Shape Band: A Deformable Object Detection Approach

Xiang Bai *
Dept. of Electronics and Info. Eng.
Huazhong Univ. of Science and Technology
xbai@hust.edu.cn; xiang.bai@gmail.com

Quannan Li,
Dept. of Electronics and Info. Eng.
Huazhong Univ. of Science and Technology
truthseeker1985@gmail.com

Longin Jan Latecki
Dept. of Computer and Information Sciences
Temple University
latecki@temple.edu

Wenyu Liu
Dept. of Electronics and Info. Eng.
Huazhong Univ. of Science and Technology
liuwy@hust.edu.cn

Zhuowen Tu
Dept. of Neurology and Dept. of Computer Science
University of California, Los Angeles
ztu@loni.ucla.edu

Abstract

In this paper, we focus on the problem of detecting/matching a query object in a given image. We propose a new algorithm, shape band, which models an object within a bandwidth of its sketch/contour. The features associated with each point on the sketch are the gradients within the bandwidth. In the detection stage, the algorithm simply scans an input image at various locations and scales for good candidates. We then perform fine scale shape matching to locate the precise object boundaries, also by taking advantage of the information from the shape band. The overall algorithm is very easy to implement, and our experimental results show that it can outperform stat-of-the-art contour based object detection algorithms.

1. Introduction

Object detection can be roughly divided into two broad categories: (1) task to detect classes of objects, e.g. face [25] and pedestrian [6]; (2) specific object retrieval and matching [2]. In the first category, a learning stage is often required to learn the detector for the class of interest. A bounding box is usually given, if an object is detected in an image, without detailed indication of the object parts. Besides the successes achieved for detecting rigid objects, such as frontal faces, this problem remains a big challenge in computer vision. The second category tries to detect/match a specific query object in a given image. This has been a long standing problem, e.g. the Generalized Hough Transform [1], and recent advances include the shape context [2], the SIFT [17], and the pyramid matching [12]. Ideally, one would have a unified systems for the two types of situations, and the difference is just in the training image set. However, such a framework dealing with general objects, with satisfactory performance, is yet not available.

In this paper we focus on the second category, to detect/match objects based on a given template. However, our objective is not just to detect/match the identical template in an input image. Rather, we hope that our system is able to capture the reasonable variation and deformation of the object in the same class. Unlike some the existing work in which query objects are highly textured [15], we emphasize shape based direction for objects without highly discriminative appearances. The problem has wide range of applications in image search, medical image analysis, and video tracking. We work on edge maps, rather than extracted interest points to explicitly explore shape based information. Our method is a shape matching and exemplar based approach without heavy learning, which is another important aspect in object detection.

Here, we propose a new method for shape-based object detection and matching: *Shape Band*. It models an object within a bandwidth of its sketch/contour. The features associated with each point on the sketch are the gradients within the bandwidth. In the detection stage, the algorithm simply scans an input image at various locations and scales

*Part of this work was done while the author was working at University of California, Los Angeles

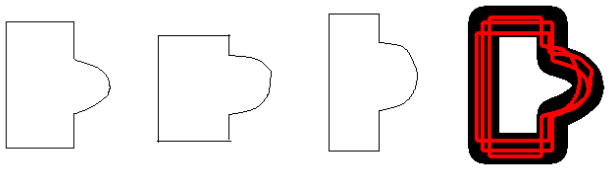


Figure 1. The proposed Shape Band allows for nonrigid contour deformations. Three different mugs (the left three contours) can be put in only one Shape Band (the right image), where the black region denotes the Shape Band, and the red curves are the transformed contours of the three mugs.

for good candidates. We view shape band as a *shape representation*, which has the tolerance for certain degree of within-class variation and deformation. Also, we can view the shape band as a search window. Though traditional sliding-window based detection algorithms use rectangles or circles as the search windows, we will show that, using the shape band as the *detection window* is more appropriate. As shown in Fig.1, although the contours of the three mugs are different, all of them can be localized in a single band when their center points are aligned. Since the global shape information is implicitly contained in the shape band, we do not need to know the order of the edges even when the local edges are not salient. Also, considering only the edge cues in the band helps to alleviate the problem of accidental alignment.

