Drawing an Automatic Sketch of Deformable Objects Using Only a Few Images

Smit Marvaniya, Sreyasee Bhattacharjee,
Venkatesh Manickavasagam, and Anurag Mittal

Indian Institute of Technology Madras, India
{smit,amittal}@cse.iitm.ac.in,
{sreya.iit,venky9111}@gmail.com

Abstract. We propose a method to automatically extract a sketch of a common object structure present in a small set of real world weakly-labeled images. Applying a part-based deformable contour matching technique gives the location of repeatable contours. An initial deformable search strategy selects a set of salient, repeatable contours robust to a large range of non-rigid deformations. A contour completion technique based on a locally greedy bi-directional search strategy is adopted to merge the repeatable contour fragments for obtaining a complete shape. The output of our algorithm is used as an input to a sketch-based object-recognizer with results that are either better, or on par with those obtained with the ground truth sketches provided with the dataset.

Keywords: Salient Contours, Part-based Deformable Contour Matching, Contour Completion.

1 Introduction

Building a contour model using edge information is an important problem in computer vision that has received considerable attention from many researchers. In this paper, we propose a technique to extract a compact sketch of an object from a few training images. The object shape is allowed to vary under conditions such as non-rigid deformations, affine transformations and a cluttered background. However, we believe that the common object shares a similar geometrical structure across the training images. We intend to capture this common shape in terms of a rough sketch through a completely automated process. While most established methods [1, 2] obtain the training images from carefully chosen dataset elements, or are captured against a uniform background, we allow the system to automatically extract a set of training images from any resource including web-engines.

Popular methods rely on either part-based [3], region based [4] or a combination of both shape and region based cues [5]. Contour-based methods are attractive since it is well-known that humans are able to identify an object simply from its contour or shape. Furthermore, while the region-based approaches [6] perform well only on good quality images, learning-based approaches [1–3, 7] are becoming increasingly popular due to their ability to handle a wide range of deformations for object recognition. However, in most cases, the basic prerequisite of these approaches is a huge set of positive training

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part I, LNCS 7583, pp. 63–72, 2012.

images with annotated bounding boxes, except an appearance-based model proposed by Bagon et.al. [8]. This requirement may be expensive and hard to obtain. The weakly supervised learning based techniques [5, 9] typically perform well in terms of learning an object model from the data by indicating the presence of an instance from a category without specifying its exact location in the training images.

Our contour based deformable matching technique helps in extracting a common shape from a handful of weakly labeled images and can therefore be treated as a cost-effective alternative. Our multi-stage automatic sketching process first attempts to identify a non-trivial commonality from training images using FDCM [10], with a localization up to rough bounding boxes. The next step involves extracting a group of repeatable, salient contours which combined together represent almost the whole structure of the common object shape. The resultant rough sketch can be used as an input to any sketch-based object-recognizer [11–13] for object recognition. The entire process is described in Figure 1.

The main contributions of this paper are: 1) Proposing a completely automatic process for drawing a sketch of the common object from a set of weakly labeled data using only contour based cues. 2) Proposing an efficient contour matching technique, which can handle a cluttered environment, scale, orientation and view point variations to a certain extent. 3) Shortlisting some repeatable contours in a deformation invariant fashion based on a novel deformable matching technique proposed by Ravishankar et.al. [11].

The rest of the paper is organized as follows: Section 2 describes the process of extracting a set of salient contours and the process of localization up to rough bounding boxes is explained in Section 3. Section 4 illustrates the process of obtaining a set of repeatable contours and Section 5 elaborates on the mechanism for completing the repeatable salient contours using a bi-directional search strategy to obtain an initial model. Finally, the experimental results are shown in Section 6.

