An RST-Tolerant Shape Descriptor for Object Detection

Chih-Wen Su[1], Hong-Yuan Mark Liao[1,2], Yu-Ming Liang[3] and Hsiao-Rong Tyan[4]
[1]*Institute of Information Science, Academia Sinica, Taiwan*
[2]*Institutes of Computer Science and Engineering, National Chiao Tung University, Taiwan*
[3]*Department of Computer Science and Information Engineering, Aletheia University*
[4]*Dept. Inf. and Comput. Eng., Chung Yuan Christian University, Taiwan*
{lucas,liao,ulin}@iis.sinica.edu.tw, tyan@ice.cycu.edu.tw

Abstract

In this paper, we propose a new object detection method that does not need a learning mechanism. Given a hand-drawn model as a query, we can detect and locate objects that are similar to the query model in cluttered images. To ensure the invariance with respect to rotation, scaling, and translation (RST), high curvature points (HCPs) on edges are detected first .Each pair of HCPs is then used to determine a circular region and all edge pixels covered by the circular region are transformed into a polar histogram. Finally, we use these local descriptors to detect and locate similar objects within any images. The experiment results show that the proposed method outperforms the existing state-of-the-art work.

1. Introduction

In this paper, we propose a new local descriptor-based algorithm to detect objects in a cluttered image. Given a hand-drawn model as a query, we can detect and locate objects that are similar to the query model without learning. Our algorithm performs partial shape matching based on a local polar histogram of edge features. Different from other feature descriptors that usually use more than one scale to perform matching, the scale of our descriptor is decided based on choosing two arbitrary high curvature points (HCPs) on edges. Then, consider the edge(s) that contains these two HCPs, we utilize a polar histogram to represent the spatial distribution of the edge(s). These two HCPs decide the pose and size of each polar histogram to ensure the invariance with respect to rotation, scaling, and translation. Finally, we locate the target objects in real images by a voting process.

The contribution of this work is three-fold. First, the proposed descriptor is invariant to RST(rotation, scaling and translation) transformation by its design nature. Second, since we only include one or two edges in each descriptor, partial matching can be efficiently performed in a cluttered image. Third, we propose a new voting scheme to determine the location of a search target.

2. Previous work

The issue of detecting a specific object in a given image has been explored for many years but a satisfactory solution is still deficient. Some early approaches such as Geometric Hashing [2] and Generalized Hough Transform [3] deal with the geometric shape (mainly composed of straight lines) extraction problem. In [4], Lowe proposed a scale invariant feature transform (SIFT) which shows great robustness in finding corresponding point between two images. However, the requirement of high dimensional feature input confines its flexibility. For example, it is hard to retrieve the objects with similar structure but slightly different patterns. In comparison with the descriptors that only detect interested points, shape descriptors are more powerful for general object detection since the latter usually extracts more semantic information than the former.

Recently, more and more researchers put their research emphasis on the issue of detecting objects in a cluttered image by a shape example. Ferrari *et al.* [7] introduced the kAS family of local contour features and demonstrated its power within a slide window-based object detector. Felzenszwalb and Schwartz [8] proposed a hierarchical representation for capturing shape information from multiple levels of resolution. Zhu *et al.* [9] extended shape contexts of selected edges to represent contour contexts on multiple scales. Most of the existing approaches proposed to solve the scaling problem by sub-sampling the original image

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.193

into a number of different scales. However, there is no systematic way to determine how many scales should be chosen to derive a best solution.

3. Proposed method

3.1. The proposed descriptor

In this paper, we propose a new method to deal with the object detection problem. First of all, we use Canny edge detector [5] to extract the edge map of an image. Similar to other contour-based methods [7-9], the redundant edges in an image may significantly affect the accuracy and computation cost. Hence, in the first step we try to eliminate isolated short edges and link those edges if the distance between the end points of two neighboring edges is short.

Once the edge map of an image is extracted, we adopt the algorithm proposed by He and Yung [10] to detect the HCPs from the extracted 2D edges. It has been pointed out in [11] that an HCP plays an important role in the recognition process of a human visual system. According to the theories reported in cognitive psychology, human beings can easily recognize an object by taking only the HCPs and joining them with straight line segments [11].