Shape band can be considered as a special window, but it is different from the typical windows using bounding boxes in [22, 9]. Given a template represented by contour/sketch, the window for shape band only considers the edges within the band, which avoids the disturbances from the other parts on the background in a cluttered image. The major advantages of the shape band are three-fold: (1) It is a simple representation but able to capture certain variation. (2) Though it is based on contour/sketch, closed object boundary is not required in detection, and thus, we avoid the complicated edge linking to get a closed contour. (3) Detecting and matching a shape using shape band can be done very efficiently, and it is not difficult to be extend to multiple templates [20, 21, 22].

2. Related Work

Early work for object detection through template matching can be dated to early 90s, e.g., the deformable template by Yuille et al. [27]. As stated before, learning is often involved to train a detector for detecting a class of object [25, 6]. Recently, a large body of research work has been proposed for shape-based object detection and matching. Shape-based approach has the advantage of being relatively robust against lighting and appearance change. A common idea to many the methods is to define a distance measures between a given shape to the objects in the image. Usually, the detection and matching steps are performed at the same time.

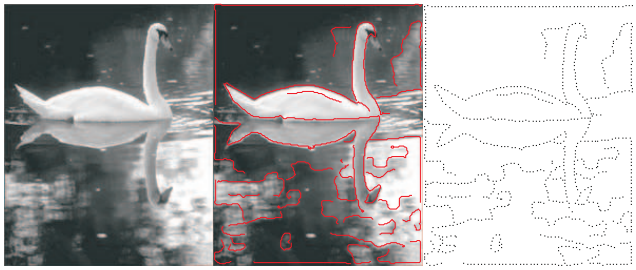
Classical methods such as Shape Contexts [2]and Chamfer Matching [23] are mostly successful in matching shapes already extracted from the images. Inner Distance Shape Contexts(IDSC) [16] improves the distance of the Shape Context for articulated shapes. However, it is not directly applicable for detecting and matching objects in cluttered images. A hierarchical shape matching method proposed by Felzenszwalb et al. [8] requires the contour points to be ordered/linked, which is not easy for real-world images. Likewise, other shape matching techniques require either good initial positions or clean images (or both) to avoid (false) local minima [19, 24, 7, 5]. In Ferrari et al. [9, 10], a shape codebook of contours is learned, followed by edge linking methods named KAS or TAS to obtain many salient segments before detection. Shotton et al. [21, 22] describes the shape of the entire object using deformable contour fragments and their relative positions. Since their distance measure using improved Chamfer Matching is sensitive to the noise, many training samples are required for boosting the discriminative shape features. Other works [29, 14, 11] decompose a given contour of a model shape into a group of contour parts, and match the resulting contour parts to edge segments in a given edge image. Zhu et al. [28] also use contour parts within a hierachical model for object detection. Recently, Wu et al. [26] proposed an Active Basis model that provides deformable template consisting of a small number of Gabor wavelet elements allowed to slightly perturb their locations and orientations.

3. Shape Band Template

In this paper, we use a contour-based template to detect objects in images and compute the similarity between the image and the template. Given an image, we extract edges [18] first and assume that most the object boundaries are contained in the edge map. The main challenge is that only a small percentage of the edge map contains edges of the target object. Further, the edges of the target objects are usually broken into many fragments and are surrounded by distractor edges resulting from texture and other objects.

Our goal is to match a template $X = \{\mathbf{x}_1, \mathbf{x}_2,...,\mathbf{x}_n\}$, where $\mathbf{x}_i$ $(i = 1, 2, ..., n)$ denote sample points of the template X, to an edge image $E = \{e_1, e_2, ..., e_{|E|}\}$ expressed as a set of edge fragments. The edge fragments are obtained by low level linking of thresholded edge pixels (we use a publicly available code [13]). The edge images are computed using [18].

We assume that the template X is centered at the origin of the coordinate system of a given image I. Based on the template X, we can easily construct the Shape Band



(a) (b) (c)

Figure 2. (a) is a gray image of a swan, (b) is the edge map of (a), and (c) shows the sample points E_s obtained from (b).

$SB(X)$ of X using a set of disks $B(\mathbf{x}_i)$ centered at each $\mathbf{x}_i$. In this paper, the radius of each disk $B(\mathbf{x}_i)$ is constant and it is equal to r.

