OBJECT RECOGNITION BASED ON DEFORMABLE EDGE SET

Haoyu Ren   *Ze-Nian Li*

Vision and Media Lab
School of Computing Science
Simon Fraser University
Vancouver, BC, Canada
{hra15, li}@sfu.ca

ABSTRACT

We aim to solve the object recognition problem by a novel contour feature called Deformable Edge Set (DES). The DES consists of several Deformable Edge Features (DEF), which is deformed from an edge template to the actual object contour according to the distribution model of pixels. Then the DES is constructed based on the combination of DEF, where the arrangement and the deformable parameters are learned in a subspace. The RealAdaBoost algorithm is further utilized to select meaningful DES to localize the object. Experimental results show that the proposed approach not only locates the object bounding boxes but also captures the object contours well. It also achieves performance competitive with the commonly-used algorithms.

Index Terms— Deformable edge, Edge template, Matching score, PCA, RealAdaBoost

1. INTRODUCTION

Object recognition using shape information is a fundamental problem of computer vision. To describe the object shape, some researchers utilized local shape templates. Marszalek et al. [1] proposed an object recognition approach that is based on the generalizations of segmentation masks, which carried information about the extent of objects. Li et al. [9] designed the omega-shape features to describe people's head-shoulder parts for efficient pedestrian detection. Lim et al. [5] proposed the sketch tokens to both learning and detecting local contour-based representations, which are learned using supervised mid-level information in the form of hand drawn contours in images. Other researchers designed the global shape model which considered the whole object silhouette as the descriptor. Bai et al. [4] used the skeleton information to capture the main structure of the object with a tree-union structure, which had particular advantage in modelling articulation and non-rigid deformation. Toshev et al. [6] proposed a bilayer segmentation method for extraction of image regions that resembled the global properties of a model boundary structure. Eslami et al. [2] adopted a type of deep Boltzmann machine for the task of modelling foreground/background and parts-based shape images.

The local shape features are relatively consistent to illumination variance, occlusions and background noises, so that they are able to classify the target object with the background [3]. In the ideal case, the local shape features should not only describe the object contour, but also classify the background edges with the foreground edges. Most of the traditional deformable edge features solve the matching problem in a local region, so that it may mismatch to inner contours or background edges. In addition, the geometric relationship between the deformable edge features is also important. Grouping the deformable edges has certain advantage because it is consistent with the perceptual grouping and recognition of human vision. It will also contribute to the description ability of continuous contours.

Inspired by these issues, we propose the Deformable Edge Set (DES). The key idea is to encode the local shape using deformable templates and then to group them to capture higher level shape characteristics. Our contributions are two folds. Firstly, the Deformable Edge Feature (DEF) is designed based on a set of edge templates and the distribution model of pixels [7]. In addition, the DES is constructed to encode the geometric relationship between neighboring DEFs. This process is achieved by efficiently learning the DEF arrangement and model parameters in a subspace. Compared to traditional shape features, the DES focuses on the continuous contours rather than general shape characteristics. The resulting classifier trained by the RealAdaBoost algorithm works well on both the object bounding box localization and the contour detection. The experimental results on ETHZ shape dataset [8] show that the proposed method achieves promising performance on all the 5 object categories.

2. DEFORMABLE EDGES

In this section, we will introduce how to represent the local shape information by DEF. DEF extraction consists of two steps, template matching and edge deformation. In the tem-

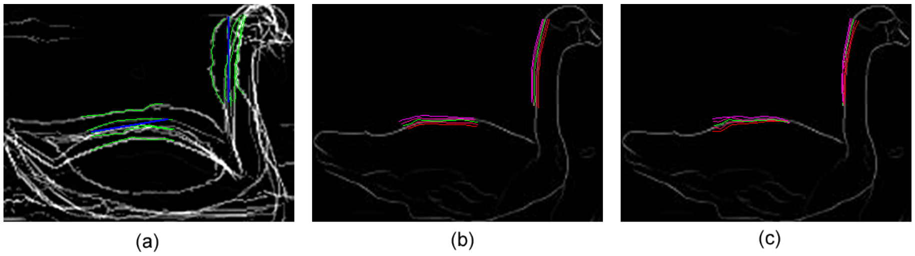

Fig. 1. Illustration of DEF. (a) Edge template (blue lines) and matched edges (green lines) in 5 swans. (b) and (c) show the deformation of the first and second principal components (the eigenvalue increases from pink to red lines).

