Abstract

In this paper, we address the problem of representing objects using contours for the purpose of recognition. We propose a novel segmentation method for integrating a new contour matching energy into level set based segmentation schemes. The contour matching energy is represented by major components of Elliptic Fourier shape descriptors and serves as a shape prior to guide the curve evolution. The contours in training dataset serve as templates and are utilized to infer the category of an unknown image based on matching. Our method is evaluated on the UCF sports dataset and Caltech 101 dataset. Experiments show that our method achieves promising recognition accuracy and is robust to noisy low-level features and background clutter.

1. Introduction

In image segmentation and shape recognition, models are usually built upon both low-level features (e.g., intensity and texture) and prior knowledge about target objects. Currently, integrating the the model-based process and data-driven process has received much attention in image segmentation. In this paper, we consider prior knowledge given by the shape of human or object, and focus on how to exploit such knowledge for shape recognition.

Level set based contour representations have become a widely used approach in image segmentation [11, 7]. They enable us to elegantly represent topological changes of the boundary which is implicitly represented. Level set segmentation methods can be used to utilize various low-level image features (e.g., intensity, color, texture, etc). However, as shown in [7], low-level features can be noisy and thus reduce segmentation performance. To overcome this problem, some studies have investigated the integration of shape prior knowledge into level set method. Leventon et al.[13] utilized shape and pose information derived by principle component analysis (PCA) on a given contour. Etyngier et al.[8] learned shape prior using a non-linear manifold learning method. Unlike [13], the algorithm in [8] extracts shape prior by manifold learning, and in particular, a diffusion map is constructed using Nyström extension. Cremers [5] approximated linear combination by a non-linear shape prior using autoregressive (AR) model. Jiang et al.[9] measured the similarity between two points sets by a TPS-RPM based method.

In this paper, we propose novel shape constraints to guide contour evolution. The flowchart of our method is shown in Figure 1. An object or a human is represented by a set of Fourier coefficients. We introduce a new curve energy function based on curvature to measure the shape similarity between a target curve and an evolving curve. The curve energy is smoothly integrated with region energy to guide segmentation. This makes the segmentation process robust to noisy low-level features. We use the obtained contours as templates for matching and recognize human actions or objects based on these contours.

2 Our Method

2.1 Contour Representation

The shape of an object can be effectively represented by Fourier series of its contour [11]. In our work, we
adopt the Elliptic Fourier descriptor (EFD) parameterized by $s$ to represent an object contour:

$$\frac{\partial x}{\partial s} = \sum_{n=1}^{\infty} \frac{2n\pi}{T} a_n \sin \left( \frac{2n\pi s}{T} \right) + \frac{2n\pi}{T} b_n \cos \left( \frac{2n\pi s}{T} \right),$$

$$\frac{\partial y}{\partial s} = \sum_{n=1}^{\infty} -\frac{2n\pi}{T} c_n \sin \left( \frac{2n\pi s}{T} \right) + \frac{2n\pi}{T} d_n \cos \left( \frac{2n\pi s}{T} \right),$$

where $x$ and $y$ are 2D coordinates on an image, and $T$ is the period.

A contour in our work is described by the first $N$ order Fourier coefficients (Figure 2). These coefficients characterize the shape of an object and are utilized as the shape prior to guide curve evolution. This type of shape prior plays an important role in curve evolution and thus facilitates the recognition task.

We use convex keypoints for initialization [4]. The keypoints are defined as multi-scale organization of curvature zero crossing points of a planar curve: $\frac{\partial^2 y}{\partial x^2} = 0, \frac{\partial^2 y}{\partial x^2} = 0$. The above conditions show that a convex keypoint is an inflection point if the curve is continuous. With alignment of such a set of convex points, our level-set based method can start.

### 2.2 Contour Evolution

Our method is based on a region-based level set scheme introduced by Chan and Vese [16]. Their method generates a segmentation of an input image with two gray value constants. However, their method relies on low-level features and may fail to converge to the desired segmentation due to unfavorable background clutter. To cope with such degraded low-level information, we propose a novel curvature matching energy function $E_{\text{curve}}$ to guide contour evolution.

Our curvature matching energy function measures shape similarity between a target object and the object to be segmented. We define the shape similarity of two objects by the curvatures on corresponding segments. Let $C(s)$ be a curve parameterized by $s$, $C_{\psi(0)}(s)$ be a target curve, and $C_{\text{evolve}}(s)$ be an evolving curve. Then the curvature of the target curve $\kappa_{\psi}$ and the evolving curve $\kappa_C$ can be given by

$$\kappa_{\psi} = \nabla \cdot \frac{\nabla C_{\psi(0)}(s)}{||\nabla C_{\psi(0)}(s)||}, \quad \kappa_C = \nabla \cdot \frac{\nabla C_{\text{evolve}}(s)}{||\nabla C_{\text{evolve}}(s)||}.$$

We use the ratio of curvature $\kappa(s) = \frac{\kappa_{\psi}}{\kappa_C}$ to denote the shape similarity of a local segment $s$. Then the shape matching energy is defined as

$$E_{\text{curve}} = \mu \int_{\psi(s) \equiv 0} |\nabla H(\psi(C(s))\kappa(s))| ds,$$

where $H(z)$ is the Heaviside function: $H(z) = 1$ if $z \geq 0$, otherwise $H(z) = 0$. $\psi(C(s))$ is used to find corresponding segments between a target curve and an evolving curve.

