Recovering Missing Contours for Occluded Object Detection

Ge Guo, Tingting Jiang, *Member, IEEE*, Yizhou Wang, *Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*—One difficult problem in practical applications is the corrupted or missing data frequently encountered in digital images. It introduces great challenges to the tasks such as object detection. This letter provides new methods for recovering missing object contours and detecting occluded objects. First, we propose an efficient contour reconstruction approach according to the Bayesian rule, utilizing global shape prior knowledge. Second, the contour reconstruction is applied to a robust detection framework for occluded objects. Based on the observed broken curves we iteratively recover object contours and propose object candidates. The experimental results demonstrate the high detection performance, localization accuracy and great advantages of our method for severe occlusion cases.**

***Index Terms*—Missing data recovery, object detection, occlusion, shape reconstruction.**

I. Introduction

In digital image processing, one frequently encountered problem is the involvement of corrupted or missing data due to occlusion, degradation and so on. The incomplete data increase the difficulties in the applications of image analysis, object detection, scene understanding, etc. For example, the contour-based object detection is an active research topic to find object contours from the cluttered digital signals in real images. However, in case of large occlusion, the unseen parts will raise the risk and cost to detect the occluded objects.

In the literature of contour-based object detection, there have been a large body of researches to recognize and locate target object outlines in edge maps, by the methods such as contour grouping [12], voting by parts [16], [24] and shape matching [13], [19], [22]. However, the problem of detecting incomplete objects has not been well solved by the state-of-the-art approaches. The missing parts, which result in the changes of object shapes, may greatly degrade the performance of contour grouping and shape matching, reduce the votes for candidates, and affect the evaluation of potential objects.

Manuscript received March 12, 2012; revised May 22, 2012; accepted June 02, 2012. Date of publication June 06, 2012; date of current version June 14, 2012. This work was supported in part by the National Science Foundation of China ((61103087, 61121002)) and by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiao-Ping Zhang.

G. Guo is with Key Lab of Intelligent Information Processing, Institute of Computing Technology and also with Graduate University, Chinese Academy of Sciences, Beijing, China (e-mail: gguo@jdl.ac.cn).

. Jiang, Y. Wang, and W. Gao are with National Engineering Lab for Video Technology, Institute of Digital Media, Key Lab of Machine Perception (MoE), School of EECS, Peking University, Beijing, China (e-mail: ttjiang@pku.edu.cn; Yizhou.Wang@pku.edu.cn; wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2012.2203592

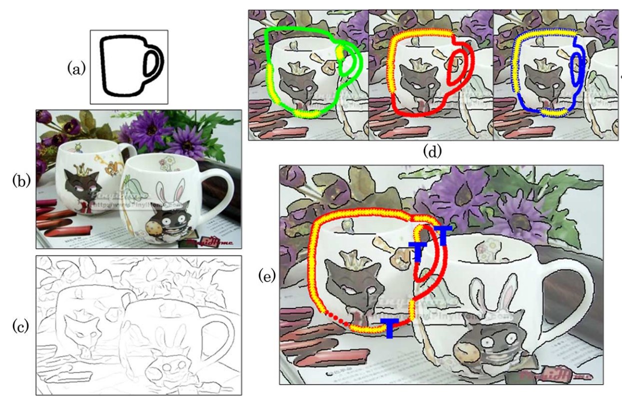

Fig. 1. Occluded object detection by recovering missing parts. (a) A hand-drawing target object template. (b) A test image and (c) its edge map. (d) Examples of three collaborative object candidates recovered based on the edge curves marked in yellow. (e) A detected object by iteratively combining edge curves of collaborative candidates and recovering new candidates. (The 'T' marks denote the detected T-junctions suggesting the occlusion boundaries.).

In this letter, we resort to contour recovery to improve object detection with occlusion. The motivation comes from the powerful reconstruction ability of human vision, which is used to rapidly recover a whole shape when it is partially occluded [18]. This facilitates object recognition. Inspiringly, it is found that recovering the missing parts do benefit the detection of target object contours in challenging situations.

