Contour-HOG: A Stub Feature based Level Set Method for Learning Object Contour

Zhi Yang
zhiyang@buffalo.edu
Yu Kong
yukong7@buffalo.edu
Yun Fu
raymondyunfu@gmail.com

Department of ECE and College of CIS
Northeastern University
Boston, MA, USA
Department of CSE
State University of New York
Buffalo, NY, USA

Abstract

In this paper, we address the problem of learning objects by contours. Toward this goal, we propose a novel curve evolution scheme which provides the classifier with more accurate contour representations. We detect edgelet feature to help localize objects in images so that the proposed evolution method can achieve more reliable contours. To capture contours of objects with large variations in pose and appearance, we adopt the similarity measure of HOG feature between the evolving contour and the contour of a class as the evaluation criteria rather than relying on strong shape priors. We encode the joint distribution of the edgelet feature, the HOG feature and the curvature feature of an object in a mixture of Gaussian model. Classification is achieved by computing the posterior of the evolved contour conditioned on the three types of features. Our method is extensively evaluated on the UCF sports dataset, the Caltech 101 dataset, and the INRIA pedestrian dataset. Results show that our method achieves improved performance for the recognition task.

1 Introduction

Automatic understanding of objects has been receiving much attention in computer vision community with applications in object recognition and action recognition. In this work, we are interested in learning objects by their contours.

An object can be effectively characterized by its contour. Caselles *et al.* [1] introduced the concept of geodesic active contours, which defines the essential function in the Riemannian space and applies the energy reducing form to acquire contours. To obtain more accurate contours for a class of objects, Leventon *et al.* [13] utilized the curvature prior as the shape prior for different classes of objects to guide contour evolution. One of the popular ideas in incorporating shape prior is to project it into a subspace. Thus, the lower dimensional representation can be learned and the energy function is minimized in the learned latent space. Etyngier *et al.* [6] proposed a non-linear manifold learning method for learning shape prior. Specifically, a diffusion map is constructed using Nyström extension to learn shapes in a low-dimensional subspace. Cremers [3] presented another non-linear shape prior

http://dx.doi.org/10.5244/C.26.15

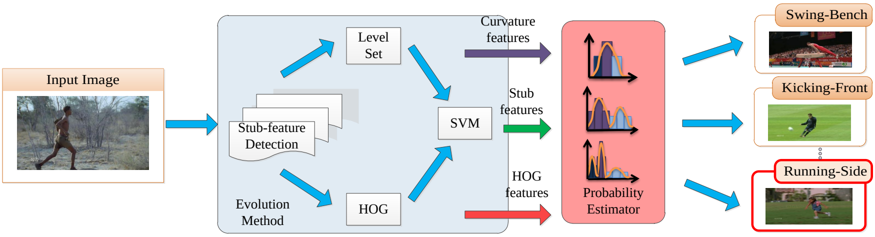


Figure 1: Flowchart of our method. Our method introduces the concept of stub features as potential edges for an object contour. The proposed evolution method uses stub features as a rough contour and iteratively approaches the target contour. HOG features are utilized as the global perspective of the contour and used as evaluation criteria of our evolution method.

using autoregressive (AR) model to approximate linear combination. Jiang *et al.* [10] utilized a TPS-RPM based method, in which TPS-RPM is used to measure similarity between two points sets. Prisacariu and Reid [16] introduced a non-linear shape prior and applied it to segmentation and object tracking system based on the elliptic Fourier shape descriptor (EFD) and Gaussian process latent variable models (GP-LVM).

Another line of work uses edges to describe objects. Edge-based models provide local perspective of an object and are robust when part of the object is occluded. However, since the global perspective is missing, the arrangement of the edge features, such as the pairwise relationships between edge features [5, 12] or the relative positions of edge features with respect to the centroid of the shape [15], is exploited to improve the edge-based models. Wu and Nevatia [20] proposed an edgelet feature based method for detection of multiple, partially occluded humans. Their method uses part of human contour as a mask, and computes the similarity of the orientation and magnitude. A probabilistic model is then employed to estimate the joint distribution of the detected edgelet response. Sabzmeydani and Mori [18] proposed the shapelet features, a set of mid-level features, which focus on low-level local gradient information that are discriminative between pedestrian and non-pedestrian classes. Gao *et al.* [9] proposed an "adaptive contour features" based method for human detection and segmentation. Their method selects seed features according to the orientations and the relative strength of features.

