

Object Localization by Propagating Connectivity via Superfeatures.

Ishani Chakraborty
Rutgers University, NJ, USA

Ahmed Elgammal
Rutgers University, NJ, USA

Abstract

In this paper, we propose a part-based approach to localize objects in cluttered images. We represent object parts as boundary segments and image patches. A semi-local grouping of parts named superfeatures encodes appearance and connectivity within a neighborhood. To match parts, we integrate inter-feature similarities and intra-feature connectivity via a relaxation labeling framework. Additionally, we use a global elliptical shape prior to match the shape of the solution space to that of the object. To this end, we demonstrate the efficacy of the method for detecting various objects in cluttered images by comparing them to simple object models.

1. Introduction

Many object localization/recognition algorithms view an object as a collection of parts [9, 3, 2]. Part relations may be viewed from being completely independent [9] to having strict dependencies [2], at two extremes. In general, we observe that local geometry of object parts is rigid at short spatial distances, while inter-part relations may vary considerably over longer spatial distances due to artifacts such as deformation, articulation etc. To this end, we propose a part-based localization that attempts to capture this variability in the inter-part relations. The approach is summarized as follows.

We use two kinds of image features as parts namely, *boundary line segments* and *image patches*. The relationship between parts is determined by their relative geometric locations. To capture part appearances and relations, we create *superfeatures* - features that are grouped together based on their local geometry. The advantages of creating superfeatures are two-fold.(a) Comparing superfeatures across images is more *discriminative* than compar-

ing individual features.(b) Superfeature groupings overlay an *implicit connectivity* between features within an image. Object localization is achieved by finding an interconnected set of image features. We exploit inter-feature similarities and intra-feature connectivities to match features in an iterative *relaxation labeling framework* [10]. Additionally, we use a *global elliptical shape prior* to restrict shape and number of localizations in an image.

There is a broad agreement in the research community that object is best represented as a sum of parts and relations [2]. Recently, some research has started exploring semi-local configurations of local features [5]. We base our semi-local description of boundary segments superfeatures on Pair of Adjacent Segments framework [3] and the keypoint superfeature on Semi-local Descriptor in [5]. Most of the parts and relations methods use iterative approaches to localize the object [12]. Relaxation labeling based feature matching has been used in [7] before. This paper builds on the framework introduced in [1] in which a multi-level contour segment based representation and matching of images was proposed. However, such a shape-only description of objects is insufficient for complex and texture-rich object classes.

2. Superfeature Description

We represent object parts by two types of features namely, line segments that approximate boundary shape and image patches (keypoints) that describe the internal appearance/texture. This combined representation of shape and appearance ensures an efficient description across object categories (Figure 2). We introduce *superfeatures* as semi-local grouping of local features. The idea is to connect spatially proximate and geometrically related features, and endow them jointly with a feature descriptor. In our framework, we generate superfeatures based on keypoints and boundary line-

segments.

- **Boundary segment superfeature:** We use Pairs of Adjacent Segment descriptor [3] for line segment superfeatures. Each line segment is paired with its closest neighbor at its end point. The descriptor vector consists of (a) The normalized relative distances between mid points of line segments, (b) Line segment orientations and (c) Normalized length of the individual line segments. A pair of superfeatures is compared by the Euclidean distance between the descriptor vectors.
- **Keypoint superfeature:** We use Scale Invariant Feature Transform (SIFT) [6] method to extract keypoints and their representations. We pair each keypoint with the N closest keypoints that occur in the first quadrant with the reference keypoint as center, similar to [5] ($N = 10$ in our experiments). Each SIFT keypoint has its associated 128 descriptor vector. The distance between two superfeatures is the average of the Euclidean distances between each ordered pair of keypoints.

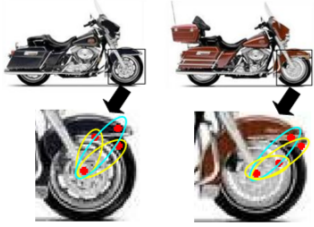


Figure 1. Keypoint superfeatures: Several pairings are possible (colorcoded, yellow), but the similarity measure for the keypoints is derived from the distance of the best matching pair (colorcoded cyan).

2.0.1 Inter-feature Distances

Each feature can belong to multiple superfeatures (Figure 1). The distance between two features is then attributed to the distance between the best matching superfeature. In effect, the inter-feature distance not only encodes the similarity between the individual features but also the neighborhood appearance and geometry. Let a be a feature in image I , α be a feature in object model O and S denote the superfeature set. Then the distance is formulated as follows.

$$d_{a\alpha} = \min_{i,j} (D(S_i^I, S_j^O) | a \in S_i^I, \alpha \in S_j^O) \quad (1)$$

2.0.2 Intra-feature connectivities

In our framework, we interconnect features within an image to identify neighborhood context. Connections are established in two ways. 1. Trivially, a superfeature connects a pair of features. 2. Features can also be indirectly connected by propagation through intermediate features. For example, if feature a forms superfeatures (a, b) and (a, c) , then b and c get connected by propagation of connectivity through a . We define two binary adjacency matrices D and M that encode the intra-feature connectivities in the image and object model respectively.