Data reconstruction is involved in a wide range of applications, e.g. signal reconstruction [1], [5], image super-resolution [6], [23], surface reconstruction [8], [21], etc.

This letter focuses on the reconstruction of edge curves/object contours from the edge maps of images. This problem is related to the curve completion topics studied from several decades ago. The fundamental problem is to compute the optimal curves that fill in the missing contour parts. A variety of curves have been utilized, and some simple generic constraints have been proposed such as isotropy, smoothness, extensibility and locality [20]. Moreover, a widely used criterion is the curvature-based constraint, such as the *elastica* model [15] and *Euler spirals* [10]. Although these approaches are good at completing simple and smooth contours, they are usually not intelligent enough to recover various contours and curve types, and especially large portions of missing parts (Fig. 1, 2).

There is another group of studies on the amodal completion, in which the occluders are blended with the background [9]. The researchers are more concerned about generic visual cues for *grouping*, e.g. continuation and proximity, rather than shape completion. In addition, grouping-based curve completion is proposed to find salient contour boundaries from cluttered edges [17]. However they usually deal with small breaks; large occlusion may be difficult to handle.

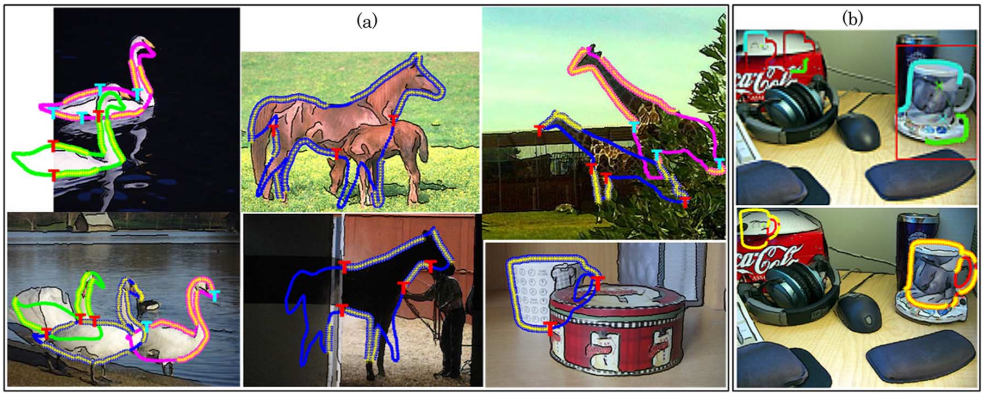

Fig. 2. (a) Results of our detected objects and recovered contours. The yellow marked curves are the observed contours in the edge map and the pink, green, blue contours are the recovered missing parts. 'T' denotes a detected T-junction, which suggests there is an occlusion boundary. (b) Comparison of results from [19] (top) and our method (bottom). The small mugs at the left-top of each image show the templates and the correspondences.

There are two main contributions in this work. Firstly, an efficient contour reconstruction method is proposed under a Bayesian framework, utilizing stronger global shape prior knowledge. The optimal complete curve is expected to fit the observed part as well as possible, and meanwhile satisfy the prior constraint. Secondly, the contour recovery is applied to occluded object detection. Complete contours are reconstructed based on edge curve segments, and considered as object candidates which vote for object locations. The candidates of similar positions and scales are collected as collaborative ones. And iteratively, we combine more collaborative curves, reconstruct object contours and vote for objects. At last occlusion relation is inferred to finalize the detected objects. Experimental results demonstrate that contour reconstruction facilitates the detection, with more robust voting and better detection performance under occlusion.

II. Recovering the Missing Contours

Given the observed curve segment P, which is considered as a part of a complete contour Q. The purpose is to estimate Q in a perceptually consistent manner, i.e. the shape of Q is unexpected to be weird, and should be consistent with P as its part. This is an under-constraint problem, which is difficult to solve, especially in the case of severe occlusion. According to the Bayesian theory, the inference of Q can be formulated as a Maximum a Posteriori (MAP) problem,

$$\max p(Q|P) \propto \max p(P|Q)p(Q) \tag{1}$$

where $p(P|Q)$ is the likelihood probability of P being a part of Q, and $p(Q)$ is the prior term of Q.