Shape energy $E_{\text{curve}}$ encodes the shape information of an object and can provide the shape prior for segmentation. In this work, we incorporate this shape energy into the region energy to guide curve evolution. Therefore, the total energy is expressed as

$$E = E_{\text{curve}} + E_{\text{region}}$$

$$= \mu \int_{\psi(s) \equiv 0} |\nabla H(\psi(C(s))\kappa(s))| ds$$

$$+ \lambda_1 \int_{\Omega_1} |\mu_0(x, y) - c_1|^2 H(\psi(x, y)) dxdy$$

$$+ \lambda_2 \int_{\Omega_2} |\mu_0(x, y) - c_2|^2 (1 - H(\psi(x, y))) dxdy,$$

where $E_{\text{region}}$ is the inhomogeneity energy [16], $\mu$, $\lambda_1$ and $\lambda_2$ are parameters, $\Omega_1$ denotes the region enclosed within $C_{\psi(0)}(s)$, $\Omega_2$ is the region outside $\Omega_1$.

### 2.3 Discussion

The advantage of the energy function in Eq.(2) is that it can remove background noise caused by local shape irregularities (see Figure 3).

Let $\hat{E}$ be an estimator. Suppose an evolving contour $X$ is estimated by an observed contour vector $Y$ and $\hat{E}[X^2] < +\infty$. We define $X, Y_1, ..., Y_n$ as random processes with finite second moments; all in the same probability space. A guiding curve $Y$ can be decomposed into harmonics of Fourier Series. Here we assume $\hat{E}[Y_i] = 0$ for all $i = 1$ and $\hat{E}[Y_i Y_j] = 0$ for $i \neq j$, then

$$\hat{E}_{\text{curve}}[X|Y_1, Y_2, ..., Y_n] = \hat{E}[X] + \sum_{i=1}^{n} \hat{E}[X - \hat{E}[X]|Y_i].$$

Evolving error is given by

$$\hat{e}_{\text{curve}} = X - \hat{E}_{\text{curve}}[X|Y_1, Y_2, ..., Y_n].$$

(Figure 2. [Best viewed in color.] Example of contour decomposition.)
Recall that local shape irregularities are often represented by short period components in frequency domain. According to Eq.(2) and Eq.(3), these irregularities can be reduced during energy optimization. Consequently, accurate contours can be obtained and thus the recognition accuracy can be improved.

2.4 Shape Matching

Many effective methods have been proposed to identify shape, such as Signed Distance Map (SDM) [13], Curvature Scale Space (CSS) [10], etc. In our work, we use the SKS algorithm [10] to compare two shapes. SKS algorithm is a robust 2D shape recognition algorithm which uses the evidence accumulation.

Let feature vector \( v = (\kappa_1, \kappa_2, \ldots, \kappa_n) \) be a vector of scaled curvatures with respect to a reference point. The similarity of two shapes can be computed by \( s_k = e^{-\frac{\|v-v_k\|}{2\sigma^2}} \). We use \( s_k \) as the shape similarity and find the best match for an unknown image from training dataset.

3 Experiments

We compare the proposed method, namely decomposed contour curvature prior (DCCP) with embedded contour curvature prior (ECCP) on UCF sports dataset and Caltech 101 dataset. First, we compare recognition accuracies of level set based methods with DCCP and ECCP on UCF dataset. Then segmentation results are compared on Caltech 101 object dataset. We assume there is only one object of interest in an image.

3.1 Results on UCF-Sports Dataset

On this dataset, we compare the recognition accuracy of our method with 6 benchmark level set based methods [1, 2, 16, 14, 12, 15]. Since these methods do not have shape prior, to conduct a fair comparison, we embed contour curvature prior ECCP into these methods using the embedding strategy described in [6].

The confusion matrix of our method is shown in Figure 4. Our method achieves 60.77% recognition accuracy. Comparison results with the other 6 methods are shown in Table 1. Results indicate that our method achieves the best overall accuracy over the other methods in comparison. The underlying reason is that our method reduces local irregularity which essentially is high-frequency noise. Our method decomposes the contour curvature into components and selects the top \( N \) components. Thus high-frequency noise can be reduced and accurate contours can be obtained. Consequently, recognition accuracy can be improved with accurate contours.

To further investigate the noise reduction in our method, we compare our method with Zhu and Chan’s method [3]. We show intermediate results (the number of iterations is set to 100) in Figure 6. Our method can efficiently remove background noise due to heterogeneity of local regions. The proposed method ignores high-frequency components which are sensitive to background noise and thus achieves robust segmentation results. With accurate contours, our method achieves higher recognition accuracy.

3.2 Results on Caltech 101 Dataset

We compare segmentation results with the method in [3]. Results are shown in Figure 5. It can be seen that our method is not sensitive to heterogeneity of local regions. Our method selects components which are insensitive to background noise while Zhu and Chan’s
Table 1. Recognition results (%) of our method and methods in [1, 2, 16, 14, 12, 15].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>44.61</td>
<td>46.92</td>
<td>50.77</td>
<td>46.92</td>
<td>53.85</td>
<td>50.77</td>
<td>60.77</td>
</tr>
</tbody>
</table>

Figure 6. Contour evolution of our method (top row) and the method in [3] (bottom row).

method considers all high-frequency components which would suffer from local irregularities.

4 Conclusion

We have proposed a decomposed contour curvature prior for level set method for segmentation and shape recognition. Our method ignores high-frequency components and can reduce background noise such as heterogeneity of local regions and irregularities. We use segmentation results to conduct shape recognition task. Results on two datasets show that our method is insensitive to background noise and achieves promising results in recognition.

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