plate matching procedure, we generate a set of lines from 6 pixels to 1/3 of the object window size as the edge templates. Given an input edge image E, each edge template t is matched to the edges in E to get the best matching result e^* following

$$e^* = \underset{|e|=|t|,e\in E}{argmin}\ D(e,t), \quad \ldots(1)$$

where $|e|$ is the length of edge e, $D(e,t)$ is the normalized distance between two edges e and t with the same length

$$D(e,t) = \frac{1}{|e|}\sum_{i=1}^{|e|}(d(e_i,t_i)+ \\ \alpha|O(e_i) - O(t_i)|). \quad \ldots(2)$$

In (2), e_i and t_i are the ith pixel of e and t respectively, $d(e_i,t_i)$ is the Euclidean distance between these two pixels, $O(\cdot)$ is the normalized orientation, and α is the constant to balance the weight of the geometric location and the orientation. Fig. 1(a) shows two templates and the matched contours in 5 swan images of ETHZ database.

In the edge deformable procedure, the distribution model of the pixels in the matched edges is utilized. For each edge template, denote the matched edge e^* in an image by a $2|e^*|$ dimensional vector based on its coordinate $\{e^*_{1,x}, e^*_{1,y}e^*_{2,x}, e^*_{2,y}, \ldots\}$, the DEF f is generated by applying PCA on all training images

$$f_{a_1,\ldots,a_p} = \mu + \sum_{i=1}^{p} a_i v_i, \quad \ldots(3)$$

where μ is the mean edge over all samples, v is the eigenvector, the parameter a_i is bounded by $|a_i| \le 2\sqrt{\lambda_i}$, and λ is the eigenvalue. The equation (3) means that DEF encodes the deformation of edges by v with the weight a_i. Fig. 1(b) and Fig. 1(c) show the deformation described by the first two components. It could be seen that the edge deforms from pink lines to red lines, which corresponds to the increasing of a_i.

In our implementation, we utilize the first three components so that it allows us to keep at least 95% energy. Given an input edge graph E and a DEF $f_{a_1,\ldots,a_p}$, the matching cost is calculated by

$$C(E,f) = \underset{i,j,k,e\in E}{min} D(e, f_{a_i,a_j,a_k}). \quad \ldots(4)$$

We quantize the a_i, a_j, a_k to 15 bins, 10 bins, and 6 bins respectively. The distance transform [3] is utilized to calculate the optimal a^*_i, a^*_j, a^*_k efficiently.

3. DEFORMABLE EDGE SET

3.1. Constructing DES

The Deformable Edge Set (DES) is defined as a set of DEF

$$F = \{f^1, f^2, \ldots, f^n\}, \quad \ldots(5)$$

where the DEFs in DES are arranged in a chain to describe the continuous contour. Each $f^i, 1 < i < n$ is adjacent to f^{i-1} and f^{i+1}, which leads to two constraints. Firstly, the two adjacent DEFs should not lay far away from each other, so one vertex of these DEFs need to be close enough. In addition, to describe the continuous contour, the other vertex of these two DEFs should not be too close. These constraints are formulated as

$$d(\mu^i_{|\mu_i|}, \mu^{i+1}_1) \le 6 \\ d(\mu^i_1, \mu^{i+1}_{|\mu_{i+1}|}) \ge \frac{1}{4}(|\mu^i| + |\mu^{i+1}|), \quad \ldots(6)$$

where μ^i is the mean of f^i over all images, μ^i_j is the jth pixel of μ^i, $|\mu|$ is the length of μ.

