Dynamic Scenario Segmentation based on Target Assumption Sort

Zifen He [1, a], Yinhui Zhang [1,b*]

[1]Kunming University of Science and Technology,Faculty of mechanical and electrical engineering, Kunming, 650500，China

[a]zyhhzf1998@163.com, [b*]yinhui_z@163.com

Keywords: Dynamic scenario segmentation;Target assumption;Markov random field;Sorting

Abstract. This paper enables robust foreground segmentation by sorting target proposals over a assumption space to achieve consistent target candidates and binary segmentation of a video sequence. This is followed by sorting the target candidates over a specific hypothesis space to yield consistent and dense object proposals. An efficient higher-order graph-cut method is adopted to optimize a Markov Random Field (MRF) model, which is instantiated by the estimated foreground hypothesis with the highest score. Compared with a state-of-the-art algorithm, our method results in better and robust segmentation performance when dealing with highly dynamic image sequences.

Introduction

Imagine a scenario where an autonomous robot is wondering in an unknown environment, the location information of foreground targets which can be inferred with an unsupervised method will certainly facilitate, to a large extent, subsequent recognition, tracking as well as visual servo manipulation of the targets. Despite the research efforts in dynamic scene segmentation, designing an algorithm that is robust for a wide variety of dynamic scenes that involve complex natural environments is still an open problem. The bottle neck in dynamic scene segmentation is the foreground object hypothesis stage. Currently, there are mainly two categories of methods for foreground hypothesis[1]. One approach takes advantage of such object appearance cues as color and texture. Typically, this approach consists of bottom-up and top-down segmentation paradigms. The bottom-up [2] paradigm exploits the appearance similarity of neighboring pixels, which are then merged into hyper pixels or image patches. Actually, bottom-up segmentation based on appearance cues always leads to over segmentation of dynamic scenes and cannot segment foreground objects from background scenes.

Target candidates generation

Due to the highly dynamic and complex nature of natural scenes, a set of target candidates are initially generated in order to form a pool of assumption. Instead of using deformable part model based object detection approaches, which produce object candidates that are represented in terms of image windows, we perform segmentation based object hypothesis by using low-level appearance features computed on super-pixels. We first compute super-pixels of the original images using SLIC algorithm[3].

Moreover, a specific energy function is formulated instead a set of parametric functions. In precise, we compute a binary segmentation by minimizing the energy:

$$E(X) = \sum_{u} U(x_u) + \sum_{(u,v)} V(x_u, x_v) \tag{1}$$

where u and v are indicx of super-pixels. U and V represent unary and pairwise potentials of the energy function E. For each super-pixel u.

Pairwise potentials are used to encode smoothness between super-pixels. Since the global probability boundary (gPb) algorithm [4] has relatively good performance in detecting salient contours in whole images, thus the standard CPMC algorithm employs gPb features to capture pairwise potentials.

Target assumption sorting

The target of sorting is to assign assumption that exhibit target-like regularities with higher sorting score. The feature consists of graph partition properties, region properties such as area, perimeter, bounding box location, major and minor axis lengths of the ellipse as well as Gestalt properties such as curvilinear continuity and convexity. Gestalt properties are computed in terms of χ^2 distance, which is based on color histogram measures between two half discs of a circular image patch.

$$\chi^2(h,g)=\sum_i \frac{(h_i-g_i)^2}{h_i+g_i} \tag{2}$$

where *h* and *g* represent color histograms of two half discs. The radius of each disc is fixed at 6 pixels, a value that was empirically seen to yield good results throughout the paper.