In this paper, we propose a novel edge-based method for learning objects. The flowchart is illustrated in Figure 1. Given an image, our method first detects edgelet features as a rough contour for an object. Edgelet features indicate potential positions for the contour and may stop curve evolution. These positions are referred to as *stub features*. Object contour is adaptively refined by the level set method. The evaluation criteria for contour evolution is defined by the similarity between the evolving contour and the target contour computed by their HOG features [4]. Therefore the curve evolution method is referred to as the *Contour-HOG* method. We formulate the joint distribution of the edgelet feature, the HOG feature and the curvature of the evolved contour in a probabilistic model, and perform classification by computing the posterior of the evolved contour conditioned on the three types of features.

The Contour-HOG method is inspired by detection and segmentation methods. Compared with previous methods [2, 8, 20], our method uses stub features to roughly localize the target object. This allows us to accurately capture the contour of the object. Moreover, the method fuses both local and global features to better describe contours and thus improves the recognition accuracy.

2 Stub Feature Detection

Our method begins by detecting edgelet features [20]. We use these features to roughly find an object from an image and then run a contour evolution method to obtain the object contour. Edgelet features indicate potential segments of the contour and can stop the evolution procedure. Thus, these features are referred to as stub features in this work. Stub features are obtained by applying a local edge mask $M_{k,s}$ and used for representing a short segment of a curve. The mask $M_{k,s}$ is a function of scale s and shape k, and maps points to 1 or 0.

With the detected stub features, we compute their similarities with the stub features in training data under a predefined edge mask $M_{k,s}$. The similarities are further used in shape classification. A stub feature can be represented by the magnitude of gradient and its orientation. In this work, we assume that they are independent from each other. Then the similarity can be computed as

$$p(x,s,k) = p_o(x,s,k)p_g(x,s,k), \tag{1}$$

where x is the 2D coordinate for a stub feature, $p_g(x,s,k)$ represents the likelihood of a local gradient and a mask having the similar magnitude, and $p_o(x,s,k)$ is the distribution of a gradient sharing the same orientation with the mask. We define $p_g(x,s,k)$ as a Gaussian distribution:

$$p_g(x,s,k) = \frac{1}{\sqrt{2\pi}|\Sigma_g|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}[g(x)-\mu]^T\Sigma_g^{-1}[g(x)-\mu]\right). \tag{2}$$

Here, μ and Σ_g are the mean and standard deviation of gradient magnitude, respectively. $g(x)$ is the magnitude of gradient vector derived by operating a mask $M_{k,s}$ on an input image I: $g(x) = |\nabla I(M_{k,s}(x))|$.

Distribution $p_o(x,s,k)$ is computed by measuring the similarity between two orientation vectors of local regions in a testing image and a training image. The orientation vector $I_q(M_{k,s}(x))$ for a testing image is computed by quantizing gradient orientations of a local region on the image indexed by mask $M_{k,s}(x)$, and the orientation vector $I_t(M_{k,s})$ for a training image is computed by quantizing gradient orientations on a training image indexed by $M_{k,s}$. Then the similarity is computed by

$$p_o(x,s,k) = \frac{\|v(I_q(M_{k,s}(x)), I_t(M_{k,s}))\|_2^2}{NT^2}, \tag{3}$$

where $\|\cdot\|_2$ is the l_2 norm, N denotes the number of pixels selected by the mask and T represents the number of orientations. We quantize gradient orientations into T directions, then the distance between two quantized directions o_1 and o_2 is defined as

$$v(o_1,o_2) = \min\{\max\{o_1,o_2\} - \min\{o_1,o_2\}, \min\{o_1,o_2\} + T - \max\{o_1,o_2\}\}. \tag{4}$$