Given an edge map E and a DES F, the matching cost of DES is the sum of the matching cost of each DEF (4) and the cost of all adjacent DEFs

$$C(E,F) = \sum_{i=1}^{n} C(E,f^i) + \beta\sum_{i=1}^{n-1} C(f^i, f^{i+1}), \quad \ldots(7)$$

where

$$C(f^i, f^{i+1}) = |D(f^i,\mu^i) - D(f^{i+1},\mu^{i+1})|, \quad \ldots(8)$$

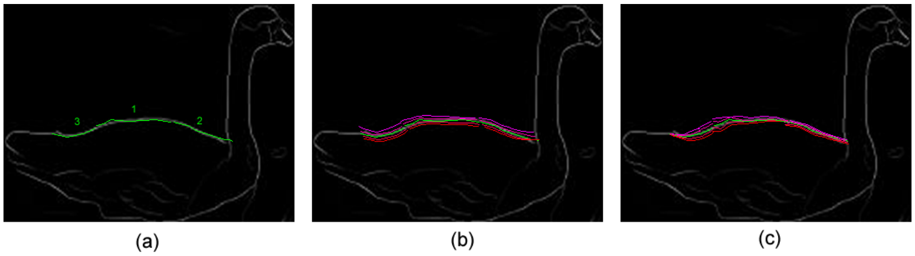

Fig. 2. Illustration of DES. (a) DES which consists of 3 DEFs (green lines). (b) and (c) show the deformation of the first and second principal components (the eigenvalue increases from pink to red lines).

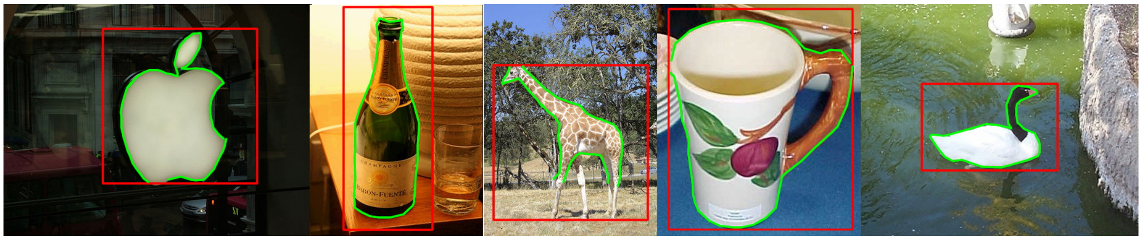

Fig. 3. Examples of object and contour location result in the ETHZ database.

β is the weight of the adjacent cost, set as 1.25 in our experiments. The adjacent cost (8) is calculated by accumulating the deformation differences of two adjacent DEFs. This cost will be small in a well-matched object since the adjacent edges in any objects has similar deformations. Fig. 2(a) gives the example of a DES which consists of 3 DEFs.

3.2. Using DES for recognition

Because we utilize 3 components in DEF extraction, for a DES $F = \{f^1, f^2, \ldots, f^n\}$, minimizing equation (7) requires us to solve the optimization problem with $3n$ parameters. Using brute force to enumerating the possible solutions is not realistic. Then we consider the $3n-$dimensional parameter space $M = \{a_{1,i}, a_{1,j}, a_{1,k} \ldots a_{n,i}, a_{n,j}, a_{n,k}\}$. Similar to the strategy use in Section 2, we move the optimization problem to the subspace by PCA. The model parameters of DES are estimated by

$$M_{b_1,\ldots,b_p} = \mu' + \sum_{i=1}^{p} b_i v_i', \quad \ldots(9)$$

where μ' is the average parameters for all training samples, v_i' is the eigenvector, the parameter b_i is bounded by $|b_i| \leq 2\sqrt{\lambda'_i}$, and λ' is the eigenvalue. We still use the top three principal components b_i, b_j, b_k and quantize them to 15 bins, 10 bins, and 6 bins respectively. As a result, the learning process of DES parameters takes $O(15 \times 10 \times 6)$ time, which is far more efficient than the brute force strategy.