2 Extracting a Set of Salient Contours

Given a set of training images $\{I_1, ..., I_k\}$, we first resize them to a predefined standard width in order to reduce the effect of large scale variance. We use the Berkeley edge detector [14] to get the edge map of the images in the training set and then use hysteresis thresholding followed by an efficient contour grouping proposed by Zhu et. al. [15] for extracting a set of potential contour groups from the output edge map. The saliency of a contour is defined using the following three components:

1. **Length of a contour**: A contour should be sufficiently long to represent some meaningful feature. Small spurious contours are eliminated as noise. The threshold on the contour length is calculated based on the size of the image.
2. **Salient points on a contour**: The number of high curvature points on a contour is an important cue to represent the descriptive power of a contour. However, too many high curvature points close to each other indicate that the contour probably originated from a cluttered background or some other kind of noise.
3. **Complexity Measure**: Complexity of a salient contour gives information about its smoothness. It is defined as the sum of the supplementary inner angles between line segments constituting the salient contour. A less complex salient contour doesn't

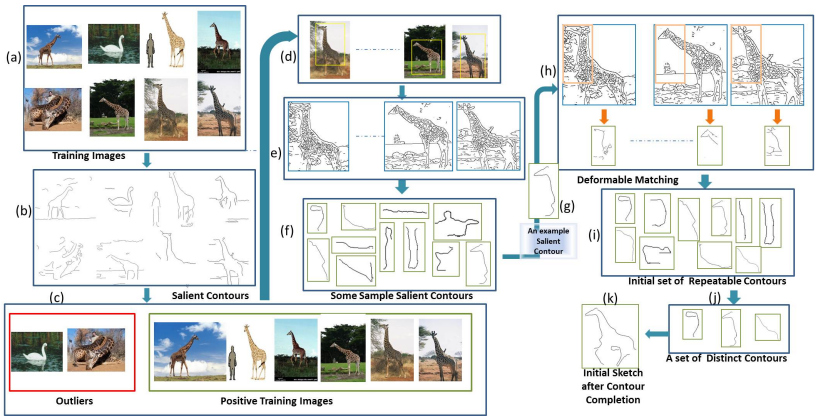

Fig. 1. (a) Training Images, (b) Salient Contours of the Training Images, (c) Classifying the training images, (d) Some example images with annotated bounding boxes obtained using method described in Section 3.1, (e) Binarized edge response of Berkeley edge detector on those images, (f) Examples of some salient contours extracted, (h) Explains the deformable matching technique adopted for getting the deformation component of the repeatability score of a salient contour using the 'neck' of a giraffe in (g), (i) Some examples of repeatable contours, a small set of distinct contours shown in (j) is obtained and used to draw an initial sketch shown in (k).

possess unique shape information while a highly complex one ends up making the system rigid.

Initially, a bunch of salient contours are identified using the first two saliency criteria. In order to reduce the effect of noise and partial matching, the salient contours are broken at every branch point into smaller fragments. We then calculate the Elastica measure [16, 17] at each branch point of the contour. Given two consecutive tangent directions, ϕ_1 and ϕ_2, the quantity $El(c_1, c_2) = 4(\varphi_1^2 \times \varphi_2^2 - \varphi_1 \times \varphi_2)$ provides a good approximation of the Elastica energy for curvature consistency at the junction of contours c_1 and c_2(see figure 3(b)). The pair of fragments having the minimum curvature inconsistency (min. Elastica cost) are then combined to resolve the branch point. The process is repeated till all the branch points in the shape have been resolved.

3 Identifying Repeated Contours Following a Deformable Matching Strategy across Images

Given a set of images represented by a collection of initial salient contours, the system attempts to extract a recurring shape structure (if any) across many of these image instances. The proposed system needs to be flexible enough to deal with the problem of occlusion, intra-class deformation and cluttered background. Our proposed matching strategy is based on the observation that, in some images at least some of the object parts are clearly visible. Clearly visible parts help the system to get the repeated salient contour set (wherever available). The salient contours of a particular image are represented at various levels of granularity - *Child Contours* and *Parent Contours*.