Theoretically Q can be any of 2-D curves, but it is almost computationally intractable. Previous work usually assumes that the filling-in curves belong to some categories of smooth curves, i.e. only smooth curves have non-zero priors. However in real applications, there are always large part of complex shapes other than simple smooth curves. Therefore traditional methods have obvious limitations in such situations.

Inspired by human visual experiences, we propose to reconstruct contours with our knowledge on common object shapes. For instance, if a partially occluded swan contour is observed, we can imagine what the rest part would be like with our knowledge of the swan shape. It is unlikely to reconstruct a visually unexperienced shape. Therefore, our prior assumption is that Q is inside the common object shape space Ω_S.

$$p(Q|P) = \int p(Q|P,T)p(T)dT, \quad T \in \Omega_S \tag{2}$$

where the template T provides the shape priors for reconstruction, and $p(T)$ follows a uniform distribution. For any considered reference shape, the reconstruction probability is

$$p(Q|P,T) \propto p(P|Q,T)p(Q|T) \tag{3}$$

There are two aspects included in this formulation. (i) The reconstruction prior $p(Q|T)$ encourages the reconstruction estimation Q to follow the shape of T; and (ii) the likelihood $p(P|Q,T)$ models the consistency between the partial observation and the global reconstruction. P is considered corresponding to certain part of T, hence a hidden variable l is introduced to indicate the corresponding location on T.

$$p(Q|P,T) \propto \int p(P|Q,T,l)p(Q|T)p(l|T)dl \tag{4}$$

According to the prior constraint, Q is modeled as a transformed version of T, through transforms such as rotation, scaling, stretching and bending. And the transform cost is expected to be as small as possible.

$$Q = \psi(T), \quad p(Q|T) \propto e^{-E(\psi)} \tag{5}$$

where ψ represents shape transforms. Inspired by the successful thin-plate-spline (TPS) model [2], we formulate ψ by rigid and non-rigid transforms (denoted by A and W respectively).

$$\psi(T) = T \cdot A + K_T \cdot W, \quad E(\psi) = trace(W' K_T W) \tag{6}$$

Here K is the kernel matrix defined in the TPS model, each entry $K_T^{ij} = \|(x_i,y_i) - (x_j,y_j)\|^2 log\|(x_i,y_i) - (x_j,y_j)\|$, where (x,y) is the point coordinates of T. This model embeds the point-set's internal structural relationship to constrain the non-rigid transforms.

For the likelihood, P corresponds to the part T_l, and should be consistent with the corresponding part of Q_l, where $Q = [Q_l Q_{\bar{l}}]'$,

$$p(P|Q,T,l) = e^{-C_1(P,Q)-\beta C_2(P,T_l)} \tag{7}$$

$$C_1(P,Q) = \|P - Q_l\|^2 + \|A_l - A\|^2$$

$$C_2(P,T_l) = trace\,(W_l' K_{T_l} W_l) \tag{8}$$

where A_l and W_l are the transforms from T_l to P, and Q_l is derived from the decomposed form according to (5), (6).

$$P = T_l \cdot A_l + K_{T_l} \cdot W_l, \tag{9}$$

$$\begin{bmatrix} Q_l \\ Q_{\bar{l}} \end{bmatrix} = \begin{bmatrix} T_l \\ T_{\bar{l}} \end{bmatrix} \cdot A + \begin{bmatrix} K_{T_l} & B \\ B' & K_{T_{\bar{l}}} \end{bmatrix} \cdot \begin{bmatrix} W_l \\ W_{\bar{l}} \end{bmatrix} \tag{10}$$

Instead of directly inferring Q, our goal is to find the best correspondence l and optimal transforms A and W, which are used to obtain Q by transforming T accordingly ((10)). The optimization problem is to minimize the total energy,

$$(A^*, W^*, l^*) = \arg\min_{A,W,l} \widetilde{E} = \arg\min_{A,W,l} (\alpha E(\psi) + C_1 + \beta C_2) \tag{11}$$

The above problem is difficult to solve. However notice that if l is known, we can get analytical solutions by the least squares. Therefore, we first find possible correspondences for which the matching score of T_l and P is below a predefined threshold. And then with each candidate l, we compute the optimal transforms and reconstruct the shape Q. Finally we choose the one that minimizes the energy. The parameters α, β balance the energy terms ($\alpha = \beta = 0.1$).

