

Geodesic pixel neighborhoods for multi-class image segmentation

Vladimir Haltakov¹

<http://campar.in.tum.de/Main/VladimirHaltakov>

Christian Unger¹

<http://campar.in.tum.de/Main/ChristianUnger>

Slobodan Ilic²

<http://campar.in.tum.de/Main/SlobodanIlic>

¹ BMW Group

Munich, Germany

² Siemens AG

Corporate Technology

Munich, Germany

Introduction

Multi-class image segmentation is a complex problem that poses several challenges: developing better classifiers, designing more discriminative features, finding efficient optimization techniques and modeling the relations between image pixels in different image regions. In this paper we focus on the last one. A common way to address the problem of structured prediction is to model it as a Conditional Random Field (CRF), but in this paper we take a different approach by using classification and integrating local and global semantic structure constraints directly in the features.

Our contribution is threefold. Firstly, we introduce a classification framework based on the concept of pixel neighborhoods, which captures structure constraints with a new histogram based neighborhood feature. Secondly, we propose a novel way to use the geodesic distance to compute the local pixel neighborhood. Thirdly, we introduce a new global rays based neighborhood, again using the geodesic distance, that can also capture global context.

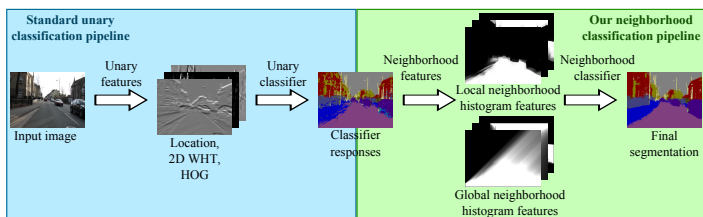


Figure 1: The blue part shows the standard unary classification process, while the green part shows the neighborhood classification pipeline.

Pixel neighborhoods

We introduce a classification framework based on the concept of pixel neighborhoods as visualized in Fig. 1. We define the neighborhood N_i of pixel i as a set of pixels that are related to the pixel i in some way. In the paper we explore several ways to define pixel neighborhoods by making use of the geodesic distance transform defined as:

$$d(i, j) = \inf_{G \in \mathcal{P}_{i,j}} \int_0^l \sqrt{1 + \gamma^2 (\nabla I \cdot \mathbf{G}'(\mathbf{s}))^2} ds, \quad (1)$$

where $\mathcal{P}_{i,j}$ is the set of all possible paths between pixels i and j , \mathbf{G} is a path from this set with length l and \mathbf{G}' is its spatial derivative. The parameter γ indicates the weight between the image gradient and the spatial distance between the two pixels. For $\gamma = 0$ the geodesic distance becomes equivalent to the euclidean distance, while for $\gamma = 1000$ it is dominated by the image gradients.



Figure 2: Visualization of the shapes of the presented neighborhoods for selected pixels (marked in black).

We introduce two types of neighborhoods: an adaptive local neighborhood and a rays based global neighborhood that are able to express local or global relations respectively (see Fig. 2). The local neighborhood

consists of the closest N pixels to the pixel of interest according to the geodesic distance. This allows the neighborhood to cover a patch around the pixel that aligns well to strong image gradients which often correspond to object edges. For the global neighborhood we shoot 8 rays at 45° from the pixel of interest to the borders of the image. For each ray we define a separate neighborhood, which again makes use of the geodesic transform to accumulate the pixels along the ray. In this way our global neighborhood is able to capture long range context relations.

Neighborhood classification framework

We first classify each image pixel individually based on features computed from the image. Then, for each pixel we compute one or more pixel neighborhoods and summarize the responses of the classifier over each neighborhood by computing a new histogram based feature. This is done by letting each pixel in the neighborhood vote for its most probable label based on the responses of the unary classifier. We then create a normalized histogram over all votes from the neighborhood and use the values of the histogram as features. This new feature is a compact representation of the whole neighborhood which allows for very fast training.

We use the neighborhood features to train a second classifier which is again used to classify each pixel, but in contrast to the first one, it integrates local and global constraints from the neighborhood features and is therefore able to improve the results of the first classifier significantly.

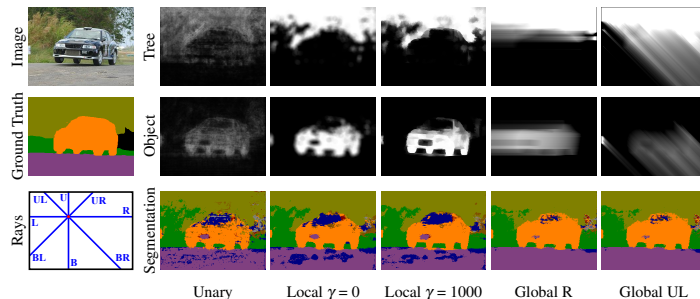


Figure 3: Histogram based neighborhood features. We show the bins corresponding to the classes TREE and OBJECT as a probability map, the segmentation from the neighborhood classifier for the local and global neighborhoods and the raw unary responses. The last two columns show two of the rays of the same global neighborhood. A high value for a pixel in the global neighborhood of a ray means that there is a region of this class in this direction.

Results

We evaluate our method on three widely used and very challenging datasets: CamVid, MSRC-21 and the Stanford background dataset. We analyze the performance of the different parts of our model and show how they contribute to increase the segmentation performance. Furthermore, we compare to two well known and strongly related methods: auto-context [2] and the robust P^n model of [1] and show an increase in performance especially around the object edges.

- [1] P. Kohli, L. Ladicky, and P. H. S. Torr. Robust higher order potentials for enforcing label consistency. In *IJCV*, 2009.
- [2] Zhuowen Tu and Xiang Bai. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. In *PAMI*, 2010.