Saliency Driven Total Variation Segmentation

Michael Donoser, Martin Urschler, Martin Hirzer and Horst Bischof
Institute for Computer Graphics and Vision
Graz University of Technology, Austria
{donoser,urschler,hirzer,bischof}@icg.tugraz.at

Abstract

This paper introduces an unsupervised color segmentation method. The underlying idea is to segment the input image several times, each time focussing on a different salient part of the image and to subsequently merge all obtained results into one composite segmentation. We identify salient parts of the image by applying affinity propagation clustering to efficiently calculated local color and texture models. Each salient region then serves as an independent initialization for a figure/ground segmentation. Segmentation is done by minimizing a convex energy functional based on weighted total variation leading to a global optimal solution. Each salient region provides an accurate figure/ground segmentation highlighting different parts of the image. These highly redundant results are combined into one composite segmentation by analyzing local segmentation certainty. Our formulation is quite general, and other salient region detection algorithms in combination with any semi-supervised figure/ground segmentation approach can be used. We demonstrate the high quality of our method on the well-known Berkeley segmentation database. Furthermore we show that our method can be used to provide good spatial support for recognition frameworks.

1. Introduction

Image segmentation is often described as partitioning an image into a set of non-overlapping regions covering the entire image. In general, one can distinguish unsupervised, semi-supervised and fully supervised methods.

Unsupervised approaches [37, 2] provide segmentation results without any prior knowledge about the image and do not require any user-interaction. One of the main directions of current research in this field is to define segmentation as finding the labeling of an image that minimizes a specific energy term. Two different approaches for finding the optimal labeling were recently popular. First, methods that define segmentation as a minimum cut or maximum flow problem through a graph as e. g. by [33] in their normalized cut framework. Second, variational approaches which evolve boundary contours as in the popular level set framework [25] or using weighted total variation [4]. Very recently also approaches for multi-label segmentation have been proposed in this context [28]. But despite the success of such energy minimization methods, still simple appearance based methods like Mean Shift [7] are considered as state-of-the-art in the field of color segmentation.

Semi-supervised methods [32, 1, 35] require a user to highlight some regions as a prior, mostly by drawing some kind of seeds into the image. These methods achieve impressively accurate results, but have the disadvantage that results heavily depend on the selection of the seeds. The correct placement of the seeds by the user needs some training and expertise, and therefore mostly cumbersome post-processing is required to correct the results.

Finally, fully supervised methods [34, 18] require labeled training data for the expected type of image, mostly for the purpose of detecting specific object categories in images. Of course, the need for accurately labeled training data limits the scope of these methods.

In this work we propose a fully unsupervised approach but borrow ideas from semi-automatic methods. The main idea of this paper is to exploit the efficiency and accuracy of semi-automatic variational figure/ground segmentations in combination with an automated seed extraction method in order to provide segmentation results in a fully unsupervised manner. We formulate segmentation as a repeated application of figure/ground segmentations steps on the same input image, each focussing on a different salient region in the image. Thereby we get accurate, highly redundant segmentations for the different salient parts of the input image. In a post-processing step, we combine all the sub-results into one composite image by making the maximum-likelihood decisions at each pixel analyzing color and texture cues. Such a combinatorial approach was recently also used for segmentation e. g. in [27, 24, 10] achieving state-of-the-art results on reference databases. However, these algorithms either suffer from long computation times or from the limited accuracy of the sub-segmentations and the

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

final combination step. We overcome these problems by proposing a fast salient region detection method and by using an accurate and efficient state-of-the-art figure/ground segmentation algorithm.

The outline of the paper is as follows. Section 2 introduces our unsupervised segmentation method in detail. Section 3 demonstrates the improved performance of the algorithm on images of the Berkeley segmentation database. Furthermore, we also show that our method is able to provide better spatial support for the task of recognition in contrast to state-of-the-art methods.

2. Unsupervised saliency driven segmentation

Our method consists of three steps. First, we automatically find salient regions in the image (Section 2.1), which are then used independent of each other as seed regions for a global optimal total variation segmentation method (Section 2.2). Finally, all segmentation results are merged to a final composite segmentation by analyzing local label certainty (Section 2.3).

