Discriminative Level Set for Contour Tracking*

Wei Li[1], Xiaoqin Zhang[1], Jun Gao[1], Weiming Hu[1], Haibin Ling[2], Xue Zhou[3]
[1]*National Laboratory of Pattern Recognition, Institute of Automation, Beijing, China*
{*weili, xqzhang, wmhu, jgao*} @*nlpr.ia.ac.cn*
[2]*Dept. of Computer and Information Sciences Temple University*
{*hbling@temple.edu*}
[3]*College of Automation, University of Electronic Science and Technology of China*
{*zhouxue@uestc.edu.cn*}

Abstract

Conventional contour tracking algorithms with level set often use generative models to construct the energy function. For tracking through cluttered and noisy background, however, a generative model may not be discriminative enough. In this paper we integrate the discriminative methods into a level set framework when constructing the level set energy function. We train a set of weak classifiers to distinguish the object from the background. Each weak classifier is designed to select the most discriminative feature space and integrated via AdaBoost according to their training errors. We also introduce a novel interaction term to explore the correlation between pixels near the object edge. This term together with the discriminative model both enhance the discriminative power of the level set. The experimental results show that the contour tracked by our approach is more accurate than the conventional algorithms with the generative model. Our algorithm successfully tracks the object contour even in a cluttered environment.

1. Introduction

Contour tracking, as one of the fundamental tasks in computer vision, aims to obtain the contour of the object in each frame rather the location. It receives more and more attention due to its crucial value in many applications such as action recognition, human-computer interfaces, and augmented reality.

Level set [1] is an implicit way to represent the contour of the object. The basic idea is to represent contour as zero level set with an implicit function defined in a higher dimension. Due to its efficiency in obtaining the contour of the object, recently years witness greatly advance of applying the level set into contour tracking area. Paragios el al. [2] use the Geodesic Active Region to track the contour of moving object. Mansouri el al. [3] propose a novel method to formulate contour tracking as a Bayesian estimation problem, with no motion model assumed. In [4], the probability density function (PDF) of texture and color features is fused in a Bayesian inference framework to construct the energy function for the level set evolution. In general, these methods share the following characters: 1) a region and edge term are combined together to design a energy function; 2) the region term incorporates the motion or color information, while the edge term explores the gradient constrain of the object. However these algorithms can not provide a reliable contour when the background is cluttered and noisy. They are apt to be distracted by similar background regions around the contour.

In [5],[6],[7], researchers consider a global shape consistency to constrain the level set evolution. The statistical property of shape prior for the object to be tracked is learned before tracking. This shape prior term is combined into the energy function to improve the performance of the level set in a cluttered and noisy background. However the shape consistency does not always hold and the learning process is time-consuming. Cremers [8] employs a dynamical statistical shape priors as the constrain of the level set. It explores the temporal coherence of the silhouettes to obtain the shape prior of current frame. This approach is only suitable for the human tracking with a regular movement. Parts of the sequence are needed before tracking to train the shape transition matrix.

Recently, discriminative models have received more and more attention because their effectiveness [9, 10].

*This work was supported in part by the NSFC (Grant No. 60825204, 60905015 and 60935002) and the National 863 High-Tech Program of China (Grant No. 2009AA01Z318).

Inspired by Avidan's work [10], we employ the discriminative model to increase the discriminative power of level set in a cluttered and noisy background. A set of weak classifiers are combined into a strong one based on their discriminative ability between the object and the background. Each weak classifier corresponds to a selected feature space which is a linear combination of R,G,B color space. So the strong classifier integrates the most discriminative feature space together to distinguish the object from the background. With this strong classifier, the probability of each pixel belonging to the object and background is computed. Then we integrate the strong classifier into the level set energy function to obtain the contour.

The main contribution of this paper concentrates on the following two parts. (1) We introduce a novel discriminative model into the level set based contour tracking. (2) A novel interaction term is proposed to take the pixel dependence into consideration. This term further increases the discriminative power of our algorithm.