3 Evolution Model

3.1 Contour Representation

The evolution model starts with the detected stub features and adaptively selects the features as the contour for an object. In this work, we adopt the Elliptic Fourier descriptor (EFD) [11] to represent an object contour. EFD depends on the first-order derivatives of Fourier series parameterized by traversal distance parameter s:

$$\frac{\partial x}{\partial s} = \sum_{i=1}^{\infty} -\frac{2n\pi}{T} a_n \sin(\frac{2\pi ns}{T}) + \frac{2\pi n}{T} b_n \cos(\frac{2\pi ns}{T}), \tag{5}$$

$$\frac{\partial y}{\partial s} = \sum_{i=1}^{\infty} -\frac{2n\pi}{T} c_n \sin(\frac{2\pi ns}{T}) + \frac{2\pi n}{T} d_n \cos(\frac{2\pi ns}{T}). \tag{6}$$

In the evolution model, the curvature force is defined by both first-order and second-order derivatives of contours. The second-order derivatives of EFD are given by

$$\frac{\partial^2 x}{\partial s^2} = \sum_{i=1}^{\infty} -(\frac{2n\pi}{T})^2 a_n \cos(\frac{2\pi ns}{T}) - (\frac{2\pi n}{T})^2 b_n \sin(\frac{2\pi ns}{T}), \tag{7}$$

$$\frac{\partial^2 y}{\partial s^2} = \sum_{i=1}^{\infty} -(\frac{2n\pi}{T})^2 c_n \cos(\frac{2\pi ns}{T}) - (\frac{2\pi n}{T})^2 d_n \sin(\frac{2\pi ns}{T}). \tag{8}$$

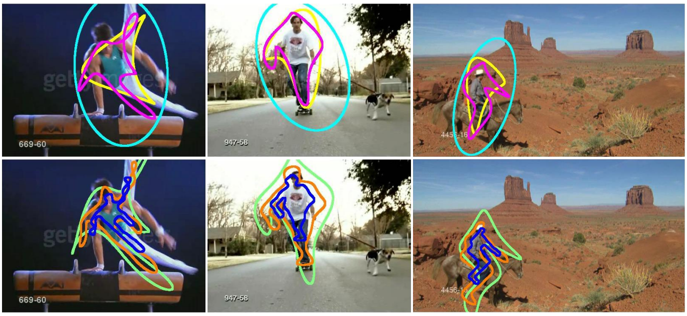


Figure 2: Examples of contour decomposition. The first row shows 1st, 4th and 5th components of human contour. The second row shows the 6th, 15th and 30th components of human contour.

3.2 Evolution Model

Our evolution model utilizes curvature force to guide contour evolution. The curvature force is defined as the ratio between the curvature of the shape prior κ_p and the curvature of the evolving contour κ_c: $f_\kappa = \frac{\kappa_p}{\kappa_c}$, where the curvature κ is given by

$$\kappa = \frac{x^{'}(s)y^{''}(s) - y^{'}(s)x^{''}(s)}{[(x^{'}(s)^2 + y^{'}(s)^2]^{3/2}}. \tag{9}$$

With the detected stub features, the contour generated by the evolution method iteratively approaches the target contour. To obtain the global perspective of an evolving contour, we compute the HOG feature of the evolving contour and measure its similarity with the contours of classes. The using of the global similarity measure of contours allows us to accurately obtain object contour. Examples of contour evolution are illustrated in Figure 2. The main steps of our shape evolution algorithm are listed in Algorithm 1.