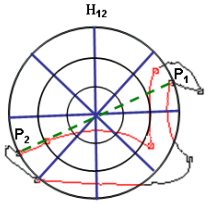

Figure 1. Green dashed line is the diameter $\overline{P_1P_2}$. Red edge segments show the edge pixels within the descriptor H_{12}.

To describe the related spatial distribution of edge(s), we use HCPs to determine the range and orientation of different descriptors. For two different HCPs, P_1 and P_2, we define a circular range C_{12} whose diameter is $\overline{P_1P_2}$. Figure 1 shows a polar coordinate mask that can be used to calculate the spatial distribution of edge pixels within C_{12}. For comparing the descriptors obtained at different scales, the number of pixels contained in each bin needs to be normalized by a corresponding diameter.

Considering the requirement of rotation invariant for some applications, one can simply use the corresponding diameter $\overline{P_1P_2}$ to adjust the orientation of the first bin of the corresponding mask towards P_1 or P_2. Figure 2 shows the two different masks (180° apart) that will result in two different polar-histograms, H_{12} and H_{21}, respectively. If the orientation is not an issue, we do not need to rotate any descriptor to make distinction between top and bottom, or left and right.

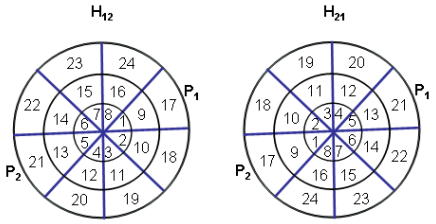

Figure 2. Rotation invariant can be achieved by using two 180° apart polar coordinate masks.

Since the number of potential descriptors is proportional to the square of the number of HCPs, redundant HCPs have to be removed to save computation time. We consider those HCPs that reside on the same edge and stay within a neighborhood distance T_L as potential redundant HCPs. For each neighborhood we choose the HCP that has the maximum curvature magnitude as a representative HCP. Through this screening process, we can remove redundant HCPs. To restrict the size of a circular mask, we set the maximum allowable diameter to be T_H.

Since the value of T_H will influence the maximum detectable size of objects in a real image, we need a strategy to deal with the situation when the distance between two HCPs is larger than T_H, we apply multi-scale sampling to convert an original image into different sizes. When at a specific resolution the largest distance between two HCPs is larger than T_H, one can take an image with smaller size to execute the algorithm. One thing to be noted is that the purpose of our multi-scale sampling process is to enhance the efficiency of the annotation and detection task. This is quite different from multi-scale sampling approaches that are designed to deal with the scaling problem.

To perform object detection in a cluttered image is very difficult due to several reasons. First, the cluttered background may severely interfere with the detection task. Second, in many cases the target objects to be detected may not look homogenous from their appearance. Their constituent components may have very different textures on the same object. Under these circumstances, for a descriptor formed by two chosen HCPs, only the edge pixels residing on the edges that cover the two HCPs will be considered valid and counted into the corresponding polar histogram. Since our method is very general, a large number of descriptors may be involved. To save computation time, some descriptors whose entropies are lower than a threshold, T_E, should be discarded. For example, when the two HCPs of a descriptor belong to two objects at a distance or they share only a straight edge inside the

descriptor, the corresponding entropy is low. The entropy E of a descriptor is defined as follows:

$$E = \sum_{i}^{N} P(i) \ln P(i) \tag{1}$$

where N denotes the number of bins of a descriptor and $P(i)$ denotes the probability of edge pixels found in the i-th bin. If edge pixels are concentrated within a small number of bins, such as the cases of a short edge or a straight edge, the descriptor will be removed.