The rest of the paper is organized as follows. Section 2 presents the definition of weak classifier and the AdaBoost algorithm to boost them. In section 3, we present a new level set energy function integrated with the AdaBoost algorithm and the novel interaction term. The experimental results are shown in section 4, which is followed by the conclusion in section 5.

2. Discriminative Model

The discriminative model focuses on maximizing the difference between the object and the background. In this paper, the difference is mainly embodied in weak classifiers. Each weak classifier is composed by a hyper-plane of a certain feature space to distinguish the object and the background. AdaBoost is adopted to boost the weak classifiers into a strong one. The weak classifiers selected by the AdaBoost are complimentary to each other.

2.1. Weak Classifier

In order to increase the discriminative power of the weak classifier, we consider each weak classifier as a feature selector. If a weak classifier is selected as a candidate for the ensemble in the AdaBoost algorithm, the feature connected with this weak classifier is the one possess the highest discriminative power. The feature pool used in this paper is a linear combination of the R,G,B color space.

$$F \in \{\alpha_1 R+\alpha_2 G+\alpha_3 B\}, \alpha_i \in \{-2,-1,0,1,2\} \quad (1)$$

After eliminating some redundant features, the feature set can finally be reduced to 49. This feature set is not casually defined. Many common features are included in this set such as intensity $R+G+B$, approximate chrominance feature $R-B$, excess color feature $2G-R-B$. The effectiveness of this feature pool is validated in [13].

Assume that $\{x_i, y_i\}_{i=1}^N$ are examples and their labels, $y_i \in \{0,1\}$, x_{f_t} is the value in feature space $f_t \in F$, N and T are the number of sample and weak classifier respectively. The weak classifier is defined as a hyperplane to separate the object from the background for feature f_t:

$$h_t(x_{f_t}) = \mathrm{sign}(h^T x_{f_t}) \quad (2)$$

The parameter h is solved using the least square regression:

$$h = (A^T w^T W A)^{-1} A^T w^T W y \quad (3)$$

where W is the matrix of the weights for each example and A is the matrix of the examples.

2.2. Strong Classifier

AdaBoost [11] is validated to be one of the most effective discriminative models at separating the positive samples from negative samples. The weak classifier selected by AdaBoost compensates the previously chosen one through adjusting the weights of the samples. The weight of each sample is computed according to the error rate of previously selected weak classifier. At the beginning of the training, each sample is initialized with a equal weight $\{w_i\}_{i=1}^N$. The training error of each weak classifier is defined as follows:

$$error_t = \sum_{i=1}^{N} w_i \mathrm{sign}(h(x_{i,f_t}) \neq y_i) \quad (4)$$

The weight α_t of each weak classifier is given by:

$$\alpha_t = \frac{1}{2}\ln(\frac{1-error_t}{error_t}) \quad (5)$$

At each iteration, the weak classifier with the minima error rate is chosen as the candidate one for boosting. At the end of tth iteration, we update the weights of the samples:

$$w_i \leftarrow w_i \exp\{-\alpha_t y_i \mathrm{sign}(error_t)\} \quad (6)$$

After T weak classifiers are selected, the strong classifier is constructed according to their weights.

$$H(x) = \sum_{t=1}^{T} \alpha_t h_t(x_{f_t}) \quad (7)$$

3. Discriminative Level Set

We introduce a new energy function integrated with the discriminative model. Meanwhile, a novel interaction term which considers the correlation between pixels is introduced.

3.1. Energy Function With Discriminative Model

The level set representation of the contour evolving with time t is given by:

$$\phi(x,t) = \begin{cases} 0 & x \in C(t) \\ -d(x, C(t)) & x \in R_{out}(t) \\ d(x, C(t)) & x \in R_{in}(t) \end{cases} \quad (8)$$

where $x \in R^2$ is the position of a pixel, $d(x, C(t)$ is the Euclidean distance from the point x to contour $C(t)$.

Denote $\Omega_{obj}, \Omega_{bck}$ as the object and background regions, $v(x)$ as the value of a pixel x. The energy function is defined from the region competition view [12]:

$$E = \Sigma_{\Omega \in \Omega_{obj}, \Omega_{bck}} \int_{\Omega} \log p(v(x)|\Omega)dx + \lambda \int_{C} ds \quad (9)$$

where the first term is the probability of the current contour, the second term is the smoothness regulation, λ is the regularization parameter. The contour C is the result of the competition between the object and the background regions.