2.1. Saliency initialization

The first step of our method is to automatically detect salient regions in the input color image which are then used as initialization for the subsequent figure/ground segmentation steps. In general any of the numerous region-of-interest (saliency) detection methods can be used, e. g. see a recent survey [17] which also provides a collaborative benchmark. We propose a simple and fast method, which as shown in the experimental evaluation achieves excellent results in combination with the figure/ground segmentation described in Section 2.2.

The underlying idea is to split the image into a set of non-overlapping blocks and to perform a simple but powerful clustering of features describing the local neighborhood of each pixel in each block. We use covariance matrices of low-level features as strong color and texture descriptor for local pixel neighborhoods as proposed by Porikli et al. [30]. The main advantage of using covariance matrices as local descriptors is that they enable efficient fusion of different types of low-level features and modalities and that in contrast to other descriptors their dimensionality is small.

We use a nine-dimensional feature vector f for constructing the covariance matrices which is defined as

$$\boldsymbol{f} = [x \;\; y \;\; L \;\; a \;\; b \;\; I_x \;\; I_y \;\; I_{xx} \;\; I_{yy}]\,, \tag{1}$$

where x and y are the normalized pixel coordinates, L, a and b are the pixel values of the Lab color space and I_x, I_y, I_{xx} and I_{yy} are the corresponding first and second-order derivatives of the image intensities. We define a fixed neighborhood size $N \times N$ for every pixel and calculate the symmetric 9×9 covariance matrix $\boldsymbol{\Sigma}$ by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{19} \\ \vdots & \ddots & \vdots \\ \sigma_{91} & \cdots & \sigma_{99} \end{pmatrix}, \tag{2}$$

$$\sigma_{ij} = \frac{1}{N^2 - 1} \sum_{n=1}^{N^2} (f_i^n - \mu_i)(f_j^n - \mu_j), \tag{3}$$

where μ_i is the mean value of the i-th feature f_i. Since covariance matrices are symmetric we get a 36-dimensional feature vector per pixel containing color and texture information of the local neighborhood. Please note, that as it was shown by [30] such local covariance matrices of pixel features can be calculated very efficiently by integral images.

Covariance matrices do not lie in an Euclidean space, but since they are positive semi-definite they can be represented in a connected Riemannian manifold, which is a topological space that is locally Euclidean. In this manifold space distances [12] can e. g. be measured by

$$\Delta\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right) = \sqrt{\sum_{d=1}^{D} \ln^2 \lambda_d\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right)}, \tag{4}$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the two input covariance matrices, λ_d are the generalized eigenvalues of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ and D is the dimensionality of the feature vector, i. e. nine in our case. This distance measure Δ fulfills all metric axioms (positivity, symmetry and the triangle inequality). This distance measure allows to compare the covariance matrices of the local pixel neighborhoods to each other.

For efficiency reason we split the image into non-overlapping equal-sized blocks. In each of these blocks we independently cluster the pixels based on our 36-dimensional feature space by applying affinity propagation clustering [15] to the affinity matrix Δ_{ij} containing all the similarities between the covariance matrices of all pixels within each block.

Affinity propagation was recently proposed by Dueck and Frey [15] and enables clustering of data points by analyzing a provided pairwise affinity matrix. It is based on iteratively exchanging messages between data points until a good solution emerges and is able to handle missing data points and non-metric similarity measures.

Affinity propagation provides a clustering of the image into several connected segments. The salient regions are then defined by further merging the obtained regions based on clustering of their mean feature vectors again by affinity propagation which further reduces the number of segments. The finally obtained clusters represent an estimate of the predominant color and texture distributions within the input

Figure 1: Salient region detection examples. Each detected region is shown in mean RGB color and the background is highlighted in blue.

image. These salient regions are eroded to avoid border effects, and the smaller regions are removed. Figure 1 shows examples of salient region detection results.