For a reference point $c \in I$, $X(c)$ denotes the template X centered at c, i.e., $X(c) = X + c$. The shape band of template $X(c)$ can be denoted as

$$SB(X(c)) = \{y \in I | \exists \mathbf{x}_i \, ||y - (\mathbf{x}_i + c)||_2 \leq r\} \quad (1)$$

where $||\cdot||_2$ denotes the l_2 norm.

The detection and matching can be considered as a process of selecting a reference point $c^* \in I$ and a subset E^* of E such that

$$D(X(c^*), E^*) = \min_{c \in I, \, E' \subset E} D(X(c), E'), \quad (2)$$

where D is a Shape Band Distance defined is Section 4. It is worth to mention that detecting multiple objects is also possible if we consider each candidate as one with the cost smaller than a threshold.

4. Shape Band Distance

The goal of this section is to identify the possible positions and scales of the the template X on the image I with respect to edge fragments E. Since the shape of the target object in the image may be significantly different from the shape of our template X, and we do not want to miss the true location of the target object, we perform a rough shape dissimilarity estimation, which we call Shape Band Distance (SBD), for every possible center point location $c \in I$ of the template X. SBD measures the similarity of sample points on $X(c)$ to the edge points inside its shape band $SB(X(c))$. To reduce the computation, each edge fragment in E is sampled with fewer points than the original number of pixels linked to form the edge fragment. Let $E_s = \{s_1, s_2, ..., s_{|E_s|}\}$ the set of all sample points obtained by sampling the edge fragments in E. Fig. 2(c) shows an example of E_s.

Let g be a function that assigns to every point $\mathbf{z}$ a quantized gradient direction: $g(\mathbf{z}) = k$ for $k = 1, 2, ..., N$ iff the gradient direction of $\mathbf{x}$ is in the interval $[\frac{k-1}{N}\pi, \frac{k}{N}\pi)$. We experimentally determined $N = 6$ as shown in Fig. 3.

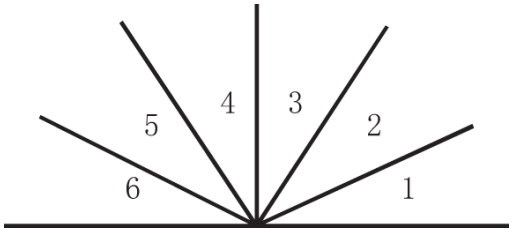


Figure 3. The angular sectors for N = 6.

Given a particular location $c \in I$ of the template X in the image I, we define the Shape Band Distance based on the similarity between the gradient direction of each template point $\mathbf{x}$ and the gradient directions of edge points within the radius of r of $\mathbf{x}$:

$$D(X(c), E_s) = \sum_{i=1}^{n} d(\mathbf{x}_i + c, E_s), \quad (3)$$

where

$$d(\mathbf{x}, E_s) = \begin{cases} 0 & \exists z \in E_s \cap B(\mathbf{x}) \;\; g(z) = g(\mathbf{x}) \\ 0.1 & \exists z \in E_s \cap B(\mathbf{x}) \;\; g(z) = g(\mathbf{x}) \pm 1 \\ 0.5 & \text{else,} \end{cases} \quad (4)$$

where ± 1 is modulo N. The equality $g(z) = g(\mathbf{x}) \pm 1$ expresses the fact that template point $\mathbf{x}$ and edge point z have adjacent (quantized) gradient directions. For example in Fig. 3, direction 6 is adjacent to 5 and 1. The meaning of equation 4 is: If there is an edge point in $B(\mathbf{x})$ with very similar tangent direction, we return 0; A relatively small value 0.1 is returned when their tangent directions are not very dissimilar, since shapes from the same class may have some locally deformation and rotation.

For a given template X, w find its optimal location in the image $c^* \in I$

$$D(X(c^*), E_s) = \min_{c \in I} D(X(c), E_s) \quad (5)$$

by computing the distance in Eq. 3 for every $c \in I$. Eq. 5 is equivalent to Eq. 2, since we can infer the subset of edges that minimizes Eq. 2 from Eq. 5, as we show below.

Since the objects from the same class may be in different scales, we use a template X in 5 scales $(w = \omega_1, \ldots, \omega_5)$ for computing SBD.