Algorithm 1: Stub feature based Contour Evolution

```
Inputs: stub feature set $S = \{S_1, ..., S_N\}$, initial contour curvature $\kappa_p$ and threshold $\theta$.
Output: potential contour set $P$.
set potential contour set $P$ as empty;
for each stub feature $S_i$
    initialize level set according to the stub feature $S_i$;
    do
        set evaluation flag $e_f \longleftarrow$ non-ready and $e_v \longleftarrow 0$;
        set candidate set as empty;
        get current contour $C_c$;
        calculate current curvature $\kappa_c$ of the evolving contour;
        calculate curvature force using Eq.(9) and do linear interpolation;
        for i = 1: length($C_c$)
            if current contour points fall into one of stub features
                put current stub feature $S_i$ into the candidate set;
                set evaluation flag $e_f \longleftarrow$ ready;
            end if
        end for
        update curvature force in level set in $\kappa_t$ and $\kappa_c$;
        if $e_f$ = ready
            $e_v \longleftarrow$ evaluate HOG of current contour $C$ using SVM;
            if $e_v > \theta$
                put $e_v$ current $C_c$ and candidate set into the potential contour set $P$;
            end if
        end if
    while level set is not converged to level 0
end for
```

4 Object Classification

4.1 Classification Model

In our work, an object is classified by the similarity between the evolved contour of the object and the contour of a target object in the training dataset. The similarity is evaluated based on their curvatures, stub features and HOG features. We compute the similarity as

$$p(\kappa_c, \kappa_t, S_{\text{hog}}, S_{\text{stub}}) = p(\kappa_c|\kappa_t, S_{\text{hog}}, S_{\text{stub}})p(\kappa_t, S_{\text{hog}}, S_{\text{stub}}), \tag{10}$$

where κ_c and κ_t denote the curvature of an evolved contour and a target contour, respectively. S_{hog} is the affinity of contour-HOG features between an evolved contour and a target contour, and S_{stub} is the affinity of stub features between them. In our work, we assume κ_t and S_{hog} are independent of S_{stub}. Then Eq.(10) can be given by

$$p(\kappa_c, \kappa_t, S_{\text{hog}}, S_{\text{stub}}) = p(\kappa_c|\kappa_t, S_{\text{hog}}, S_{\text{stub}})p(\kappa_t, S_{\text{hog}})p(S_{\text{stub}}). \tag{11}$$

Here, we define $p(\kappa_c|\kappa_t, S_{\text{hog}}, S_{\text{stub}})$ as a Gaussian distribution:

$$p(\kappa_c|\kappa_t, S_{\text{hog}}, S_{\text{stub}}) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{||\kappa_c - \kappa_t||_2^2}{2\sigma^2}), \tag{12}$$

where σ is the standard deviation. In Eq.(11), distribution $p(\kappa_t, S_{\text{hog}})$ is the similarity of HOG features between an evolved contour and a contour of a class computed in the LINE 18 of Algorithm 1. $p(S_{\text{stub}})$ is the similarity of an edge segment and a predefined edge mask computed in Eq.(1).

4.2 Model Training

We use a likelihood function $\Lambda(Y;\theta)$ to measure the likelihood of a particular model with N training samples:

$$\Lambda(Y;\theta) = \prod_{n=1}^{N} f(y_n;\theta), \tag{13}$$

where y_n is the n-th training sample and θ represents a vector of parameters for the estimation function f. We decompose a training contour into harmonics of Fourier series, implying that the training contour can be viewed as summation of all harmonics. In this work, we define the joint likelihood function as a Gaussian mixture model (GMM):

$$\Lambda(Y;\theta) = \prod_{n=1}^{N} \sum_{k=1}^{K} w_k G(y;\mu_k,\delta_k), \tag{14}$$

where K is the number of components, w_k denotes the weight for the k-th component in the GMM, G represents the Gaussian distribution, μ_k is the mean of the k-th component, and δ_k is covariance of the k-th component. Our parametric model is learned as follows:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\{\Lambda(Y;\theta)\}. \tag{15}$$