2.2. Total variation segmentation

The salient region detection step of Section 2.1 provides a set of regions where each is passed as an initialization to a figure/ground segmentation method. We exploit the efficiency and accuracy of a global optimal total variation segmentation approach to calculate an accurate segmentation for each of the identified salient regions. We use a hybrid algorithm which combines a geodesic active contour (GAC) as proposed by Caselles *et al*. [5] with a region model that incorporates color and texture information into an energy functional E. According to [5] we establish the energy minimization problem

$$\min_C\Big\{E\Big\} = \min_C\Big\{\underbrace{\int_0^{L_C} g(\nabla I(C(s)))\,ds}_{\text{GAC}} + \lambda \underbrace{\int_\Omega uf\,d\Omega}_{\text{Region}}\Big\}, \tag{5}$$

where C is the evolving contour, *i.e*. the boundary of the evolving region, L_C is its Euclidean length, g is an edge indicator function that vanishes at object boundaries of I, the image under consideration, u is a characteristic function, and f is a function that encompasses the region information. However, due to its non-convexity we replace the GAC-term in this equation with the weighted total variation TV_g minimization problem

$$\min_u\Big\{E_{TV_g}\Big\} = \min_u\Big\{\underbrace{\int_\Omega g\,|\nabla u|\;d\Omega}_{\text{TV}_\text{g}} + \lambda \underbrace{\int_\Omega uf\,d\Omega}_{\text{Region}}\Big\}, \tag{6}$$

which was introduced by Bresson *et al*. in [4]. They proved that if u is a characteristic function 1_{Ω_C} of a set Ω_C whose boundary is denoted C, and u is allowed to vary continuously between $[0, 1]$, (5) and (6) describe the same energy. The advantage of the latter formulation over the previous one is its (non-strict) convexity, making it possible to derive a globally optimal solution for u. By selecting an arbitrary levelset of u a binary segmentation is obtained. For the weighting g we use $g(\nabla I) = e^{-\eta|\nabla I|^\kappa}$, an edge measure that is optimized for natural images [16]. With the parameter λ the influence of the TV_g-term and the region-term on the segmentation result can be controlled. A small λ means that the result is primarily determined by the TV_g term, whereas a large λ makes the region-term to the main contributor. We define the function f as a probability ratio that forces the segmentation to partition the image into homogeneous regions. The involved probabilities encompass color and texture information by

$$\min_u\Big\{E_{TV_g}\Big\} = \min_u\Big\{\underbrace{\int_\Omega g\,|\nabla u|\;d\Omega}_{\text{TV}_\text{g}} + \lambda \underbrace{\int_\Omega u \log\frac{p_{bg}}{p_{fg}}\,d\Omega}_{\text{Region}}\Big\}, \tag{7}$$

where the probabilities p_{fg} and p_{bg} represent the foreground region, i.e. the current segmentation region under investigation, and the background region. To obtain these probabilities we use Gaussian models in a feature space consisting of color (*Lab* space) and texture (gradient) features. Note that covariance matrices as described in Section 2.1 cannot be directly used as features since they do not form a Euclidean space. The model parameters are derived from all pixels that are currently inside the region. Having built the Gaussian model for a region then enables us to calculate the probability that a pixel belongs to the foreground, respectively background.

Solving total variation models is a demanding task due to the non-differentiability of the L_1-norm in the TV-term at zero. Many different approaches exist, ranging from explicit time marching algorithms to graph cut methods. Based on Aujol *et al*. [3] we modify our model using an approximation variable v

$$\begin{aligned}\min_{u,v}\Big\{E_{TV_g\,Region}\Big\} = \min_{u,v}\Big\{&\int_\Omega g\Big|\nabla u\Big|\,d\Omega + \\ &\frac{1}{2\theta}\int_\Omega (u-v)^2\,d\Omega + \lambda\int_\Omega v\log\frac{p_{bg}}{p_{fg}}\,d\Omega\Big\}.\end{aligned} \tag{8}$$

Introducing v leads to a third term in the TV_g-region model, the connection term $\frac{1}{2\theta}\int_\Omega (u-v)^2\,d\Omega$. For $\theta > 0$

Equation (8) is a strictly convex approximation of (7), with θ controlling how well v has to approximate u. It is chosen small to enforce u and v being similar. The minimization task is now split into two steps. First, the energy is minimized in terms of u with v being fixed, and then the energy is minimized in terms of v with u being fixed. These two steps are iterated until convergence:

$$1. \quad \min_u \Big\{ \int_\Omega g\,|\nabla u|\,d\Omega + \frac{1}{2\theta}\int_\Omega (u-v)^2\,d\Omega \Big\} \tag{9}$$

$$2. \quad \min_v \Big\{ \frac{1}{2\theta}\int_\Omega (u-v)^2\,d\Omega + \lambda \int_\Omega v \log\frac{p_{bg}}{p_{fg}}\,d\Omega \Big\} \tag{10}$$