3.2. Location Voting

Our proposed descriptor can be regarded as a scalable part-based descriptor. To locate and verify a target object in a real image, a voting process is executed to find a correct scale and location of the object. For those applications that need rotation invariant, the rotation angle, A_{ab}^{cd}, between the two corresponding diameters, $\overline{P_a P_b}$ and $\overline{P_c P_d}$, of the two descriptors, H_{ab} and H_{cd}, should be calculated. As to the scale invariant issue, the ratio R_{ab}^{cd} between the scale of H_{ab} and that of H_{cd} needs to be computed through computing the ratio between the lengths of their corresponding diameters. Therefore, the degree of similarity, S_{ab}^{cd}, between H_{ab} and H_{cd} can be determined by a modified histogram intersection operation as follows:

$$S_{ab}^{cd} = \frac{\min(H_{ab}, H_{cd})}{\max(H_{ab}, H_{cd})} \tag{2}$$

We give a vote to (A_{ab}^{cd} , R_{ab}^{cd}) if the degree of similarity is larger than a predefined threshold, T_S. The agglomerative hierarchical clustering with single linkage proposed by Hastie *et al.* [6] is used to find the group of votes concentrating within a small range, and the distance measure of clustering is defined as follows:

$$D(V_i, V_j) = \begin{cases} \infty, \ if \ \ D_R(V_i, V_j) > T_R \ \ or \ \ D_A(V_i, V_j) > T_A \\ \sqrt{D_R^2(V_i, V_j) + \alpha D_A^2(V_i, V_j)}, \ otherwise \end{cases} \tag{3}$$

where $D(V_i, V_j)$ is the distance between votes V_i and V_j. D_R and D_A are the distances corresponding to scale and angle, respectively. α is a constant weight for D_A and we set it to 0 for those applications that do not need to be rotation invariant. In our experiment, we set T_R and T_A to 0.2 and $\pi/4$, respectively. Figure 3 shows an example of scale and rotation voting. It is obvious that the votes congregated within a small range.

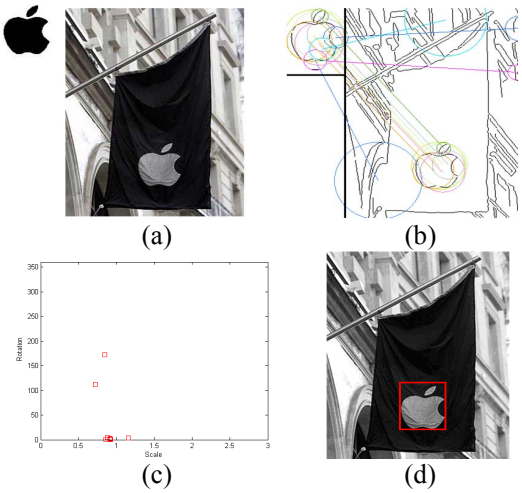

Figure 3. (a)A shape model and a test image. (b)Descriptor matching between the shape maps of (a). (c)The corresponding votes of (b). (d) The detection result.

At last, the spatial relationship between any two votes in a cluster will be further checked by looking at the overlapping area of the two descriptors. To do this, we project a rectangular boundary of model onto the image according to the relation of scale and location between the two descriptors for each vote. Suppose $B(Vi)$ denotes the projected boundary of a vote Vi in a cluster C, we remove it from C if the following function is satisfied.

$$\frac{B(V_i) \cap B(V_j)}{B(V_i) \cup B(V_j)} < T_B, \ \ \forall V_j \in C, \ j \neq i, \tag{4}$$

where T_B is set to 0.5 in our experiments. Finally, all clusters are ranked according to the number of legal votes inside. Figure 3(d) shows the detection result of Figure 3(a). Red boundary denotes the average boundary of votes for the cluster that received the highest score.

4. Experimental Result

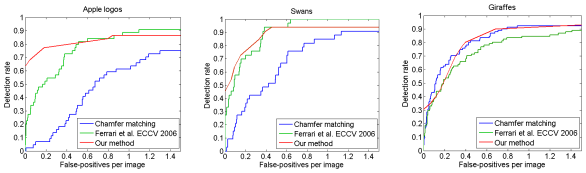

Figure 4. Object detection performance.