By repeating the computation in Eq. 5 for the different scales, we obtain the optimal model position $c^* \in I$ and optimal scale $w*$:

$$D(X(c^*, w^*), E_s) = \min_{c \in I, w} D(X(c, w), E_s), \quad (6)$$

where $X(c, w)$ denotes template X centered at c and with scale w.

However, in a heavily clutter image, the optimal position $c^* \in I$ of the template, may not be the true position of the target object. For the examples shown in Fig. 4(a) and (d), the optimal positions are not the true positions. Especially, in Fig. 4(d), the edge segments (in red) near to the optimal position are very similar to the input model (in green). This case can be considered as an accidental matching. Therefore, we define a distance map DM over the image I as

$$DM(c) = D(X(c, w^*), E_s) = \min_w D(X(c, w), E_s) \quad (7)$$

We then select a small set of possible target object positions in image I as the local minima of DM. The local minim are computed with respect to eight neighbors in I. We denote by $C = \{c_1, c_2, ..., c_{|C|}\}$ the set of local minima of DM, and with $W = \{w_1, w_2, ..., w_{|C|}\}$ the set of corresponding optimal scales. These are our candidate detection positions, and as demonstrated by our experimental results, the true position of a target object usually belongs to the set C; see also Fig. 4(b) and (e). In our experiments, the number of the candidate detection positions in each image is usually between 10 and 30.

5. Segments Selection and Shape Matching

While the goal of the Shape Band Distance in Section 4 was to filter out the unlikely center point positions of template X, and reduce it to a small set of likely positions C, the goal of this section is to detect true positions and scales of the target object. We achieve this goal with a global shape similarity method.

For each $c_i \in C$, we first consider all edge segments around $X(c_i, w_i)$. Then, we refine the set of these edge segments so that the shape similarity between $X(c_i, w_i)$ and the reduced set of segments is maximized.

For each edge segment e_j in E, we compute its minimum distance to the shape template $X(c_i, w_i)$, and select these segments with minimum distance below a threshold $T = \alpha r$ $(\alpha > 1)$ as the segments adjunct to the detection position. The minimum distance of e_i to X is defined as

$$minDist(e_j, X(c_i, w_i)) = \min_{\mathbf{x}_i \in X, \mathbf{p} \in K} ||p - (w_i \mathbf{x}_i + c_i)||_2, \quad (8)$$

where K is the point set consisting of the two end points and the center point of segment e_j.

If $minDist(e_j, X(c_i, w_i)) < T$, then e_j is adjunct to $X(c_i, w_i)$. For each template position $c_i \in C$, we denote the set of adjunct segments as $AS(c_i)$. One example can be seen in Fig.5

Then we evaluate the configuration of the adjunct segments of each candidate detection point c_i using the global shape context descriptor. Our goal is to select the subset of edge segments from $AS(c_i)$ that maximizes the shape similarity to $X(c_i, w_i)$. Let SC denotes the shape context distance between two sets [2]. We define the minimal shape distance as

$$SD(c_i) = \min_{S \subseteq AS(c_i)} SC(\bigcup S, X(c_i, w_i)) \quad (9)$$

and the subset $S \subseteq AS(c_i)$ that minimizes $SD(c_i)$ as $MS(c_i)$ for *matching segments* at template position c_i. The term $SC(\bigcup S, X(c_i, w_i)$ in Eq. 9 evaluates the shape context distance between the shape formed by a subset S of edge segments and template $X(c_i, w_i)$. Since each segment can be *selected* or not *selected*, there are $2^{|AS(c_i)|}$ possible configurations, which prohibits brute force computation of $MS(c_i)$.

Therefore, we propose a simple heuristics, which as our experimental results demonstrate is successful, since the set of adjunct segments $AS(c_i)$ provides a good initialization that is close to the optimal set $MS(c_i)$.

To minimize Eq. 9, we utilize a simple iteration process by deleting a segment from $AS(c_i)$ at each iteration that makes the shape context distance decrease most until the distance does not decrease any more. An example is shown in Fig. 6, since the initial segment selection in $AS(c_i)$ provides a very good initialization, the final result is correct.

If we search for a single instance of the template shape, then the final detection position fp is the one with the smallest shape distance

$$fp = argmin_{c_i \in C} SD(c_i) \quad (10)$$

Clearly, we can also set a threshold on shape distance SD to obtain multiple object instances or to conclude that the target object is not present in the image.