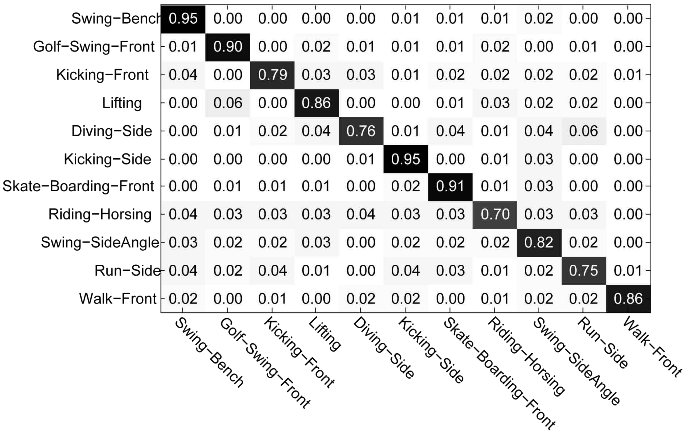
In this work, we use Expectation-Maximization (EM) algorithm to train the model.

5 Experiment

We test our method on three datasets, the UCF Sports dataset [17], the Caltech 101 dataset [7], and the INRIA pedestrian dataset [4]. UCF Sports dataset consists of 11 actions collected from broadcast media platform such as BBC and ESPN. Caltech 101 dataset contains 101 object classes. We select 10 classes from this dataset to test the performance of the method. Each class consists of 100 images (1000 images in total). On these two datasets, we evaluate the recognition accuracy of our method. We also apply our method to object detection and evaluate its performance on the UCF Sports dataset and the INRIA pedestrian dataset.

5.1 Recognition Results on UCF Sports Dataset

We test the recognition performance of our method on the UCF sports dataset. We randomly select 10 frames from a video as representative frames and use voting on these frames to classify the video. Confusion matrix is shown in Figure 3. Our method achieves 84.09% accuracy on this dataset. This result is encouraging since the method not only can recognize actions but also can segment people from images. Our method can recognize human actions with large variations in view-point and appearance. This is mainly due to the fact that it describes contours by multiple features and thus accurately captures contours of deformable



	Swing−Bench	Golf−Swing−Front	Kicking−Front	Lifting	Diving−Side	Kicking−Side	Skate−Boarding−Front	Riding−Horsing	Swing−SideAngle	Run−Side	Walk−Front
Swing−Bench	0.95	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.00	0.00
Golf−Swing−Front	0.01	0.90	0.00	0.02	0.01	0.01	0.01	0.02	0.00	0.01	0.00
Kicking−Front	0.04	0.00	0.79	0.03	0.03	0.01	0.02	0.02	0.02	0.02	0.01
Lifting	0.00	0.06	0.00	0.86	0.00	0.00	0.01	0.03	0.02	0.02	0.00
Diving−Side	0.00	0.01	0.02	0.04	0.76	0.01	0.04	0.01	0.04	0.06	0.00
Kicking−Side	0.00	0.00	0.00	0.00	0.01	0.95	0.00	0.01	0.03	0.00	0.00
Skate−Boarding−Front	0.00	0.01	0.01	0.01	0.00	0.02	0.91	0.01	0.03	0.00	0.00
Riding−Horsing	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.70	0.03	0.03	0.00
Swing−SideAngle	0.03	0.02	0.02	0.03	0.00	0.02	0.02	0.02	0.82	0.02	0.00
Run−Side	0.04	0.02	0.04	0.01	0.00	0.04	0.03	0.01	0.02	0.75	0.01
Walk−Front	0.02	0.00	0.01	0.00	0.02	0.02	0.00	0.01	0.02	0.02	0.86

Figure 3: Confusion matrix of our method on the UCF Sports dataset.

objects. Moreover, the evolution method segments target objects from cluttered images and provides the classifier with noiseless shape features.

We also illustrate the intermediate results given by the evolution method in Figure 4. Results demonstrate that the contours of people in large pose variations can be captured by the proposed evolution method. This method utilizes HOG feature similarity as the evaluation criteria rather than relying on a strong shape prior for each action class. This allows us to obtain accurate contours of people in drastic motion (e.g. Skate-Boarding). With accurate contours, our method is able to accurately infer the class for an unknown example.