Spatio-temporal MRF segmentation

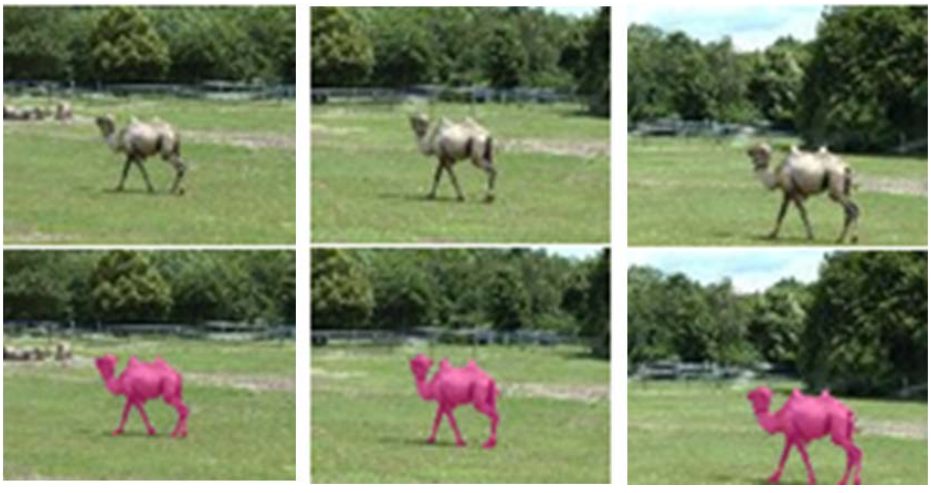
The popular spatio-temporal Markov random field (MRF) is used to model the correlation in spatial domain an along temporal axis. In particular, the MRF model consists of two unary terms and two pairwise terms. The unary terms exert appearance and position constraints on foreground object hypotheses. Following the method in [1], we formulate MRF model as

$$E(X)=\sum_{t,u} A^t(x_u)+\alpha_1\sum_{t,u} L^t(x_u)$$
$$+\alpha_2\sum_{u\overset{s}{\longleftrightarrow}v} V^t(x_u,x_v)+\alpha_3\sum_{u\overset{t}{\longleftrightarrow}v} W^t(x_u,x_v) \tag{3}$$

where *A* denotes appearance potential that encourage super-pixels with similar color values in the RGB color space to cluster in the same category.

Experiments

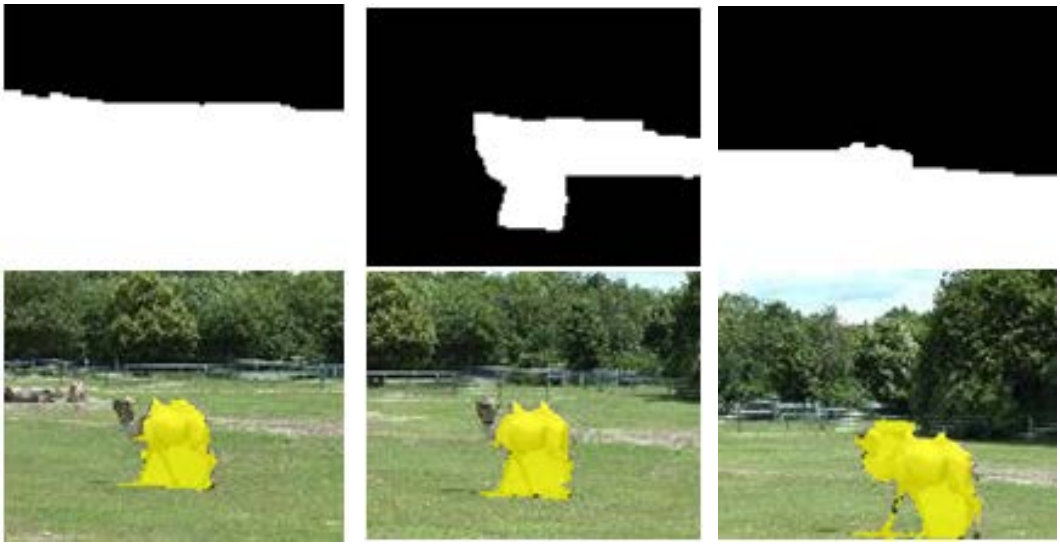
To demonstrate the effect of the proposed method, we use the Freiburg-Berkeley Motion Segmentation Dataset [5], which contains of 59 video sequences captured from real-world scenes. The camel01 data set is selected to test the algorithm because this sequence contains typical challenges such as background clutter, various and confusing object and camera motion, as well as dramatic shape articulation. There are 100 frames in this data set and 3 frames (#20, #60, #100) are manually annotated. A sample of the original image sequence and their corresponding ground truth masks are shown in Fig 1.