The whole method can be considered as a coarse-to-fine procedure for object detection. The Shape Band distance is used as the coarse process to select the candidate points of detection while the Shape Context distance is used as the fine process to decide the optimal point of detection. Candidate point selection by Shape Band distance is a process

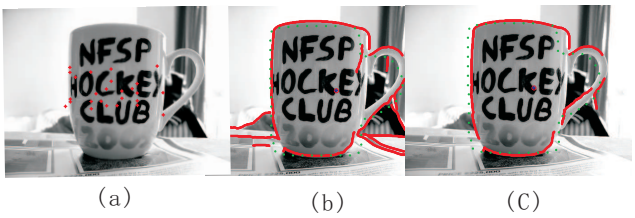


Figure 5. (a) shows the candidate center point positions (in red) of a mug template obtained as local minima of DM, (b) shows the selected segments for the center position in blue, and (c) is the final matching result based on the selected segments of (b) by minimizing Eq. 9.

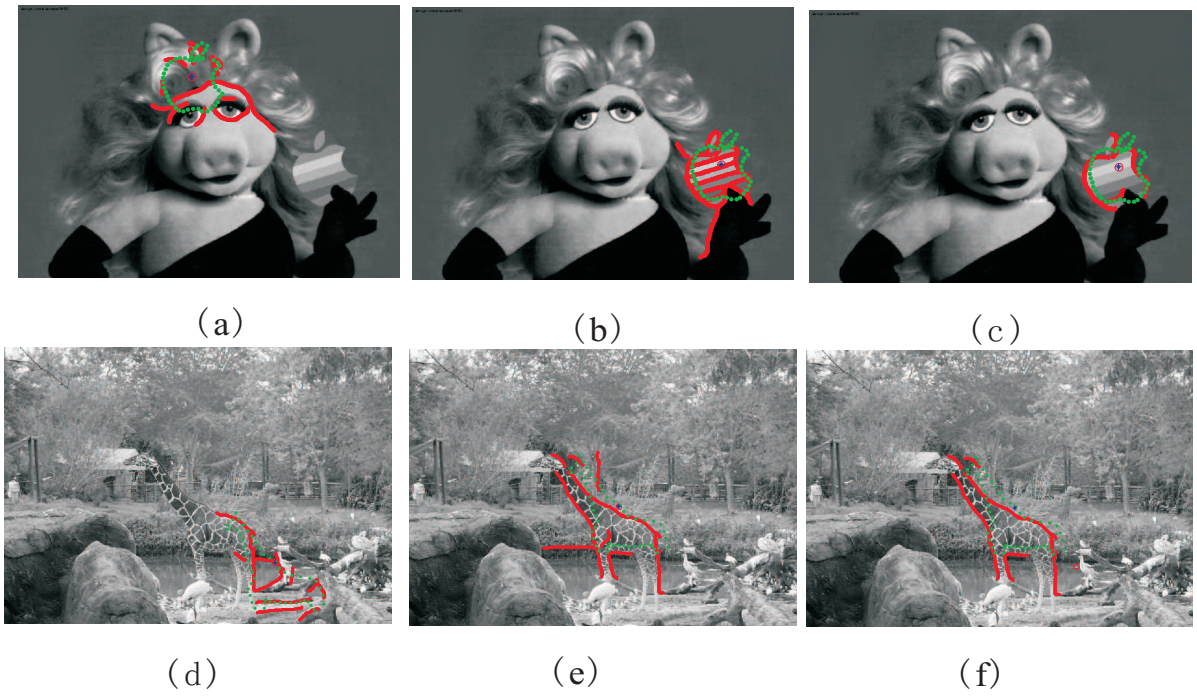


Figure 4. In each image, the green points are the points of a contour model with a reference center point (in blue), and the segments (in red) are the edge segments adjacent to the model. (a) and (d) show detection results at local minima of DM that are not the true positions of the objects. (b) and (e) show detection results with at local minima of DM that are the true positions of the objects. (c) and (f) are the final matching results.