In the following section we apply the contour reconstruction to object detection. A hand-drawing object outline is given as the target, which just provides the reference template for contour recovery.

III. Detection of Occluded Objects

As the challenges of discovering targets with missing data, the proposed reconstruction method is particularly important to cope with such situations. On one side, to facilitate the detection of occluded contours, we use the inferred complete contours by the reconstruction from the original partial curves to vote for objects. On the other side, to suppress the false positives, a potential occluded object contour should be supported by heuristics such as occlusion boundaries.

The work flow of the proposed detection framework is as follows. The first step is *edge extraction*, in which the sophisticated edge detectors as in [14] is adopted to generate edge maps, and a number of curve segments are extracted. The second is the *recovery proposals of missing parts*. In this step, the curve segments are taken as potential parts of target object contours. We perform the reconstruction method as suggested in Section II, to recover missing parts given the curve segments as observations. The reconstructions propose object candidates which may be occluded. Then thirdly we implement the step *voting for occluded hypotheses*. Different from the traditional Hough transform methods (such as in [11]), we combine the reconstruction into this successful framework.

$$\begin{aligned} p(O, \mathcal{P}|T) &= \sum_i p(O|Q_i, \mathcal{P}_i, T) p(Q_i|\mathcal{P}_i, T) \\ &= \sum_i p(O|Q_i) p(Q_i|\mathcal{P}_i, T) \end{aligned} \tag{12}$$

where O denotes the estimation of object position; it is voted by a set of reconstructed complete contour proposals Q_i, each one voting independently; Q_i is recovered from the observed partial curves $\mathcal{P}_i = \{P_{i_1}, P_{i_2}, \ldots, P_{i_n}\}$; T is a given target contour template. Here $p(O|Q_i)$ is the probabilistic Hough votes. The voted center and scale are derived from the reconstruction since we have estimated the transformations as in Section II. Each vote is weighted by the confidence of the reconstruction, which is defined by the amount of non-occluded part of the contour. $p(Q_i|\mathcal{P}_i, T)$ is the reconstruction probability defined in (3).

There is an iterative process of the second and third step to gradually combine the collaborative parts, whose reconstructions vote for the same object. Initially only a few curves are in the set $\mathcal{P}_i$ to reconstruct Q_i in the second step. And after the voting in the third step, we collect those collaborative parts for the next iteration.

The final step is *hypothesis confirmation* with occlusion relation interpretation. For each hypothesis, we compute the coverage of the object contour, i.e. the rate of the curve length of the non-occluded contour part v.s. that of the total contour. If the coverage $C(\mathcal{P})$ is below a threshold ($\theta = 0.8$ in our experiments), we assume the hypothesis is a potential occluded object, and we should check whether there exist reasonable occlusion boundaries. There are two aspects for hypothesis evaluation, computed by the reconstruction score and the occlusion score:

$$S(x) = S_r(Q) + S_o(Q) \tag{13}$$

$$S_r(Q) = \log p(Q|\mathcal{P}, T) \tag{14}$$

$$S_o(Q) = \begin{cases} \frac{1}{K} \sum_k (\mathbf{1}(J_{k_1}) + \mathbf{1}(J_{k_1}) - 2) & if\ C(\mathcal{P}) < \theta \\ 0 & otherwise \end{cases} \tag{15}$$

For an occluded contour segment $b_k \in \{Q - \mathcal{P}\}$ we check its two endpoints J_{k_i}, $i = 1, 2$. An indicator function $\mathbf{1}(J_{k_i})$ is used to describe whether there is a occlusion boundary at the endpoints. (Those very short b_ks are unnecessary for check.)