According to Chambolle *et al*. [6] Equation (9) can be solved by using a dual variable $\boldsymbol{p} = \frac{\nabla u}{|\nabla u|}$ under the assumption that v is constant:

$$\begin{aligned} u^{n+1} &= v + \theta \operatorname{div} \boldsymbol{p} \\ \boldsymbol{p}^{n+1} &= \frac{\boldsymbol{p}^n + \frac{\tau}{\theta}\nabla u}{1 + \frac{\tau}{\theta}\frac{|\nabla u|}{g}} \end{aligned} \tag{11}$$

In practice the timestep τ has to be less or equal than $\frac{1}{4}$ in order to achieve convergence. To solve (10) under the assumption that u is constant we obtain:

$$\begin{aligned} &\frac{1}{\theta}(v-u) + \lambda \log\frac{p_{bg}}{p_{fg}} \stackrel{!}{=} 0 \\ &\Rightarrow v = u - \lambda\theta \log\frac{p_{bg}}{p_{fg}}, \qquad v = \max(0, \min(1, v)) \end{aligned} \tag{12}$$

After every iteration of the region evolving process the probabilities p_{fg} and p_{bg} are updated according to the actual segmentation. However, for efficiency reasons new probability values are only calculated after significant changes in the region size. The final result of this procedure is a figure/ground segmentation u.

2.3. Segmentation fusion

The method described in Section 2.2 returns a figure/ground segmentation for every provided salient region. Since these results may overlap we make a region based maximum likelihood decision in overlapping areas. We use the finally estimated probability maps from the figure/ground segmentation as a local measure for the segmentation quality. The lower the probability ratio the more similar the local neighborhood is to the background. To increase robustness we again analyze local pixel neighborhoods of size $N \times N$.

All overlapping segmentation areas are analyzed and we locally assign the label with the highest mean probability value in the overlap area which provides a first composite segmentation result. Since some regions will still be unassigned we additionally have to perform some post-processing. If we have a large unassigned region we can assume that this area covers a missed salient region in the image. In such a case, we simply do another figure/ground segmentation with this region as initialization. All remaining small unassigned regions are added to the neighbor with the most similar feature covariance matrix (see Equation 4). Thus, in the end every pixel of the image is assigned a unique ID representing the final segmentation result.

3. Experiments

We implemented the proposed method in Matlab, where a single run of the total variation method requires about a second. The fully unsupervised segmentation takes a few seconds, depending on the number of salient regions. But please note, that our approach can be easily parallelized. Furthermore, as e. g. shown by Pock *et al*. [29] it is possible to implement total variation methods on modern GPUs, which significantly reduces the runtime of the method.

In all experiments we used the same default parameters for our method. The local window size was set to $N = 15$ and the parameters of the total variation figure/ground segmentation were fixed to $\lambda = 0.5$, $\theta = 0.01$, $\eta = 0.1$ and $\kappa = 0.55$.

In Section 3.1, we first demonstrate the quality of our proposed segmentation method on the well-known Berkeley image database in comparison to several state-of-the-art methods. We further show in Section 3.2 that our method also provides good spatial support for recognition frameworks outperforming several other methods.

3.1. Berkeley segmentation database

We first benchmark our proposed algorithm on the well-known Berkeley image database [21]. Berkeley provides 300 images and corresponding ground truth data (at least 4 human segmentations per image). For evaluation of the overall segmentation quality we follow a recent trend [19] of using four different quality measures: the Probabilistic Rand Index (PRI) [36], the Variation of Information (VoI) [23], the Global Consistency Error (GCE) [22] and the Boundary Displacement Error (BDE) [14].

We compare the scores calculated for our algorithm to ten state-of-the-art color segmentation algorithms (in chronological order): the Mean Shift method (*Mshift*) [7], the standard normalized cut algorithm (*Ncuts*) [33], the JSEG algorithm (*Jseg*) [9], the pixel affinity based method (*Affin*) [13], the spectral clustering method (*SpectClust*) [39], the graph based segmentation (*GraphBased*) [11], the multi-scale normalized cut approach (*Mscuts*) [8], the seeded graph cuts method (*Seed*) [24], the MSER-based segmentation method (*ROI-Seg*) [10] and the