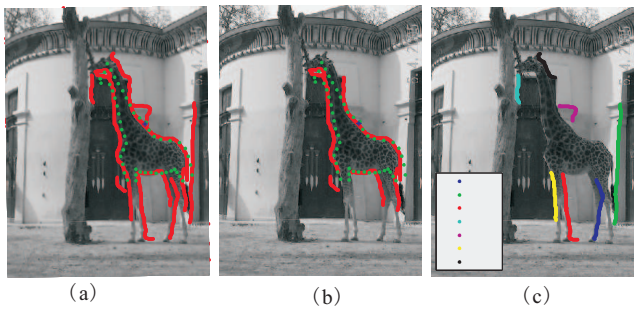


Figure 6. (a) shows a set of segments (in red) selected with a candidate center point marked in blue. The giraffe contour template is shown in Fig. 7. (b) shows the final matching results after deleting several redundant segments, and (c) shows the removed segments in different colors The colors of the points in the bottom left subimage illustrate the order of removing these segments by our algorithm.

of pruning to prune points that are clearly not points of detection. We used Shape Band as a initialization for Shape Contexts matching for two reasons: 1) Shape Band removes many redundant segments for improving the matching speed. 2) Without a good selection of segments, SC matching cannot give good results.

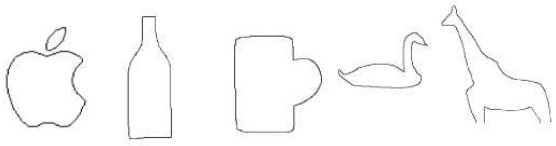


Figure 7. The contour models used as the queries in our experiments on ETHZ dataset [9].

6. Experiments

We tested our system on the challenging ETHZ dataset [9]. It has 5 different object categories with 255 images in total. All categories have significant intra-class variations, scale changes, and illumination changes. Moreover, many objects are surrounded by extensive background clutter and have interior contours. Similar to the experimental setup in [29], we use only a single hand-drawn model as shown in Fig. 7 for each class.

The results of our algorithm are summarized in Fig.11. We also compare it to the results in [10] and in [29]. Precision vs Recall (P/R) curves are used for quantitative evaluation. To compare with the results in [10] that are evaluated by detection rate (DR) vs. false positive per image (FPPI),

we use the translation of their results into P/R values done by [29]. As argued in [29], P/R is a more objective measure than DR/FPPI, because DR/FPPI depends on the ratio of the number of positive and negative test images and hence is biased. Fig. 11 shows P/R curves for the Contour Selection method in [29] in black, the method in [10] in green, and our method in blue. Our approach is significantly better than the method in [10] on all five categories, and it significantly outperforms the Contour Selection method [29] on four categories.

We also compare the precision to [29] and [10] at same recall rates in Table 1. The precision / recall pairs of [29] and [10] are quoted from [29]. Notice that our result for class Mugs are much better than the others, the main reason is that for the class of Mug, there are many clutters inside their boundaries. Using Shape Band allows us to ignore the clutters inside the contour, e.g., see Fig. 5, which is not possible when using classical rectangular windows.

From the sample results shown in Fig. 8, we can observe that even though the contour templates are not very similar to the objects, the final results are very robust and the our method performs well even if target objects are slightly rotated. The last row in Fig. 8 shows some of our worst examples, but we observe that the output edges are similar to the templates.

Some details about the experiments are mentioned here: the number of the sample points n for each template is 50, the radius of each band r is set as 30, and α is 1.5. For N, if N is smaller, more local minima will be found but can not be distinguished, while if N is larger, less local minima will be found and the true position maybe be missed. In our experiments, we found $N = 6$ is appropriate. Although scanning with Shape Band is very fast (no more than 3 seconds), the average time cost for the whole detection framework is between 1 and 2 minutes due to the iterative matching with Shape Contexts.

In addition we show that Shape Band has a potential to be applied in part-based detection/matching. Fig. 9 shows the contour parts used for testing on Weizman Horse dataset [3]. We use Canny edges [4] as the edge maps here. The detection results are shown in Fig. 10, and we notice that the part templates and the corresponding parts of the objects are not equal. We did not use Shape Contexts for the refinement here, since we want to show that Shape Band Distance is a promising direction in multi-template, contour part based object detection. For horse images, n is 25 for each contour part, and r is 20, the other parameters are the same with before.