1. **Child contour**: The salient contours are represented in terms of a set of child contours. Child contours are constructed by first decomposing all the salient contours according to the first two saliency criteria mentioned in Section 2. Of those contours that do not satisfy both criteria, those which satisfy the length criterion alone are also considered as child contours.
2. **Parent contour**: Parent contours are formed by merging consecutive child contours, where the extent of merging is governed by the complexity measure. This process ends when the complexity measure reaches a predefined threshold Cp_{th}.

The similarity values between all pairs of images computed using the proposed matching strategy, explained in Section 3.1, are finally stored in our repository.

3.1 Matching Strategy

In order to retrieve a mutual commonality among images, the basic matching algorithm using edges as features should be tolerant to small deformations of shapes and fragmentations of edges. In this work, we use Fast Directional Chamfer Matching (FDCM) [18] which works much better in such a scenario. The local maxima for a given matched contour in an image are determined by non-maximal suppression and represented on a map overlaid on the image by the location of their mid and the two end-points. The FDCM score M_{dc}, evaluating the goodness of a match found in the image, is also retained for each subsegment. From such information, a rough scale and rotation angle of the match are also precomputed for later use.

Given a *Parent contour* P_k originating from an underlying image I, FDCM is used to extract a few smaller windows (if any) as a set of potential matched locations. We add Parent contour P_k into the Matched Contour Set MC by adding its constituent child contours.We then identify its nearest neighboring child contour c_j from MC_{P_k} in I. The relative location of c_j with respect to MC_{P_k} can extract a roughly similar region in $I^{'}$ (Figure 2), which is searched for a potential match for c_j using FDCM. A match is declared as reasonably good, if c_j satisfies the goodness measure $G(c_j)$. We again add the constituting child contours c_j of P_k into MC_{P_k} in I. The above steps are repeated until no nearest neighbor is found and the region in $I^{'}$ with respect to c_j is marked using dynamic programming to prevent it from being matched multiple times to the contour in the model image. The Matched Contour Set MC_{P_k} grows with each match.

$$G(c_{i,j}^{N_c}) = w_a \times M_{dc}(c_{i,j}^{N_c}) + w_b \times A(c_{i,j}^{N_c}) + w_c \times T(c_{i,j}^{N_c}) \tag{1}$$

$$A(c_{i,j}^{N_c}) = [1 - e^{-\frac{\triangle\theta_{i,j}}{180} * \pi}] \tag{2}$$

$$T(c_{i,j}^{N_c}) = \frac{\sqrt{((\triangle x_l)^2 + (\triangle y_l)^2)}}{D_l} \tag{3}$$

where N_c is the total number of child contours in I_i, $\triangle\theta_{i,j} = \theta_{i,j}^{'} - \theta_{i,j}$ accounts for local angular inconsistency where $\theta_{i,j}$, $\theta_{i,j}^{'}$ represents the relative spatial information in between two neighboring Child contours in I and $I^{'}$ respectively. $T(c_{i,j}^{N_c})$ represents a quantitative measure for translational inconsistency to evaluate the amount of deviation

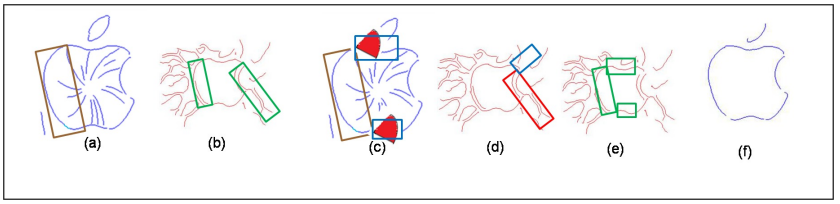

Fig. 2. (a) Model image highlighted with Parent Contour. (b) Target image shown with multiple matches with respect to parent contour. (c) Identifying the nearest child contour using bidirectional search technique. In (d) and (e), with respect to the already estimated locations for Parent contour, similar regions are explored in the target image to obtain a suitable match for the Child contour. The System failed to obtain a suitable match for Child contour in (d). The final Repeated Contours are shown in (f).