Method / Score	PRI	VoI	GCE	BDE
Mshift [7]	**0.7958**	**1.9725**	0.1888	**14.41**
Ncuts [33]	0.7242	2.9061	0.2232	17.15
Jseg [9]	0.7756	2.3217	0.1989	**14.40**
Affin [13]	X	X	0.2140	X
SpectClust [39]	0.7357	2.6336	0.2469	15.40
GraphBased [11]	0.7139	3.3949	**0.1746**	16.67
Mscuts [8]	0.7559	2.4701	0.1925	15.10
Seed [24]	X	X	0.2090	X
ROI-Seg [10]	0.7599	2.0072	0.1846	22.45
NormTree [37]	0.7521	2.4954	0.2373	16.30
Our method	**0.7758**	**1.8165**	**0.1768**	16.24

Table 1: Comparison of different methods (in chronological order) on Berkeley image database by Probabilistic Rand Index (PRI), Variation of Information (VoI), Global Consistency Error (GCE) and Boundary Displacement Error (BDE). The best two results are always highlighted in bold. Our method shows competitive results compared to state-of-the-art, e. g. achieving the lowest Variation of Information score of all methods.

normalized partitioning tree (*NormTree*) [37]. Results were calculated with publicly available implementations or taken from [24], [10] and [37]. Table 1 summarizes the scores for all algorithms. As can be seen our proposed algorithm shows competitive results compared to state-of-the-art. We rank in the top two for three different measures and e. g. achieve the best Variation of Information score of all.

3.2. Spatial support For recognition

The second part of our experimental evaluation demonstrates the high potential of our method to provide good spatial support for recognition methods. Recently many authors confirmed that using unsupervised segmentation results allows to substantially improve recognition performance [31, 20, 26], especially for objects that cannot be well approximated by a rectangle (as it is common in the sliding window approaches). These papers all agree that in general unsupervised segmentation provides the necessary spatial support, but also clarify that none of the state-of-the-art methods is able to provide an optimal result for all different kinds of images.

Our method finds segmentations that visually are quite similar to the outlines that humans identify (see Figure 4). This makes our algorithm perfectly suited as a basis for different recognition methods. To demonstrate the improved performance of our method in contrast to several other state-of-the-art unsupervised segmentation methods for the task of recognition, we follow an experimental setup proposed by Malisiewicz and Efros [20].

They compared different unsupervised segmentation algorithms concerning their capability of providing a good level of spatial support for the task of semantic segmentation of images. They estimated a spatial support score defined as follows. Given a ground truth labeling of an image into a set of N non-overlapping regions G_n and the corresponding segmentation by a query algorithm into a set of M overlapping regions Q_m, an overlap score OS is calculated between every ground truth and every query region by

$$OS\left(G_n, Q_m\right) = \frac{G_n \cap Q_m}{G_n \cup Q_m}. \tag{13}$$

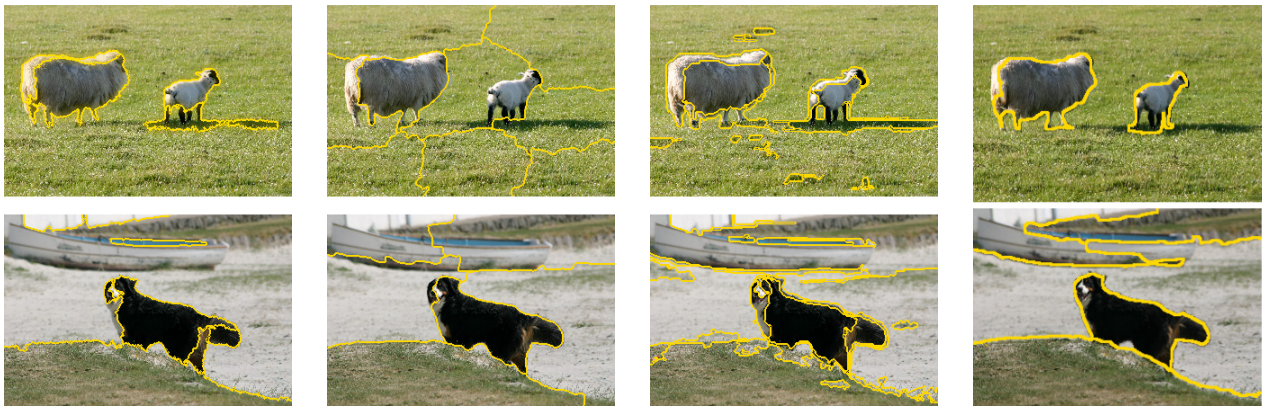
Then for every ground truth region G_n the Best Spatial Support (BSS) score is calculated by

