, vol. 164.

[6] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features", IEEE Transactions on Pattern Analysis and Machine Intelligence, **(2007)**, vol. 29, no. 4, pp. 701-717.

[7] P. Ferrari, T. Tuytelaars and L. V. Gool, "Integral histogram: A fast way to extract histograms in cartesian spaces", in Computer Vision and Pattern Recognition, **(2005)**, vol. 1, pp. 829-836.

[8] V. Ferrari, T. Tuytelaars and L. Van Gool, "Object detection by contour segment networks", European conference on computer vision, **(2006)**, vol. 3953, pp. 14-28.

Authors

Xiaofeng Zhang received the B.S. degree in computer science from Soochow University, China, in 2002, the M.S. degree from Shanghai Normal University, China, in 2009. He is currently pursuing the Ph.D. degree in East China Normal University, China. His research interests focus on computer vision and its applications.

Qiaoyu Sun received the B.S. degree in computer science from Tianjin University, China, in 1996, the M.S. degree from East China Normal University, China, in 2003. She is currently pursuing the Ph.D. degree in East China Normal University, China. Her research interests focus on document image recognition and retrieval.

Yue Lu received his B.S. degree in wireless technology and M.S. degree in telecommunications and electronic system, both from Zhejiang University in 1990 and 1993 respectively, and his PhD degree in pattern recognition and intelligent system from Shanghai Jiao Tong University in 2000. In 2004, he joined the Department of Computer Science and Technology, East China Normal University, and is currently a Professor. His research interests include document image recognition and retrieval, natural language processing, biometrics, and intelligent system development.