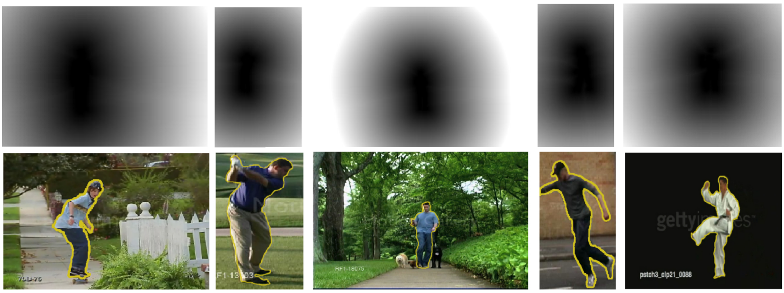


Figure 4: Intermediate results of our evolution model. The first row shows the detected stub features before starting level set evolution. The second row shows the results of segmentation. This figure is best viewed in color.

We compare the proposed method with methods in [17, 19, 21] on the UCF sports dataset. Results in Table 1 indicate that our method achieves comparable results with the dense HOG3D method in [19] but outperforms [17, 21]. It should be noted that, compared with these methods, our method can jointly recognize actions and segment target people from images. However, other comparison methods can only perform recognition.

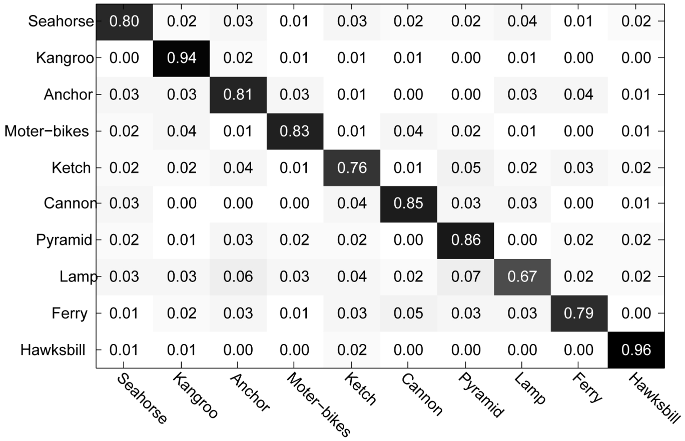
Table 1: Comparison results on the UCF sports dataset. Seg. and Rec. are short for segmentation and recognition, respectively.

Methods	Ours	Wang *et al.* [19]	Yeffet & Wolf [21]	Rodriguez *et al.* [17]
Accuracy	84.09%	85.06%	79.2%	69.2%
Func.	Seg. & Rec.	only Rec.	only Rec.	only Rec.

5.2 Recognition Results on Caltech 101 Dataset

We also evaluate the performance of our method on object recognition using Caltech 101 dataset. This dataset consists of 101 object classes. 10 classes are selected from this dataset to build a new dataset for evaluation: Seahorse, Kangaroo, Anchor, Motorbike, Ketch, Cannon, Pyramid, Lamp, Ferry, and Hawksbill. There are 100 images for each class, to provide 1000 images in total.

Our method achieves 82.7% recognition accuracy on Caltech 101 dataset. Confusion matrix is displayed in Figure 5. It should be noted that the method distinguishes different object classes by shapes of objects rather than popularly used local salient features such as SIFT features [14]. The extraction of contours allows us to jointly recognize and segment objects. Our method fuses multiple features including the curvature features, local stub features and global HOG features for recognition. These features capture different perspectives of shapes and thus reduce misclassification rates. Moreover, level set applied in our method provides clean shape features and thus benefits the recognition task. Most of the confusion is due to occlusions and illumination variations. Another reason for misclassification may be due to similar parts are shared in two different classes, e.g. Lamp and Pyramid. Figure 6 shows segmentation results given by the evolution model. Results show that our evolution method precisely extracts objects of interest from cluttered images. Thus, clean shape features can be computed from the extracted objects and further facilitate the recognition.