IV. Experiments and Discussions

To the best of our knowledge, there is still not a commonly used occlusion dataset for the contour-based object detection. We have collected images with occluded objects from the internet of four object classes, horses, mugs, swans and giraffes (approximately 20 images for each class)—just as the categories of the ETHZ shape datasets [4], which are popularly used but occlusion seldom included.

The proposed methods provide an efficient way of both occlusion recovery and occluded object detection & localization. For one thing, large portions of missing contour parts are recovered in a consistent deformation manner as that of the observed parts (Fig. 2). For another, by simultaneous reconstruction and detection, our method has obvious advantages compared with the previous contour-based detection approaches which have not addressed the occlusion problem.

We compare our method—voting by reconstruction proposals, with one of the state-of-the-art approaches as in [19]—voting by curve segments (We use the same partial matching strategy as in [19] to propose a set of potential correspondences between curve segments and the template.) It is shown in Fig. 2(b) that our method finds much more complete and accurately localized contours. The detection performance on our occlusion dataset is greatly improved by our method v.s. [19], as illustrated by the Precision/Recall (P/R) curve (average of different classes) in Fig. 3, and the Interpolated Average Precision (AP) computed from the P/R curves (as in PASCAL VOC challenge [3]) in Table I. Also our method achieves more accurate object localizations than [19], according to the average

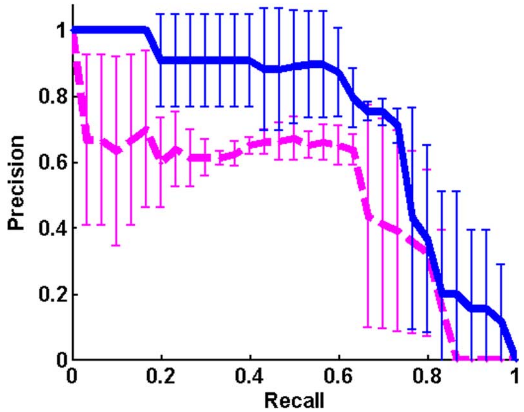

Fig. 3. Detection performance by Precision/Recall curve (the dotted pink for the result of [19] and the solid blue for ours), averaged by all the object classes of our occluded dataset, the error bars indicating the standard deviation at different recall rates.

TABLE I
THE INTERPOLATED AVERAGE PRECISIONS (AP) AND AVERAGE BOUNDING BOX ACCURACY ON OUR OCCLUSION DATASET

	Interpolated AP		Average BB accuracy	
	Riemensch-neider[19]	Ours	Riemensch-neider[19]	Ours
horses	0.6100	0.8131	0.5272	0.6359
swans	0.4240	0.6024	0.6658	0.7782
giraffes	0.5642	0.6560	0.5377	0.6927
mugs	0.5686	0.7395	0.7316	0.8052
mean	0.5417	0.7028	0.6056	0.7280

TABLE II
CONTOUR RECOVERY ERRORS ON SYNTHETIC OCCLUSION DATASET

applelogos	bottles	mugs	giraffes	swans	mean
0.1540	0.1294	0.2175	0.2856	0.3966	0.2366

area rate of the intersection vs. union of the ground-truth and detected object bounding boxes (BB) (Table I).

Additionally, we experiment on the synthetic occlusion datasets by randomly placing ellipse boards of random size and aspect ratio into the test images of ETHZ datasets. Table II shows the quantitative evaluation of contour recovery errors measured by the Hausdorff distances of the ground-truth outlines and the recovered results (all normalized with the longest side of BB scaled into $[-1, 1]$).

The object shape priors enable our method to better deal with complex shape structures than the traditional methods with smoothness priors [7]. Whereas our performance depends on the consistency of the test instance and object prior. Generally speaking, the more consistency between the two, the more accurate the recovered contour. In a worst case of totally unavailable object prior, we can only use general shape prior such as the smoothness to recover the contour. However the recovery results based on the general prior are much worse than those with specific object prior [7].