$$BSS\left(G_n\right) = \max_{Q_m}\left(OS\left(G_n, Q_m\right)\right), \tag{14}$$

i. e. the query segmentation region with the best spatial support provides its overlap score as the BSS for the ground truth region. In such a way an overall spatial support score is calculated by taking the mean BSS over all ground truth regions in all test images.

In [20] Normalized Cut [33], Felzenszwalb's algorithm [11] and Mean Shift [7] were compared. It was shown that Mean Shift provides the best single segmentation results. Therefore, we only compare to Mean Shift, the results for the other two methods can be found in [20]. We follow the same experimental setup and use the 21 class database of MSRC to calculate an overall spatial support score for our algorithm. MSRC is one of the few databases providing a pixel-wise segmentation and labeling for all images. For this reason, we segmented the 591 color images using our algorithm with default parameters to provide a single labeling result for every image. Figure 2 shows example segmentations for the MSRC database in comparison to reference methods. Same images as in [20] are given.

Figure 3 shows our class specific overall spatial support scores in comparison to three other approaches: the results of the Mean Shift algorithm, the best possible rectangle simulating a bounding box approach and a superpixel limit. The superpixel limit was calculated by segmenting the image in small sub-regions and choosing the best combination of several superpixels as a limit for the spatial support score. Results were taken from [20]. As can be seen we outperform Mean Shift in several classes like sky, water and flowers significantly and we get a 5% improvement for the overall spatial support score. Please note further, that the results for Mean Shift are achieved by segmenting the images several times with different parametrization and that finally from this large set of segmentations the best one is chosen to provide the BSS score. In contrast, we applied our algorithm with default parameters only once (!) per image and achieved these improved results, which is a much stricter experimental setup.



(a) Mean Shift [7] (b) Normalized Cuts [33] (c) Felzsenswalb [11] (d) **Our method**

Figure 2: Segmentation results for MSRC semantic segmentation database. Images of reference methods are from [38].

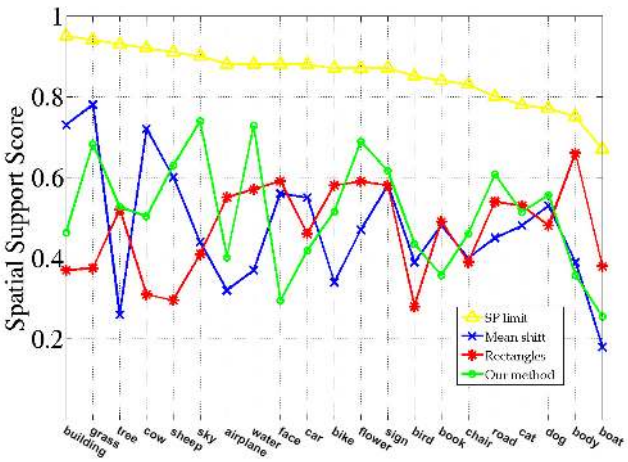


Figure 3: Spatial Support Scores [20] calculated on the 21 class MSRC database. Our method provides an on average 5% better score with significant better performance for classes like sky, water or flowers. In addition, our results were achieved with a single run with default parametrization (!), whereas for Mean Shift images were segmented several times with different parameters and the best segmentation per image was chosen for calculating the score.

Figure 4 gives a better insight into segmentation results achieved by our method. We show selected images for each of the 20 image categories of the MSRC database. For every category the two images with the highest spatial support score and the image with the worst score are shown.

To sum up, results prove the applicability of our method in recognition frameworks. Please note, that as was shown by [31, 20, 26] the best recognition results are mostly achieved by combining several differently parameterized algorithms, which may be the focus of future work.

4. Conclusion

In this paper we introduced a fully unsupervised segmentation method, which is based on the idea of combining several figure/ground segmentations (each focussing on a different salient part of the image) into one composite segmentation result. We described a global optimal total variation framework that allows to provide accurate and efficient figure/ground segmentations and showed how the required initializations are provided by a novel salient region detection process. Experimental results showed that competitive results are obtained on the Berkeley image database and that results can be used to provide good spatial support in recognition methods.