7. Conclusion

In this paper, we exercise the idea of performing shape-based object detection and matching using a single template. We have proposed a new representation, *shape band*,

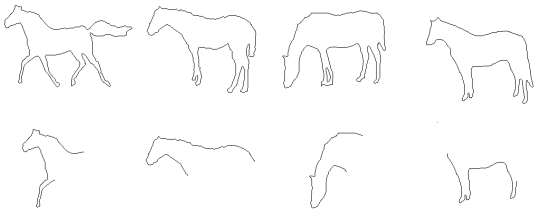


Figure 9. The second row shows the contour parts used as the inputs for part-based detection, which are taken from the original contours in the first row.

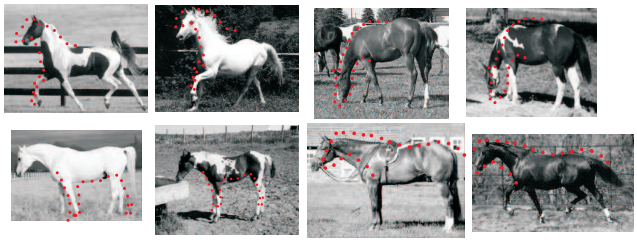


Figure 10. Example detection results for contour parts as input templates. The points (in red) denote part templates at the positions with the minimal Shape Band Distance.

which models an object within a bandwidth of its contour/sketch and is natural to think of. The shape band can also be viewed as another type of window for detection. It is robust against large degree of intra-class appearance variation and certain extent of deformations. Our algorithm is related to the shape context algorithm [2] and the active basis work [26]. We show significantly improved results for object detection and matching over many the existing algorithms. It takes, typically, a few seconds in the detection stage. Our algorithm is easy to implement and has a wide range of vision applications. It has the particular advantage of working on cluttered images with extracted edges in fragments. In addition, we also show promising results using shape band for detecting object parts. One of our future directions is to build a graphical model, still under a single template, on object parts and detect objects of articulation and non-rigid deformation.

Acknowledgements

This work is in part supported by Office of Naval Research Award, No. N000140910099. Any findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Office of Naval Research. This work was also supported in part by NSF Grant IIS-0812118, DOE Grant DE-FG52-06NA27508, NSFC(No.60873127), NSFC(No.60803115) and Education Ministry Doctoral Re-

Figure 8. Selected results on ETHZ dataset [9]. The points (in green) denote the contour template, and the segments (in red) are the final obtained edges. The last row are selected worst results.

search Foundation of China (No.20070487028). Xiang Bai's research is in part supported by MSRA Fellowship.

References

[1] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition

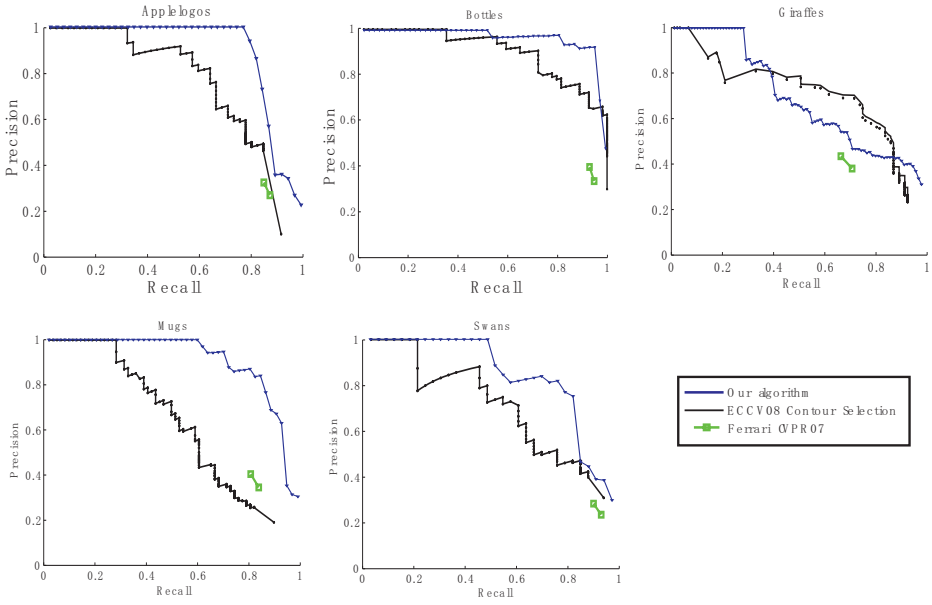


Figure 11. Comparison of Precision vs Recall (P/R) curves on the ETHZ dataset.