We use the sigmoid function to define probability of each pixel x given the contour:

$$p(v(x)|\Omega) = \frac{1}{1 + \exp\{-\sum_{t=1}^{T} \alpha_t h_t(x_{f_t})\}} \quad (10)$$

For each iteration of contour evolution, only the pixels around the contour contribute to the evolution. A narrow band around the contour instead of the whole object and background regions is adopted for the construction of the energy function. Thus the energy function with level set representation is defined as follows:

$$E(\phi, \Omega_{obj}, \Omega_{bck}) = \int_{0}^{k} \Upsilon(\phi) \log p(v(x)|\Omega_{obj})dx +$$

$$\int_{-k}^{0} (1 - \Upsilon(\phi)) \log p(v(x)|\Omega_{bak})dx + \lambda \int_{\Omega} \nabla |\Upsilon(\phi)| dx \quad (11)$$

where k is the size of band, Υ is a Heaviside function.

3.2. The Interaction Term

In the conventional level set energy function, pixels are treated as independent of each other. However the value of each pixel is always influenced by the pixels around it. The values between two nearby pixels can not vary too much due to the consecutive constrain. So we introduce a novel interaction term into the energy function to explore the dependence between pixels. This term further increases the discriminative power of our algorithm which is validated in the experiments.

The probability of each pixel is revised with:

$$\widetilde{p}(v(x_i)|\Omega) = \frac{1}{1 + \exp\{-\frac{1}{2}(H(x_i) + \frac{1}{N} \sum_{j \in N_i} y_{i,j} H(x_j))\}} \quad (12)$$

where N_i is the 8-neighborhood of pixel i, $y_{i,j} = |y_i - y_j|$ is a edge label, y_i is defined as follows:

$$y_i = \text{sign}(\sum_{t=1}^{T} \alpha_t h_t(x_{i f_t})) \quad (13)$$

From the definition of the interaction term, the probability of pixel x is influenced by the pixels around it. If the pixels near x possess the same label with it, the probability changes little. However if the pixels near x are not identical with it, pixel x is considered as a noisy one. Its probability is changed according to the pixels near it.

3.3. Evolution Function

Differentiating the energy function with respect to C results in the evolution equation of the level set function:

$$\frac{d\phi(x,t)}{dt} = \frac{dE}{dC} = \delta(\phi)(2 \log \widetilde{p}(v(x)|\Omega) + \lambda k)| \bigtriangledown \phi| \quad (14)$$

where k is the curvature of the contour.

With the speed function, the contour evolves to the desired boundary by modifying ϕ iteratively:

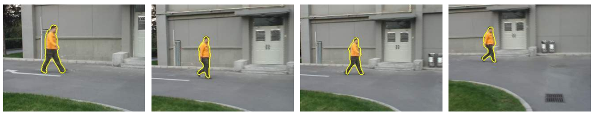
$$\phi_t = \phi_t + \Delta t(\log \widetilde{p}(v(x)|\Omega) + \lambda k)| \bigtriangledown \phi| \quad (15)$$

where Δt is the parameter to control the speed.

4. Experiments

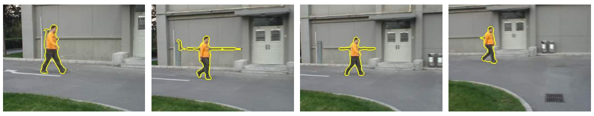
In order to show the effectiveness of our algorithm at improving the discriminative power of the level set evolution, we evaluate the proposed model on three sequences. The parameters of the weak classifiers are trained in the first frame using the contour given by hand.

In the first experiment, we track the contour of a pedestrian under an outdoor environment. The legs of the people share similar color with the ground especially the shoes. As illustrated in Fig 1, the first row



(a) Our approach

(b) Conventional level set



(c) Our approach without the interaction term

Figure 1. Tracking results of three different methods.