References

[1] P. Arbelaez and L. Cohen. Constrained image segmentation from hierarchical boundaries. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[3] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition–modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.

[4] X. Bresson, S. Esedoglu, P. Vandergheynst, J. P. Thiran, and S. J. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.

[5] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79, February 1997.

[6] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

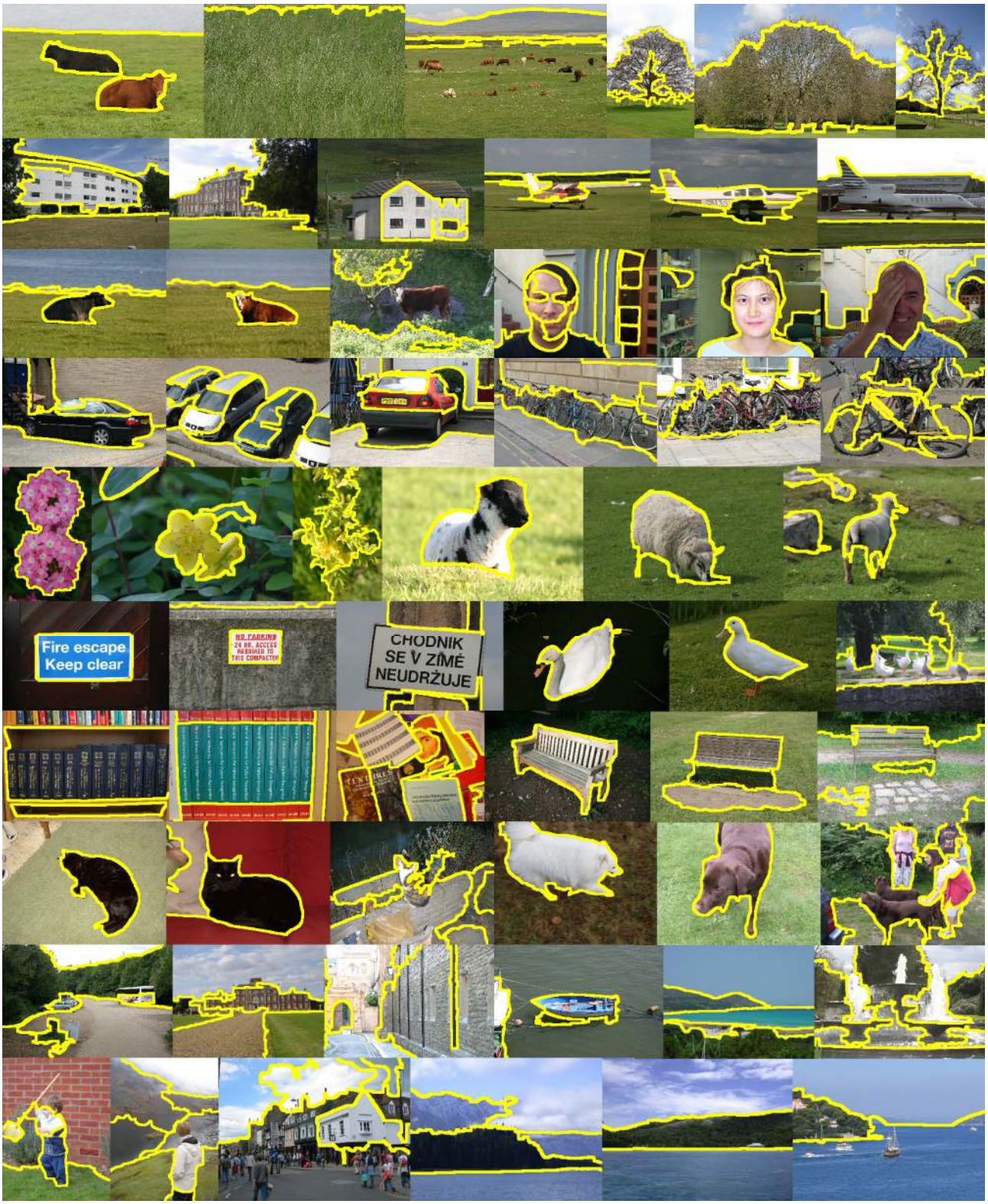


Figure 4: Segmentation results of our method with default parameters per image for MSRC database. For each of the 20 image categories, the two images with the best spatial score and the image with the worst score are shown (two categories per line).