To test the effectiveness of our proposed method, we adopted ETHZ Shape Classes used by Ferrari *et al.* [7] to conduct experiments. This dataset contains five diverse object categories with 255 images in total and it has been widely used for testing the performance of object detection by a single hand-drawn model. The main challenges it offers are clutter, intra-class shape

variability, and scale changes. In the experiments, we only chose the models of three object categories out of the five since there are only very few HCPs in the models of the other two object categories. Figure 4 shows the comparison of the performances among chamfer matching [1], Ferrari *et al*. [7] and our method. It is clear that our method received higher detection rate with fewer false positives. In other words, our method has less false detections in higher ranked results. This characteristic can help people search more accurate returned instances from the top tens of responding detection results. Figure 5 shows part of the successfully detected instances through searching the ETHZ dataset. In these cases, the positions and the ranges of the target objects were precisely located. Note that some of the detected shapes were deformed and some of them were interfered with a cluttered background.

Figure 6 shows some unsuccessful cases. According to our observation, most of the missing detections were caused by the undetected HCPs and poor edge detection result. Sometimes over smoothed edge boundaries may reduce the number of detectable HCPs, such as the left column of Figure 6. In the middle column and the right column of Figure 6, fragmental edges and missing edges are also the causes of miss detection.

Figure 5. Some successfully detected results based on the ETHZ dataset.

5. Conclusion

In this paper, we propose a new object detection method that does not need to introduce a learning mechanism. To ensure the invariance with respect to RST, we utilize HCPs on edges to determine the region of a polar histogram and counting the related edge pixels inside. Since we only consider one or two edges in each local descriptor, partial matching can be efficiently performed in a cluttered image. Finally, we locate the target objects in real images by a voting process. The experiment results show that our proposed method outperforms the existing state-of-the-art work.

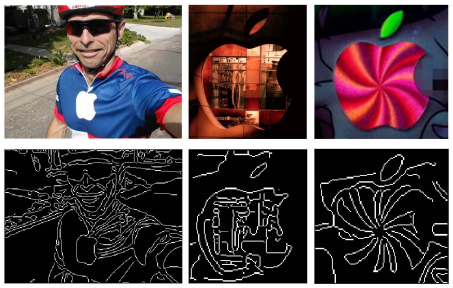

Figure 6. Some unsuccessfully cases. The upper row shows three undetected cases and the bottom row shows their corresponding Canny edge maps.

Acknowledgement

This research was supported in part by Taiwan E-learning and Digital Archives Programs (TELDAP) sponsored by the National Science Council of Taiwan under NSC Grant: NSC99-2631-H-001-020.

References

[1] D. M. Gavrila and V. Philomin, “Real-time object detection for smart vehicles,” in *IEEE International Conference on Computer Vision*, 1999.

[2] H. J. Wolfson, and I. Rigoutsos, “Geometric Hashing: An Overview,” *IEEE Computational Science and Engineering*, vol. 4, no. 4, pp. 10-21, 1997.

[3] D. H. Ballard. “Generalizing the hough transform to detect arbitrary shapes,” *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981.

[4] D. G. Lowe, “Object recognition from local scale-invariant features”. *Proceedings of the ICCV*, vol. 2. pp. 1150–1157, 1999.

[5] J. Canny, "A Computational Approach To Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-714, 1986.

[6] T. Hastie, R. Tibshirani and J. Friedman, “14.3.12 Hierarchical clustering,” *The Elements of Statistical Learning (2nd ed.)*. Springer, pp. 520–528. 2009.

[7] V. Ferrari, T. Tuytelaars, and L.J. Van Gool. “Object detection by contour segment networks.” In *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 3, pp. 14–28, 2006.

[8] P. Felzenszwalb and J. Schwartz, “Hierarchical matching of deformable shapes.” In *IEEE onference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[9] Q. Zhu, L. Wang, Y. Wu and J. Shi, “Contour Context Selection for Object Detection: A Set-to-Set Contour Matching Approach,” *ECCV*, vol. 2, pp. 774-787, 2008.

[10] X. C. He and N. H. C. Yung, “Curvature Scale Space Corner Detector with Adaptive Threshold and Dynamic Region of Support,” in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 2, pp. 791-794, Aug. 2004.

[11] J. Feldman, M. Singh, “Information along contours and object boundaries”, *Psychological Review*, vol. 112, no. 1, pp. 243-252, Jan. 2005.