In the training process, we begin with an initial DEF and incrementally grows it into a DES. In each iteration, we search all adjacent DEFs following the constraint (6). For each candidate DEF, the matching cost function (7) of adding it to current DES is updated, and the model parameter M is calculated to update the parameter subspace (9). Then the DEF with the minimum cost will be added to current DES. This process is repeated until the subspace fail to keep 95% of the energy or the total number of DEF exceeds 12. Fig. 2(b) and Fig. 2(c) illustrate a DES deformed by the first and second principal component.

We utilize DEF and DES as shape features for object recognition in RealAdaBoost [10] framework. The minimum matching cost (7) is utilized as the feature response. The objects are detected using the classifier trained to 10^{-6} false positive rate. After the detection, the object contours will be located in the detected bounding box. Given a pixel x in a detection window, the probability of x being in any edges is the sum of the confidence of DEFs including x

$$P(x) = \sum_{x \in f^i} \frac{W^+_{C_x}}{W^+_{C_x} + W^-_{C_x}}, \quad \ldots(10)$$

where C_x is the matching score corresponding to f^i, $W^{\pm}$ is the distribution on positive samples and negative samples respectively. The final contours are obtained by averaging (10) in the sliding windows of multiple scales.

4. EXPERIMENTS

We show the performance of the proposed method using the ETHZ shape dataset [8], which consists of 5 classes including applelogo, bottle, giraffe, mug and swan. This dataset is challenging since the objects appear in a wide range of scales

‘

Table 1. Comparison of Average Precision (AP) with commonly-used algorithms on ETHZ shape database.

	Felzenszwalb [14]	Ma [15]	Wang [12]	Srinivasan [13]	Li [16]	DEF	DES
Applelogos	89.1	88.1	88.6	84.5	82.3	90.1	**92.0**
Bottles	95.0	92.0	**97.5**	91.6	90.0	92.0	**97.5**
Giraffes	60.8	75.6	**83.2**	78.7	69.2	80.4	81.8
Mugs	72.1	86.8	84.3	88.8	**98.0**	86.8	89.8
Swans	39.1	**95.9**	82.8	92.2	81.0	88.2	92.6
Mean	71.2	87.7	87.3	87.2	84.1	87.5	**90.2**

‘

Table 2. Accuracy of localized object boundaries. Each entry is the AC/AP.

	Bounding boxes [7]	Ferrari [7]	DEF	DES
Applelogos	42.5/40.8	91.6/**93.9**	91.4/91.5	**92.7**/93.2
Bottles	71.2/67.7	83.6/84.5	83.7/83.1	**84.5/87.2**
Giraffes	26.7/29.8	68.5/77.3	65.0/75.2	**73.3/79.2**
Mugs	55.1/62.3	**84.4**/77.6	82.7/77.6	83.9/**80.1**
Swans	36.8/39.3	77.7/77.2	74.3/75.2	**80.2/84.4**

with intra-class shape variations. We follow the training and testing protocol in [8]. For applelogos and giraffes, 80×80 windows with 6,154 edge templates are utilized. For mugs, 100×90 windows with 6,982 edge templates are adopted. 100×60 windows and 5,610 edge templates are used for swans, while 40×100 windows and 2,772 edge templates for bottles. The α in equation (2) are set to 4 for applelogos, swans and giraffes, 6 for mugs, and 3 for bottles.

In Table 1, we compare the Average Precision (AP) of our algorithm with the commonly-used algorithms [12][13][14] [15][16]. These algorithms achieve the state-of-the-art results on this database. It could be seen that DEF shows comparable result with these algorithms. The DES achieves about 3% better accuracy compared to DEF, which also shows the best result on 2 categories, and the second best results on the other 3 categories. This signifies the advantage of grouping DEF to capture the continuous contour information.