of the present position of $c_{i,j}^{N_c}$ from its estimated position with respect to its parent segment $c_{i,j+1}^{N_c}$ where D_l is the l_2-diagonal length of the image I.

$$SC(\mathcal{S}_{i,j}^{c_0}, P_k) = \frac{1}{n}\sum_{n} G(c_{i,j}^{n}) + \sum_{m} S_{const} \times w_m + M_{dc}(P_k) \times w_p \tag{4}$$

where n and m are the number of matched and skipped child contours respectively, w_m is the defined as $\frac{1}{N_c}$, w_p is defined as $\frac{1}{N_{p_c}}$ where N_{p_c} is the number of Child contours constituting the parent contour P_k and S_{const} is the penalty for those child contours that do not find a match in the target image. We take w_a=1, w_b=0.5, w_c is set based on the image size. We repeat the above process for each parent contour and the Matched Contour Set MC of those with the minimum dissimilarity score is considered a Repeated Contour Set. Finally, the pairwise image score $PS(i,j)$ is defined as follows:

$$PS(i,j) = min\left\{SC(\mathcal{S}_{i,j}^{c_0}, P_k)\right\}_k \tag{5}$$

3.2 Bounding Box for the Positive Images

For getting the bounding box for image I_i, we calculate the repeatable contours $RC(I_i)$ across the remaining training images in the training set which satisfy the repeatability threshold R_{th}.

$$RC(I_i) = \left\{Rep(c_{i,j}^{N_c}) > R_{th}\right\}_k \tag{6}$$

where $Rep(c_{I_i,j}^{n})$ is the repeatability of j^{th} child contour in i^{th} image and $1 \leq i \leq k$. The bounding box for I_i is the tight bounding box for $RC(I_i)$.

4 Extracting a Set of Candidate Foreground Contours from Images

Due to the influence of noise and other external clutter, the prediction of the bounding box described above is rough. In order to reduce the effect of scale, images are

cropped along their annotated bounding boxes and resized in a standardized frame for further consideration. Given an image, only salient contours lying within its cropped sub window are retained for describing it.

4.1 Repeatability of a Contour

The repeatability measure of a salient contour is defined using its Deformable Matching Score, a Shape Context [19] based similarity score and its length. In the following subsections we will discuss each of these components in detail.

Deformable Matching Score: In order to be repeatable, a salient contour c_j, originating from I_j should have a potential match at a similar location in another training image I_i. Unlike most existing methods that rely on rigid matching, we attempt to handle shapes that may undergo an amount of non-rigid deformation. In order to do so, we use the deformable *Fine Matching* strategy proposed by Ravishankar et.al. [11] to deal with a large set of non-rigid deformations and assign a deformable matching score to every salient contour from an image. The algorithm had shown to achieve among the best results on the standard ETHZ dataset, if only a sketch was available.

The tight bounding box around c_j extracts a patch P_j from I_j. Patch P_i from a similar location in the edge map of one of the remaining training images is found. While treating P_j as a model sketch of the contour, the deformable fine-matching strategy was adopted to find its good match in P_i. Each model contour is broken at high curvature points to be represented in terms of k-segments and a dynamic programming based matching strategy finds a suitable match in the target image. The resulting comprehensive matching cost(Q) takes into account inter-segment scale and orientation variations as well as edge strength and intra-segment bending deformations of the matched contour segment in P_i. Finally, the repeatability score R_j (e^{-Q}) is computed as a function of Q, ensuring a goodness measure for c_j. By its very definition, the value of Q always lies within a tractable range, ensuring an effective repeatability score for c_j.