	Seahorse	Kangroo	Anchor	Moter−bikes	Ketch	Cannon	Pyramid	Lamp	Ferry	Hawksbill
Seahorse	0.80	0.02	0.03	0.01	0.03	0.02	0.02	0.04	0.01	0.02
Kangroo	0.00	0.94	0.02	0.01	0.01	0.01	0.00	0.01	0.00	0.00
Anchor	0.03	0.03	0.81	0.03	0.01	0.00	0.00	0.03	0.04	0.01
Moter−bikes	0.02	0.04	0.01	0.83	0.01	0.04	0.02	0.01	0.00	0.01
Ketch	0.02	0.02	0.04	0.01	0.76	0.01	0.05	0.02	0.03	0.02
Cannon	0.03	0.00	0.00	0.00	0.04	0.85	0.03	0.03	0.00	0.01
Pyramid	0.02	0.01	0.03	0.02	0.02	0.00	0.86	0.00	0.02	0.02
Lamp	0.03	0.03	0.06	0.03	0.04	0.02	0.07	0.67	0.02	0.02
Ferry	0.01	0.02	0.03	0.01	0.03	0.05	0.03	0.03	0.79	0.00
Hawksbill	0.01	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.96

Figure 5: Confusion matrix of our method on the Caltech 101 dataset.

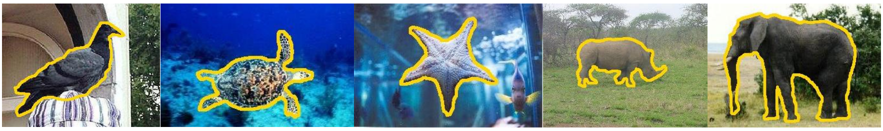
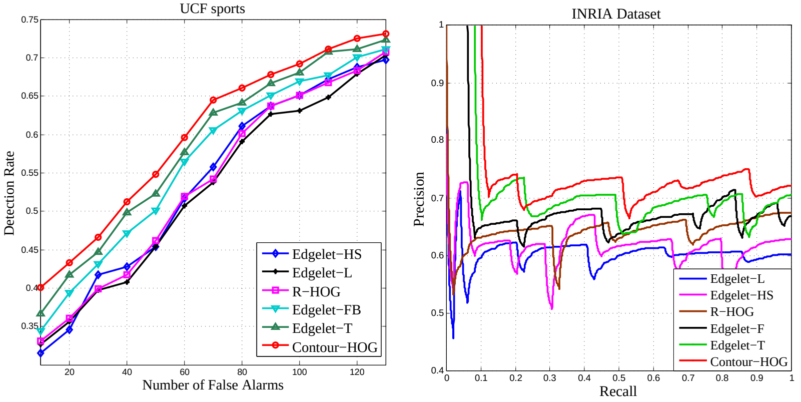


Figure 6: Segmentation results on Caltech 101 dataset. This figure is best viewed in color.

5.3 Detection Results

To assess the performance of our method on detection, we evaluate it on the UCF sports dataset and the INRIA pedestrian dataset. Both datasets contain variations in scales, poses, and clothing styles. We compare our method with four variant detectors in [20]: edgelet of head shoulder (Edgelet-HS), legs (Edgelet-L), full body (Edgelet-FB) and torso (Edgelet-T),



(a) ROC curves of evaluation as detector on the UCF sports dataset.

(b) Precision-recall curves of methods on the INRIA dataset.

Figure 7: Detection results of methods on the UCF sports dataset and INRIA dataset.

and rectangular HOG detector (R-HOG) in [8]. Wu and Nevatia [20] used edge features and Felzenszwalb *et al*. [8] focused on both local and global features. Our method applies level set to obtain contours and uses support vector machine as the detector.

In the first experiment, we test detection performance of methods on the UCF sports dataset. The experiment mainly focuses on detecting single pedestrian, and we ignore the case if multiple pedestrians are detected. There are 13 types of actions in the dataset. We randomly select 200 images from each action class to provide 2600 images in total. The experiment is conducted by randomly selecting one image from the dataset, and calculate the intersection of the detected bounding box and the ground truth over the whole area of ground truth, until the number of false alarms reaches upper bound.