3. Matching via Relaxation Labeling

We frame object localization as a feature matching problem in which the inter-feature distances and intra-feature connectivities are used in an iterative relaxation labeling framework [10]. Each iteration consists of two steps - 1. Matching step. 2. Update step. We define a *probability matrix* Q that denotes the probabilities of assignment of each image feature to every object feature. Through the iterative process, we update these probabilities until a stable solution is reached.

The matching step computes a one to one correspondence between image and object features. This is accomplished by using the Hungarian algorithm [4] on the probability matrix Q . The iteration starts by setting Q equal to the inter-feature similarities ($d_{a\alpha}$) that induces an initial matching.

Next, in the update step, Q is updated based on the intra-feature connectivities (D and M). Each probability value of an image feature is enhanced or reduced through a *support function* ($S_{a\alpha}$) based on its connections to the matched features in the matching step. The update at iteration $i + 1$ is

$$Q_{a\alpha}^{i+1} = \frac{Q_{a\alpha}^i S_{a\alpha}^i}{\sum_b \sum_{\beta} Q_{b\beta}^i S_{b\beta}^i} \quad (2)$$

Finally, we describe the support function. Lets assume that an image feature b is connected to a and similarly, an object feature β to α . We define a compatibility indicator $r_{b \rightarrow \beta}^{a \rightarrow \alpha}$ between 2 neighboring features a and b . It measures the hypothesis that “ a and b are both assigned to object features that are also connected to one another”.

$$r_{b \rightarrow \beta}^{a \rightarrow \alpha} = 1, \text{ if } D_{ab} = 1, M_{\alpha\beta} = 1, a \rightarrow \alpha, b \rightarrow \beta \\ = 0, \text{ otherwise}$$

We formulate the support function probabilistically using a noisy-OR model [8]. It gates in the probability scores of matched features via the com-

patibility indicators.

$$S_{\alpha\alpha}^i = 1 - \left\{ \prod_b \prod_{\beta} (1 - Q(b, \beta))^{r_{b \rightarrow \beta}^{\alpha}} \right\} \quad (3)$$



Figure 2. Detections using boundary segments (left), keypoints (center) and combining boundary segments and keypoints (right).

4. Adding Shape Prior

A shape prior is used to restrict the shape of the solution space and speed up the convergence of the algorithm. In effect, object localization is confined to a single region in the input image (Figure 4). We use an elliptical shape prior, similar to [11]. The 2D covariance parameters of the ellipse is learnt from the spatial distribution of the feature coordinates in the model. The shape structure is enforced during detection to ascertain that the matched features respect the general elliptical spread of the object. In order to accommodate features appearing at the fringe of detected space, the ellipse is set to a size 3 standard deviations from the mean.

During testing phase, the *median* of the detected features is estimated as the centroid. The elliptical prior is fitted on this centroid. Any point within the ellipse is given a weight of 1 and any point outside has a weight 0. We model the shape probability as a weighted sum of uniform prior and the elliptical prior. The probability matrix Q is scaled by the shape probability. The weighting coefficient λ is incremented as the confidence in the localization increases in each iteration. Specifically, the weights at each feature are computed as follows.

$$W_{prior}(x, y) = \lambda * ShapePrior(x, y) + (1 - \lambda) * \mathbf{1} \quad (4)$$

5. Experiments and Results

We evaluate our localization algorithm on three object categories, namely Giraffe class [from ETHZ dataset] (87 images), Motorbike class [TUD/ETHZ] (115 images) and the Helicopter class [Caltech101] (41 images). The object categories were selected to evaluate on diverse appearances and part-relations. The Giraffe class has a repeated, textured pattern

but a unique, well constrained outline contour. The motorbike class has a characteristic appearance but variable contours. The helicopter class has a moderately consistent appearance and shape.

We compare each of the three classes with a *single model* of each class. Contour segments and keypoints are extracted using publicly available codes.¹² The stopping criterion for the relaxation labeling process is when Q remains unchanged in two consecutive iterations. The number of iterations was empirically found to not exceed 12 iterations, hence the λ of the global shape prior is initialized to 0.1 and is scaled by 1.2 after each iteration. The contour segments and points detected in the final iteration are framed with a bounding box. The output is a unique bounding box due to the shape prior. The localization is correct if the overlap between true and detected bounding boxes is greater than 50% of the union.