Accumulated Weighted Repeatability Score (AWR Score): Any true object contour should have reasonably good matches in many training images which in turn would increase its repeatability score (defined above in Section 4.1). The accumulated weighted repeatability (AWR) score of a contour c_j is thus computed as follows:

$$R(c_j) = \sum_i ((\underbrace{1 - we^{-L(C_i^{'})}}_{LC(i)}) \times SC_j(i) \times R_j(i)) \tag{7}$$

where $LC(i)$ works as a weighing term providing a positive bias to longer matched contours. $SC_j(i)$ computes a Shape Context [19] based similarity score obtained from matching c_j in i^{th} image. While R_j computes a goodness measure from a deformable point of view, the shape context based matching score aims to extract the amount of overall similarity between c_j and its match in each training image. These measures prove to be an effective combination for achieving a robust AWR score. A smaller group of salient contours having nonzero repeatability scores are shortlisted as $\mathcal{C}_{rep}$.

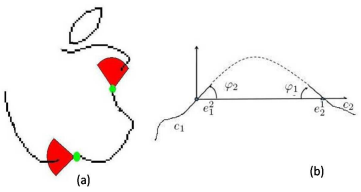

Fig. 3. (a) An example of the neighborhood search for contour completion using oriented sectors, (b) Contour continuity at the adjoining end points e_1^2 and e_2^1 of two contours c_1 and c_2 is measured using Elastica completion cost

However, a contour completion strategy is required to verify and merge the potential candidate contours to extract a fully (or mostly) complete sketch for an object category.

5 Contour Completion Using Neighborhood Search

In order to evaluate the relative structural configurations of a set of salient contours $\mathcal{C}_I(\in \mathcal{C}_{rep})$ originating from I, we propose a deformable contour completion that results in a sub-sequence of contours from $\mathcal{C}_I$. Given a repeatable primary contour c originating from an image I, a bi-directional search process (redrawn from Ravishankar et.al [11] as shown in Figure 3(a)) explores neighborhoods at both end points of a contour in parallel. In a one-Vs-many matching strategy, if there is a similar spatial contour layout observed in many training images, we declare that contour extension to be valid. The entire contour chain is built in steps. The comprehensive compatibility score corresponding to the goodness of an extended contour is dependent on its length, curvature continuity at the connections computed using Elastica Completion cost [16, 17] and an average repeatability score computed as follows:

$$CC(c_1, c_2) = El(c_1, c_2) \times \frac{L(c_1)}{(L(c_1) + L(c_2))} \times R_{mean}(c_1, c_2) \tag{8}$$

where, $CC(c_1, c_2)$ evaluates the goodness score at the connection point of c_1 and c_2. In the case that there are multiple candidates for extension, we follow a locally greedy approach to choose the locally best neighbor for merging (Figure 3(b)). The search process at both its end points is continued until we reach a stage where there is no reasonably compatible neighbor to extend it further. A similar iterative process is repeated until $\mathcal{C}_I$ is empty. However at this stage it would be unrealistic to assume that only one complete contour would be able to cover the entire object shape. In contrast, we may land up achieving a set (S) of contour chains. Initiated with the highest repeatable contour chain, contours are iteratively merged with all the other elements of S if their relative spatial configuration is mostly similar in many images.

6 Experimental Results

Given a user-specified object category, we performed experiments using different number (n) of training images. We conducted tests on automatically downloaded images