	Apple logos	Bottles	Giraffes	Swans	Mugs
Our precision/recall	**58.8%/86.4%**	**95.0%/92.7%**	56.0%/70.3%	**44.1%/93.9%**	**83.3%/83.4%**
Precision/recall in [29]	49.3%/86.4%	65.4%/92.7%	69.3%/70.3%	31.3%/93.9%	25.7%/83.4%
Precision/recall in [10]	32.6%/86.4%	33.3%/92.7%	43.9%/70.3%	23.3%/93.9%	40.9%/83.4%

Table 1. Comparison of the precision at the same recall rates.

using shape contexts. *PAMI*, 24(4):509–522, 2002.

[3] E. Borenstein, E. Sharon, and S. Ullman. Combining top-down and bottom-up segmentation. In *IEEE Workshop on POCV*, 2004.

[4] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI*, 8(6):679–698, 1986.

[5] J. M. Coughlan and S. J. Ferreira. Finding deformable shapes using loopy belief propagation. In *ECCV*, 2002.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of CVPR*, pages 886–893, 2005.

[7] P. F. Felzenswalb. Representation and detection of deformable shapes. *IEEE Trans. PAMI*, 27(2):208–220, 2005.

[8] P. F. Felzenszwalb and J. D. Schwartz. Hierachical matching of deformable shapes. In *CVPR*, 2007.

[9] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *PAMI*, 30(1):36–51, 2008.

[10] V. Ferrari, F. Jurie, and C. Schmid. Accurate object detection combining recognition and segmentation. In *CVPR*, 2007.

[11] D. Gavrila. A bayesian, exemplar-based approach to hierarchical shape matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(8):1408–1421, 2007.

[12] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *Proc. of ICCV*, 2005.

[13] P. D. Kovesi. MATLAB and Octave functions for computer vision and image processing. School of Computer Science & Software Engineering, The University of Western Australia, 2008. Available from: <http://www.csse.uwa.edu.au/∼pk/research/matlabfns/>.

[14] L. Latecki, C. Lu, M. Sobel, and X. Bai. Multiscale random fields with application to coutour grouping. In *NIPS*, 2008.

[15] S. Lazebnik, C. Schmid, and J. Ponce. A maximum entropy framework for part-based texture and object recognition. In *Proc. of ICCV*, pages 832–838, 2005.

[16] H. Ling and D. Jacobs. Using the inner-distance for classification of articulated shapes. *PAMI*, 29(2):286–299, 2007.

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[18] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, colour and texture cues. *PAMI*, 26(5):530–549, 2004.

[19] G. McNeill and S. Vijayakumar. Part-based probabilistic point matching using equivalence constraints. In *NIPS*, 2006.

[20] A. Opelt, A. Pinz, and A. Zisserman. A boundary-frament-model for object detection. In *ECCV*, 2006.

[21] J. Shotton, A. Blake, and R. Cipolla. Contour-based learning for object detection. In *ICCV*, 2005.

[22] J. Shotton, A. Blake, and R. Cipolla. Multi-scale categorical object recognition using contour fragments. *PAMI*, 30(7):1270–1281, 2008.

[23] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla. Shape contexts and chamfer matching in cluttered scences. In *CVPR*, 2003.

[24] Z. Tu and A. Yuille. Shape matching and recognition using generative models and informative features. In *ECCV*, 2004.

[25] P. A. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[26] Y. Wu, Z. Si, H. Gong, and S. Zhu. Active basis for modeling, learning and recognizing deformable templates. *IJCV*, 2009.

[27] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *IJCV*, 8(2):99–111, 1992.

[28] L. Zhu, C. Lin, H. Huang, Y. Chen, and A. Yuille. Unsupervised structure learning: Hierarchical recursive composition, suspicious coincidence and competitive exclusion. In *ECCV*, 2008.

[29] Q. Zhu, L. Wang, Y. Wu, and J. Shi. Contour context selection for object detection: A set-to-set contour matching approach. In *ECCV*, 2008.