V. CONCLUSIONS

Focusing on missing data recovery and occluded object detection, we propose a new contour reconstruction algorithm, which is able to complete object contours with various shape deformation, and robust to large portions of occlusion. Moreover, the reconstruction is demonstrated to greatly facilitate the detection of occluded objects, which achieves high performance with well localized and recovered object contours.

REFERENCES

[1] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, p. 489C509, 2006.
[2] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, no. 2–3, pp. 114–141, 2003.
[3] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object class (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.
[4] V. Ferrari, T. Tuytelaars, and L. V. Gool, "Object detection by contour segment networks," in *Eur. Conf. Computer Vis.*, 2006 [Online]. Available: Dataset:www.vision.ee.ethz.ch/ferrari
[5] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, 1997.
[6] B. K. Gunturk and M. Gevrekci, "High-resolution image reconstruction from multiple differently exposed images," *IEEE Signal Process. Lett.*, vol. 13, no. 4, pp. 197–200, 2006.
[7] G. Guo, T. Jiang, Y. Wang, and W. Gao, "2-D shape completion with shape priors," Tech. Rep., Peking Univ., Beijing, China, 2011.
[8] S. Ilic and P. Fua, "Implicit meshes for modeling and reconstruction," in *IEEE Conf. Computer Vis. and Patt. Recognition*, 2003.
[9] G. Kanizsa and W. Gerbino, , J. Beck, Ed., "Amodal completion: Seeing or thinking?," in *Organization and Representation in Perception*. Hillsdale, NJ: Lawrence Erlbaum, 1982.
[10] B. Kimia, I. Frankel, and A. Popescu, "Euler spiral for shape completion," *Int. J. Comput. Vis.*, vol. 54, no. 1/2, pp. 157–180, 2003.
[11] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *Int. J. Comput. Vis.*, 2008.
[12] C. Lu, L. J. Latecki, N. Adluru, X. Yang, and H. Ling, "Shape guided contour grouping with particle filters," in *Int. Conf. Computer Vis.*, 2009.
[13] T. Ma and L. J. Latecki, "From partial shape matching through local deformation to robust global shape similarity for object detection," in *IEEE Conf. Computer Vis. and Patt. Recognition*, 2011.
[14] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, 2004.
[15] D. Mumford, "Elastica and computer vision," *Algebr. Geom. Applicat.*, pp. 491–50, 1994.
[16] B. Ommer and J. Malik, "Multi-scale object detection by clustering lines," in *Int. Conf. Computer Vis.*, 2009.
[17] X. Ren, C. Fowlkes, and J. Malik, "Learning probabilistic models for contour completion in natural images," *Int. J. Comput. Vis.*, vol. 77, no. 10, pp. 47–63, 2008.
[18] R. Rensink and J. Enns, "Early completion of occluded objects," *Vis. Res.*, vol. 38, pp. 2489–2505, 1998.
[19] H. Riemenschneider, M. Donoser, and H. Bischof, "Using partial edge contour matches for efficient object category localization," in *Eur. Conf. Computer Vis.*, 2010.
[20] W. Rutkowski, "Shape completion," *Comput. Graph. Image Process.*, vol. 9, pp. 89–101, 1979.
[21] A. Sharf, M. Alexa, and D. Cohen-Or, "Context-based surface completion," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 878–887, 2004.
[22] P. Srinivasan, Q. Zhu, and J. Shi, "Many-to-one contour matching for describing and discriminating object shape," in *IEEE Conf. Computer Vis. and Patt. Recognition*, 2010.
[23] K.-H. Yap, Y. He, Y. Tian, and L.-P. Chau, "A nonlinear L1-norm approach for joint image registration and super-resolution," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 981–984, 2009.
[24] P. Yarlagadda, A. Monroy, and B. Ommer, "Voting by grouping dependent parts," in *Eur. Conf. Computer Vis.*, 2010.