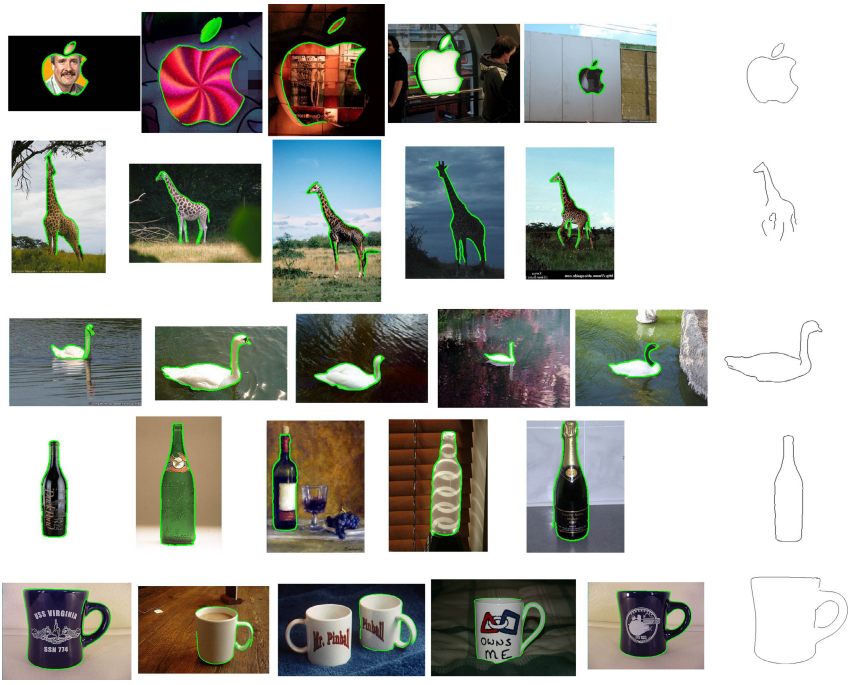

Fig. 4. Some results using a small training set of five images, taken from ETHZ dataset

Fig. 5. Some results using a small training set of four images, taken from Caltech101 dataset

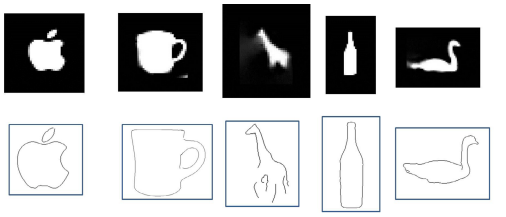

Fig. 6. Our results are shown in the second row

from a search engine and also on images from the ETHZ [20]/Caltech101 dataset. The comparative study is reported on the ETHZ dataset which has different classes of objects: Apple logo (40), Bottle (48), Giraffe (87), Mug (48) and Swan (32). The objects in the images are at various scales, orientations, illumination changes and a substantial amount of intra-class variations which make it a difficult dataset to work on. For each value of n in the range 4-10, results of some experiments on five training images of each category chosen from the ETHZ dataset and on four training images from Caltech101 dataset are shown in Figures 4 and 5 respectively. As seen in Figure 6, the sketch obtained by Bagon et.al [8] loses important information in the deformable parts of the object, such as giraffe's legs and swan's beak, while we are able to obtain such details due to a *deformable* approach.

The best extracted sketch is treated as the output of our system. The parameters used to obtain such a sketch are set independent of the image category. A sketch obtained from a particular iteration was used for object recognition using the algorithm of Ravishankar et.al [11] on the ETHZ dataset and the corresponding detection rate was used to evaluate the quality of the sketch. Table 1 shows the detection rates at 0.4 and 0.3 FPPI averaged over 100 iterations. We have also referred to other state of the art results for completeness. However, the important observation is that the same algorithm proposed by Ravishankar et.al [11] sometimes performs better than the original sketches due to some additional information extracted by our system. In other cases, the performance achieved by Ravishankar et.al using some hand-drawn sketches remains the same. Our system's improved performance on the giraffe category is due to the more complete sketch obtained by it. It was partly successful in obtaining 'legs' that enabled it to achieve a better result, rather than by using the ground truth model provided along with the dataset. We have achieved a detection rate of 93.4% at a FPPI as low as 0.1 on giraffe images. The results on swan category were again marginally better due to the better sketches obtained by our system. We allowed it to include some outliers at random so that its robustness to outliers would systematically evolve(ref. Table 1).