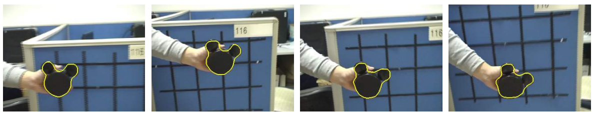
is the tracking performance of our algorithm, where the obtained contour tightly encloses the person we tracked. The second row gives the results of a state-of-art algorithm [4]. In [4] the generative model is employed to construct the energy function. In order to give a fair comparison, the interaction term is also used. It is clear that the generative model loses the discriminative power when the object shares the similar appearance with the background. The leg part of the person is not accurately tracked due to its similarity with the ground. In the third row, we implement our algorithm without the edge term. The tracking results show the contour is distracted by the bar in the wall and not tightly enclosed.

The second sequence is a Mickey head moving in a man-made circumstance. This experiment is designed to show the effect of the interaction term in resisting the distraction of the noisy background. The contour of the object is apt to be distracted by bars whose colors are the same with the object. As shown in Fig 2.(b), the contour is distracted by the bars in the background without using the interaction term. On the contrary, the tracking results in Fig 2.(a) illustrate that the actual contour is available with the interaction term .

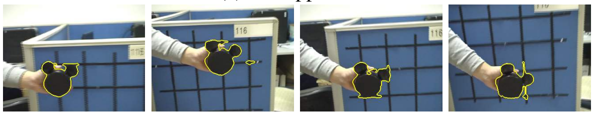
The last experiment is designed to show the ability of our algorithm to track the contour under a noisy environment. The car is moving in a frog weather and the background is noisy and cluttered. The tracking results in Fig 3 show our algorithm successfully tracks the contour of the car.

5. Conclusion

In this paper, the discriminative power of the level set is increased through introducing the discriminative model into the level set energy function. Also a new interaction term is introduced to make the level set robust against the noisy background. Experiments illus-



(a) Our approach



(b) Our approach without the interaction term

Figure 2. The comparison between our approach with the edge term and without the interaction term.

Figure 3. Vehicle tracking against noisy and cluttered backgrounds.

trate that our algorithm successfully increases the discriminative ability and obtains a better contour than the generative based one under a complex environment.

References

[1] S.Osher and J.Sethian, "Fronts propagation with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations", Journal of Computational Physics, vol.79, pp.12-49, 1988.1, 2

[2] N. Paragios and R. Deriche. "Geodesic active regions and level set methods fpr motion estimation and tracking", in CVIU, 2005.

[3] A.R. Mansouri, "Region tracking via level set pdes without motion computation", IEEE Trans.on PAMI, vol. 24, no. 7, pp. 947-961, 2002.

[4] A.Yilmaz, X.Li and M.Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras", IEEE Trans.ON PAMI, vol.26, pp.1531-1536, Nov. 2004.

[5] N.Paragios, M.Rousson, "Shape Priors For Level Set Representation", ECCV, pp.78-92, 2002.

[6] M. Rousson and D. Cremers, "Efficient kernel density estimation of shape and intensity priors for level set segmentation", MICCAI, volume 1, pages 757-764, 2005.

[7] M. Leventon, W. Grimson, and O. Faugeras, " Statistical shape influence in geodesic active contours", CVPR, volume 1,pages 316-323, 2000.

[8] D. Cremers, "Dynamical statistical shape priors for level set based tracking" in IEEE Trans", PAMI, vol.28, pp. 1262-1273, 2006.

[9] X. Zhang, W. Hu, S. Maybank, and X. Li, "Graph based discriminative learning for robust and efficient object tracking", In ICCV 2007.

[10] S. Avidan, "Support vector tracking", PAMI, 26:1064C1072, 2004.

[11] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of Computer and System Sciences, 55(1):119C139, 1997.

[12] S. C. Zhu, T. S. Lee, and A. L. Yuille, "Region competition: unifying snakes, region growing, energy/bayes/MDL for multi-band image segmentation", in Proc. ICCV, pp. 416C423,1995.

[13] R. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features", PAMI, 27(10):1631C1643, Oct. 2005.