[7] D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 750–755, 1997.

[8] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[9] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *Trans. on Pattern Analysis and Machine Intelligence*, 23(8):800–810, 2001.

[10] M. Donoser and H. Bischof. ROI-SEG: Unsupervised color segmentation by combining differently focused sub results. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[11] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[12] W. Foerstner and B. Moonen. A metric for covariance matrices. Technical report, Department of Geodesy and Geoinformatics, Stuttgart University, 1999.

[13] C. Fowlkes, D. Martin, and J. Malik. Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 54–61, 2003.

[14] J. Freixenet, X. Munoz, D. Raba, J. Martí, and X. Cufí. Yet Another Survey on Image Segmentation: Region and Boundary Information Integration. In *European Conf. on Computer Vision (ECCV)*, pages 408–422, 2002.

[15] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315:972–976, 2007.

[16] J. Huang and D. Mumford. Statistics of natural images and models. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 541–547, 1999.

[17] T.-H. Huang, K.-Y. Cheng, and Y.-Y. Chuang. A collaborative benchmark for region of interest detection algorithms. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 296–303, June 2009.

[18] J. Jiang and Z. Tu. Efficient scale space auto-context for image segmentation and labeling. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[19] Y. Ma, H. Derksen, W. Hong, and J. Wright. Segmentation of multivariate mixed data via lossy data coding and compression. *Trans. on Pattern Analysis and Machine Intelligence*, 29(9):1546–1562, Sept. 2007.

[20] T. Malisiewicz and A. A. Efros. Improving spatial support for objects via multiple segmentations. In *British Machine Vision Conf. (BMVC)*, September 2007.

[21] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *International Conf. on Computer Vision (ICCV)*, pages 416–423, 2001.

[22] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *Trans. on Pattern Analysis and Machine Intelligence*, 26(5):530–549, 2004.

[23] M. Meila. Comparing clusterings by the variation of information. In *Journal of Multivariate Analysis*, pages 173–187. 2003.

[24] B. Micusik and A. Hanbury. Automatic image segmentation by positioning a seed. In *European Conf. on Computer Vision (ECCV)*, volume 2, pages 468–480, 2006.

[25] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging,Vision,and Graphics*. Springer New York, Inc., 2003.

[26] C. Pantofaru, C. Schmid, and M. Hebert. Object recognition by integrating multiple image segmentations. In *European Conf. on Computer Vision (ECCV)*, pages 481–494, 2008.

[27] J. C. Pichel, D. E. Singh, and F. F. Rivera. Image segmentation based on merging of sub-optimal segmentations. *Pattern Recognition Letters*, 27(10):1105–1116, 2006.

[28] T. Pock, A. Chambolle, D. Cremers, and H. Bischof. A convex relaxation approach for computing minimal partitions. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[29] T. Pock, M. Unger, D. Cremers, and H. Bischof. Fast and exact solution of total variation models on the gpu. In *CVPR Workshop on Visual Computer Vision on GPUs*, 2008.

[30] F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on lie algebra. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[31] B. C. Russell, A. A. Efros, J. Sivic, W. T. Freeman, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[32] J. Santner, M. Unger, T. Pock, C. Leistner, A. Saffari, and H. Bischof. Interactive texture segmentation. In *British Machine Vision Conf. (BMVC)*, 2009.

[33] J. Shi and J. Malik. Normalized cuts and image segmentation. *Trans. on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[34] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[35] A. Sinop and L. Grady. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *International Conf. on Computer Vision (ICCV)*, 2007.

[36] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *Trans. on Pattern Analysis and Machine Intelligence*, 29(6):929–944, 2007.

[37] J. Wang, Y. Jia, X.-S. Hua, C. Zhang, and L. Quan. Normalized tree partitioning for image segmentation. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[38] A. Yang, J. Wright, Y. Ma, and S. S. Sastry. Unsupervised segmentation of natural images via lossy data compression. *Computer Vision and Image Understanding*, 2007.

[39] S. Yu and J. Shi. Multiclass spectral clustering. In *International Conf. on Computer Vision (ICCV)*, pages 313–319, 2003.