Table 1. Comparison of detection rates of objects at 0.4 FPPI / 0.3 FPPI

Ref	Applelogo	Bottle	Giraffe	Mug	Swan
Ravishankar et al. [11]	97.7/95.5	92.7/90.9	93.4/91.2	95.3/93.7	96.9/93.9
Lu et al. [13]	92.5/92	95.8/95.8	92.0/86.2	85.4/83.3	93.8/93.8
Zhu et al. [21]	80.0/80.0	92.9/92.9	68.1/68.1	74.2/64.5	82.4/82.4
Riemenschneider et al. [22]	93.3/93.3	97.0/97.0	81.9/79.2	86.3/84.6	92.6/92.6
Our System	97.7/97.7	92.7/90.9	93.4/93.4	95.8/95.8	96.87/96.87

7 Conclusion

We have demonstrated a method for drawing an automatic sketch from a very small set of training images that may have outliers. Such a sketch was found to be effective for object recognition. Apart from a visually appealing output, our work can form a part of a complete object recognition system that can automatically find objects in images obtained using a text query from a search engine.

References

1. Ferrari, V., Jurie, F., Schmid, C.: From images to shape models for object detection. International Journal of Computer Vision 87(3), 284–303 (2010)
2. Srinivasan, P., Zhu, Q., Shi, J.: Many-to-one contour matching for describing and discriminating object shape. In: CVPR (2010)
3. Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1627–1645 (2010)
4. Todorovic, S., Ahuja, N.: Extracting subimages of an unknown category from a set of images. In: CVPR, pp. 927–934 (2006)
5. Lee, Y.J., Grauman, K.: Shape Discovery from Unlabeled Image Collections. In: CVPR (2009)
6. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. International Journal of Computer Vision 61(1), 55–79 (2005)
7. Wu, B., Nevatia, R.: Simultaneous object detection and segmentation by boosting local shape feature based classifier. In: ICCV, pp. 1–8 (2007)
8. Bagon, S., Brostovski, O., Galun, M., Irani, M.: Detecting and sketching the common. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 33–40 (2010)
9. Prest, A., Schmid, C., Ferrari, V.: Weakly supervised learning of interactions between humans and objects. IEEE Trans. Pattern Anal. Mach. Intell. 34(3), 601–614 (2012)
10. Liu, M.Y., Tuzel, O., Veeraraghavan, A., Chellappa, R.: Fast directional chamfer matching. In: CVPR, pp. 1696–1703 (2010)
11. Ravishankar, S., Jain, A., Mittal, A.: Multi-stage Contour Based Detection of Deformable Objects. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 483–496. Springer, Heidelberg (2008)
12. Bai, X., Latecki, L.J., Li, Q., Liu, W., Tu, Z.: Shape band: A deformable object detection approach. In: CVPR (2009)
13. Lu, C., Latecki, L.J., Adluru, N., Yang, X., Ling, H.: Shape guided contour grouping with particle filters. In: ICCV, pp. 1–8 (2009)
14. Martin, D., Fowlkes, C., Malik, J.: Learning to detect natural boundaries using local brightness, color and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 530–549 (2004)
15. Zhu, Q., Song, G., Shi, J.: Untangling cycles for contour grouping. In: CVPR (2007)
16. Kokkinos, I., Yuille, A.: Inference and learning with hierarchical shape models. International Journal of Computer Vision, 1–25 (2010)
17. Mumford, D.: Elastica and computer vision. In: Bajaj, C.L. (ed.) Algebraic Geometry and its Applications. Springer, New York (1994)
18. Liu, M.Y., Tuzel, O., Veeraraghavan, A., Chellappa, R.: Fast directional chamfer matching. In: CVPR (2010)
19. Belongie, S., Puzhicha, J., Malik, J.: Shape matching and object recognition using shape contexts. IEEE Transactions. on Pattern Analysis and Machine Intelligence 24, 509–522 (2002)
20. Ferrari, V., Tuytelaars, T., Van Gool, L.: Object Detection by Contour Segment Networks. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 14–28. Springer, Heidelberg (2006)
21. Zhu, Q.-H., Wang, L.-M., Wu, Y., Shi, J.: Contour Context Selection for Object Detection: A Set-to-Set Contour Matching Approach. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 774–787. Springer, Heidelberg (2008)
22. Riemenschneider, H., Donoser, M., Bischof, H.: Using Partial Edge Contour Matches for Efficient Object Category Localization. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 29–42. Springer, Heidelberg (2010)