Results in Figure 7(a) show that our method outperforms all the comparison methods. Compared with methods in [8, 20], our method applies level set method to obtain the contour of an object and thus provides the detector with much cleaner features. It should be noted that methods in [8, 20] offer the bounding box of an object while our method gives the detection results at contour-level which provide more accurate feature for an object.

Another test set is selected from INRIA pedestrian dataset which contains 1132 positive samples and 453 negative samples. All samples are mixed together and one image is tested at a time. We compare our method with methods in [8, 20] on this dataset. Precision-Recall curves are illustrated in Figure 7(b). Results also show that our method outperforms all the other comparison methods.

6 Conclusions and Future Work

We have proposed a novel edge-based method for object recognition. Our method introduces the concept of stub features which can be utilized for roughly localizing objects in images. This allows our contour evolution method to accurately obtain the contour of an object. We use the HOG feature in the evolution method as the evaluation criteria to achieve the contour of a deformable object with large variations. The contour is described by multiple features: the stub feature, the HOG feature and the curvature feature. We characterize those features in a joint distribution. Classification of an object is computed by the posterior of its contour.

Our method is extensively evaluated on three datasets and shows promising results.

In this paper we have assumed that there is only one object in an image. We are going to extend the method to jointly recognize and segment multiple objects in future work.

Acknowledgement

This research is supported in part by the NSF CNS 1135660, Office of Naval Research award N00014-12-1-0125, and U.S. Army Research Office grant W911NF-11-1-0365.

References

[1] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31, 1993.

[2] Jason Chang and John W. Fisher III. Efficient mcmc sampling with implicit shape representations. In *CVPR*, 2011.

[3] D. Cremers. Nonlinear dynamical shape priors for level set segmentation. In *CVPR*, 2007.

[4] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 2, pages 886–893, 2005.

[5] S. M. Ali Eslami, Nicolas Heess, and John Winn. The shape boltzmann machine: a strong model of object shape. In *CVPR*, 2012.

[6] Patrick Etyngier, Florent Ségonne, and Renaud Keriven. Active-contour-based image segmentation using machine learning techniques. In *MICCAI*, pages 891–899, 2007.

[7] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *PAMI*, 28(4):594–611, 2006.

[8] P. Felzenszwalb, David McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, 2008.

[9] Wei Gao, Haizhou Ai, and Shihong Lao. Adaptive contour features in oriented granular space for human detection and segmentation. In *CVPR*, pages 1786–1793, 2009.

[10] Tingting Jiang, Frédéric Jurie, and Cordelia Schmid. Learning shape prior models for object matching. In *CVPR*, 2009.

[11] F. P. Kuhl and C. R. Giardina. Elliptic fourier features of a closed contour. *Computer Graphics and Image Processing*, 18(3):236–258, 1982.

[12] M. Leordeanu, M. Hebert, and R. Sukthankar. Beyond local appearance: Category recognition from pairwise interactions of simple features. In *CVPR*, 2007.

[13] Michael E. Leventon, W. Eric L. Grimson, Olivier Faugeras, and William M. Wells III. Level set based segmentation with intensity and curvature priors. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 4–11, 2000.

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

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

[16] Victor Adrian Prisacariu and Ian Reid. Nonlinear shape manifolds as shape priors in level set segmentation and tracking. In *CVPR*, 2011.

[17] M. Rodriguez, J. Ahmed, and M. Shah. Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *CVPR*, 2008.

[18] Payam Sabzmeydani and Greg Mori. Detecting pedestrians by learning shapelet features. In *CVPR*, 2007.

[19] H. Wang, M. Ullah, A. Klaser, I. Laptev, and C. Schmid. Evaluation of local spatio-temporal features for action recognition. In *BMVC*, 2009.

[20] Bo Wu and Ram Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *ICCV*, pages 90–97, 2005.

[21] L. Yeffet and L. Wolf. Local trinary patterns for human action recognition. In *CVPR*, 2009.