#20 #60 #100

Fig.1 Original video sequences and ground truth.

The segmentation result is shown in the second row of Fig 2. From the segmentation result we can find that, qualitatively, most of the foreground objects are segmented correctly using the unsupervised method. The first row of Fig 2 shows the target assumption result with the highest score using the proposed sorting method. From the assumption result we observe that frame #20 and frame #60 have good results for object candidates.The object region hypotheses in frame #20 and frame #100 classify the whole grass area as the foreground candidate. Following the same definitions as in [1], we initialize the unary and pairwise potentials of the spatio-temporal MRF using the target assumption with the highest score shown in the first row.



#20 #60 #100

Fig 2 The first row: object hypothesis with highest ranking score; The second row: segmentation results after MRF energy minimization.

To assess the segmentation result quantitatively, we use four segmentation criteria, which include Area Under Curve (AUC), Average Precision (AP), Correct Rate (CR) and CR in percent (CRp). The AUC and AP is derived from precision recall curve by using vl-feat toolbox [6] over each frame. Then the AUC and AP are averaged over six frames: #20,#60 and #100. The CR is computed using XOR operation of ground truth and segmented images and then minus by total number of pixels in each frame. The CRp is computed by dividing CR over the total number of pixels in each frame. The third row of Table 1 shows the quantitative segmentation result using the proposed algorithm and the respective standard division is shown in the last row. Compared with the result of a state-of-the-art Fast Object Segmentation (FOS) method, the result demonstrates that our method achieves robust and improved segmentation accuracy of a highly dynamic image sequence, the AP improved by 48.15% and 96.52% pixels are correctly classified.

Summary

In this paper, we propose a foreground hypothesis framework for object segmentation in presence of highly dynamic scene. This framework enables foreground segmentation by ranking object hypotheses over spatial space to achieve a robust binary segmentation of a video sequence in an unsupervised manner. Object candidates derived from spatial features in each frame are used to rank the object candidates over a specific hypothesis space to yield object proposals with varying ranking scores. Based on the hypothesis with the highest ranking score, a Markov Random Field (MRF) model is use to segment foreground regions. We demonstrate the performance of our approach through experimental evaluation on a difficult dynamic scene segmentation benchmark from Freiburg-Berkeley Motion Segmentation Data-set, and show that our method achieves robust segmentation performance when dealing with a typical highly dynamic image sequence.

Acknowledgements

This work was supported by Project 61302173 and 61461022 of the National Science Foundation of China and Foundation of Kunming University of Science and Technology under Grant 14118777, KKZ3201401003.

References

[1] A. Papazoglou and V. Ferrari, Fast object segmentation in unconstrained video, IEEE International Conference on Computer Vision (ICCV), (2013), pp. 1777

[2] P. F. Felzenszwalb and D. P. Huttenlocher, Efficient graph-based image segmentation, International Journal of Computer Vision, vol. 59, no. 2, (2004).

[3] A. Radhakrishna, S.Appu, S. Kevin, L. Aurelien, F. Pascal and S. Susstrunk, Slic superpixels. Technical Report 149300 EPFL, (2010).

[4] P. Arbelaez, M. Maire, C. Fowlkes and J. Malik, Contour detection and hierarchical image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, no. 5, (2011) pp. 898

[5] http://lmb.informatik.uni-freiburg.de/resources/datasets/moseg.en.html

[6] D. Hoiem, Y. Chodpathumwan and Q. Dai, Diagnosing error in object detectors. ECCV, (2012).