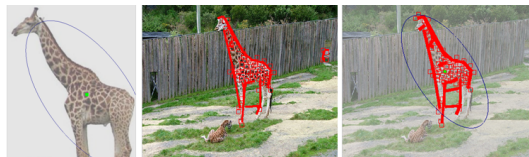


Figure 4. Effect of elliptical prior(left) to remove extraneous detections.

Our results show that all the three classes are detected with high confidence (Figure 3). For example, the localization of Giraffe category outperforms the baseline result of approx. 64% (at 0.3 FPPI in [3]). Other categories are incomparable with other methods because of the difference in the type and number of object models used. We notice that *combining appearance and shape cues significantly improves* the localization in all the categories. The influence of the additional global shape prior is more in the giraffe and motorbike categories than helicopter, because of higher amounts of cluttered background in these images. In general, the results demonstrate that our approach is highly successful in performing generic object localization even with a very simple model.

6. Conclusions

In this paper, we proposed a part-based approach to localize objects in cluttered images. We used superfeatures to encode discriminative

¹<http://www.vision.ee.ethz.ch/calvin/>

²<http://www.vlfeat.org/vedaldi/code/sift.html>

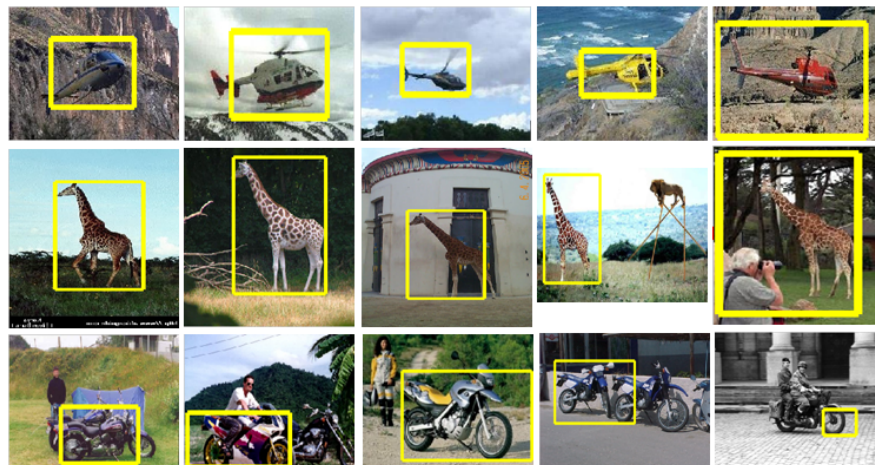


Figure 3. Detection bounding boxes. The last column shows incorrect detections.

inter-feature distances and *intra-feature connectivity*. Additionally, a global shape prior is used to restrict the shape of the detected space. The success of this method is demonstrated by the overall localization rate of 84% across three different categories.

Acknowledgements

This research has been partially funded by NSF award number 0803732.

	Giraffe	Motorbike	Helicopter
<i>Segments(S)</i>	75.8%	64.3%	85.3%
<i>Keypoints(K)</i>	82.7%	53.9%	78.0%
<i>False positives w/o prior</i>	0.27%	0.33%	0.07%
<i>(S+K+prior)</i>	82.7%	81.7%	87.8%

Table 1. Localization rates.

References

- [1] I. Chakraborty and A. Elgammal. Contour segment matching by integrating intra and inter shape cues of objects. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2009.
- [2] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages II: 264–271, 2003.
- [3] V. Ferrari, T. Tuytelaars, and L. Van Gool. Object detection by contour segment networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages III: 14–28, 2006.
- [4] H. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistic Quarterly*, 2:83–97, 1955.
- [5] Y. Lee and K. Grauman. Foreground focus: Finding meaningful features in unlabeled images. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2008.
- [6] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, November 2004.
- [7] B. Luo and E. Hancock. Structural graph matching using the em algorithm and singular value decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 23(10):1120–1136, October 2001.
- [8] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.
- [9] F. Perronnin, C. Dance, G. Csurka, and M. Bressan. Adapted vocabularies for generic visual categorization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages IV: 464–475, 2006.
- [10] A. Rosenfeld, R. Hummel, and S. Zucker. Scene labeling by relaxation operations. *IEEE Trans. Systems, Man and Cybernetics*, 6(6):420–433, June 1976.
- [11] G. Slabaugh and G. Unal. Graph cuts segmentation using an elliptical shape prior. In *Proceedings of the International Conference on Image Processing (ICIP)*, pages II: 1222–1225, 2005.
- [12] Q. Zhu, L. Wang, Y. Wu, and J. Shi. Contour context selection for object detection: A set-to-set contour matching approach. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages II: 774–787, 2008.