In Table 2, we compare the performance of the boundary location with [7] using both the AP and the Average Coverage (AC) as the evaluation protocol. It shows that only using DEF, the accuracy is slightly lower than [7]. Combining DEF to DES, the accuracy is significantly improved, especially for the swans and giraffes. This result is reasonable since compared to DEF, DES is able to filter some false matches on the inner or background edges. For the swans images which include a lot of background edges in the water surface, or the giraffes images where the giraffes are surrounded by the forest and prairie, the advantage of using DES is clear. Fig. 3 shows some recognition examples and localized contours.

5. CONCLUSION

In this paper, we propose a novel local shape features set DES which consists of several deformable edge features DEF for object recognition. The local shape information is encoded by the pixel distribution model and further grouped based on the geometric relationship to describe continuous contour. The experimental results show that the boosted classifiers trained on DEF and DES work well on both the object recognition and boundary localization tasks.

6. ACKNOWLEDGEMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under the Grant RGP36726.

7. REFERENCES

[1] Marszałek, Marcin and Schmid, Cordelia, “Accurate object recognition with shape masks,” in *International journal of computer vision*. 2012, vol. 97, pp. 191–209.

[2] Eslami, SM Ali and Heess, Nicolas and Williams, Christopher KI and Winn, John, “The shape boltzmann machine: a strong model of object shape,” in *International journal of computer vision*. 2014, vol. 107, pp. 155–176.

[3] Shotton, Jamie and Blake, Andrew and Cipolla, Roberto, “Multiscale categorical object recognition using contour fragments,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008, vol. 30, pp. 1270–1281.

[4] Bai, Xiang and Wang, Xinggang and Latecki, Longin Jan and Liu, Wenyu and Tu, Zhuowen, "Active skeleton for non-rigid object detection," in *IEEE International Conference on Computer Vision*. 2009.

[5] Lim, Joseph J and Zitnick, C Lawrence and Dollár, Piotr, "Sketch tokens: A learned mid-level representation for contour and object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. 2013.

[6] Toshev, Alexander and Taskar, Ben and Daniilidis, Kostas, "Object detection via boundary structure segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*. 2010.

[7] Ferrari, Vittorio and Jurie, Frederic and Schmid, Cordelia, "From images to shape models for object detection," in *International journal of computer vision*. 2010, vol. 87, pp. 284–303.

[8] Ferrari, Vittorio and Fevrier, Loic and Schmid, Cordelia and Jurie, Frédéric and others, "Groups of adjacent contour segments for object detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008, vol. 30, pp. 36–51.

[9] Li, Min and Zhang, Zhaoxiang and Huang, Kaiqi and Tan, Tieniu, "Rapid and robust human detection and tracking based on omega-shape features," in *IEEE International Conference on Image Processing*. 2009.

[10] Schapire, R. and Singer, Y., "Improved boosting algorithms using confidence-rated predictions," in *Machine Learning*. 1999, vol. 37, pp. 297–396.

[11] Yang, X. and Latecki, L.J. "Weakly Supervised Shape Based Object Detection with Particle Filter," in *European Conference on Computer Vision*. 2010.

[12] Wang, Xinggang and Bai, Xiang and Ma, Tianyang and Liu, Wenyu and Latecki, Longin Jan, "Fan shape model for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. 2012.

[13] Srinivasan, Praveen and Zhu, Qihui and Shi, Jianbo, "Many-to-one contour matching for describing and discriminating object shape," in *IEEE Conference on Computer Vision and Pattern Recognition*. 2010.

[14] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., "Object detection with discriminatively trained part based models," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010, vol. 32, pp. 1627–1645.

[15] Ma, Tianyang and Latecki, Longin Jan, "From partial shape matching through local deformation to robust global shape similarity for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. 2011.

[16] Li, Tao and Ye, Mao and Ding, Jian, "Discriminative Hough context model for object detection," in *The Visual Computer*. 2014, vol. 30, pp. 59–69.

[17] Chui, Haili and Rangarajan, Anand, "A new point matching algorithm for non-rigid registration," in *Computer Vision and Image Understanding*. 2003, vol. 89, pp. 114–141.