ELEKTROTEHNIŠKI VESTNIK 83(3): 138–143, 2016
ORIGINAL SCIENTIFIC PAPER

A regularization-based approach for unsupervised image segmentation

Aleksandar Dimitriev[1,†], **Matej Kristan**[1,2]

[1] *Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia*
[2] *Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia*

[†] *E-mail: ad7414@student.uni-lj.si*

Abstract. We propose a novel unsupervised image-segmentation algorithm aiming at segmenting an image into several coherent parts. It requires neither user input nor supervised learning phase and assumes an unknown number of segments. It achieves this by over-segmenting the image into several hundred superpixels iteratively joined on the basis of a discriminative classifier trained on color and texture information obtained from each superpixel. The output of the classifier is regularized by a Markov random field that lends more influence to the neighbouring superpixels that are more similar. In each iteration, the similar superpixels fall under the same label, until only a few coherent regions remain in the image. The algorithm is tested on a standard evaluation data set, where it performs on par with the state-of-the-art algorithms in terms of precision and greatly outperforms the them by reducing the oversegmentation of the object of interest.

Keywords: image segmentation, Markov random field, computer vision

Regularizirana nenadzorovana segmentacija slik

Predlagamo nov algoritem za nenadzorovano segmentacijo slik na več vizualno skladnih delov. Novi algoritem ne zahteva inicializacije s strani uporabnika, ne potrebuje nadzorovane učne faze in ne predpostavlja znanega števila segmentov. To doseže tako, da najprej sliko pre-segmentira na nekaj sto superpikslov, nato pa jih iterativno združuje s sprotnim nenadzorovanim učenjem diskriminativnega modela, ki temelji na barvi in teksturi. Rezultat diskriminacije je regulariziran z Markovim slučajnim poljem, moč regularizacije pa se avtomatsko določa preko vizualne podobnosti sosednjih superpikslov. Algoritem smo analizirali na standardni podatkovni zbirki in ugotovili, da je po preciznosti primerljiv z najsodobnješimi algoritmi za segmentacijo in jih močno prekaša v konsistentnosti segmentacije, saj bistveno manjkrat objekt razgradi v več regij.

1 INTRODUCTION

Image segmentation is a popular branch of computer vision, used to partition an image into a number of non-overlapping different segments or labels that correspond to some meaningful regions of the image. Depending on the size and number of the resulting regions, the final result can be called a segmentation or an oversegmentation into superpixels.

Superpixels are usually very homogeneous in color and texture and respect strong edges.

Received 23 December 2015
Accepted 10 March 2016

Broadly speaking, the image is said to be comprised of superpixels if the number of segments is in the hundreds, though there is no precise definition. Most methods can produce both coarse and fine segmentations, which can sometimes be controlled by adjusting the method parameters. We use the words segment, region and label interchangeably, since segments are defined as all the pixels belonging to the same label.

Figure 1. From left to right: input image, an example of the MRF pairwise consistency encoded by color similarity, and final segmentation.

Our focus in this paper is on non-semantic unsupervised segmentation with an unknown and variable number of segments, i.e. we make no assumptions about the meaning of the regions produced, we do not require any user input and the number of final regions is determined by the output of our algorithm. Therefore, even though a large part of image segmentation concerns supervised segmentation, foreground-background delineation and region labelling, all the work indicated here concerns unsupervised and object-agnostic segmentation unless

otherwise noted.

Several approaches to image segmentation have been developed. Among the first are graph-partitioning methods such as Shi and Malik's normalized cuts (NCC) [1] and Felzenszwalb and Huttenlocher's graph cut in nearest-neighbour graphs (FH) [2], which are frequently used as a preprocessing step in more sophisticated algorithms. Also well-known is the Grabcut [3] method, though it requires user input and only outputs a foreground/background segmentation. Another approach is mode-seeking algorithms such as Mean Shift (MS) [4] or Quick Shift [5] used in a color space such as RGB or Lab.

Arbelaez et al. [6] tackle the problem of image segmentation through contour detection, but the contours are not always valid segmentations because they are not necessarily closed. Other unsupervised methods start with a mixture of a large number of Gaussians [7] and gradually reduce this number by removing degenerated Gaussians and merging those that are closer in the feature space than some specified threshold.

Popular methods also include agglomerative clustering, i.e. a bottom-up aggregation of either the image pixels [8], [9], [10] or superpixels [11], [12]. However, these methods are greedy and suffer from error propagation, where incorrectly merged regions are propagated into subsequent iterations, although recent work by Wang et al. [13] tries to alleviate such errors by using multiple merge steps.

Although many methods do not use the spatial structure of the image, ignoring this information can sometimes lead to non-smooth segmentations. To combat this, many segmentation methods [14], [15], [16], [17] use MRFs as a way to enforce spatial consistency in the neighboring pixels in the image. However, these methods work on the pixel level, which is becoming increasingly more computationally intensive due to the rise of high-resolution images. Another benefit of this approach is that by using a superpixel segmentation algorithm that can output a fixed number of superpixels, all images require roughly the same computational effort regardless of their underlying pixel resolution.

Therefore, in more recent work, MRFs have been used on the superpixel level instead as a way to reduce the computational cost and achieve a speed-up [18]. Similarly, Fulkerson et al. [19] impose an MRF on a preliminary oversegmentation from QS [5]. In contrast to our work, [5] require offline pre-training and do not perform a segmentation but rather a region proposal generation for object detection. Some methods, such as Li et al. [20] and Wang et al. [21] use graph partitioning on the superpixel level, employing methods such as MS [4] and FH [2] as a preliminary step. However, both methods require the number of the final segments to be known.

Our contributions and approach: Our work is influenced by several of the aforementioned approaches. We adopt the preprocessing step of oversegmenting the image into several hundred superpixels as in [20], [21], [19]. There are many methods able to provide superpixel initialization such as SLIC [22], MS [4] and FH [2]. We find that SLIC works best for our purposes. From each of these superpixels we extract color and texture features using COLOR-CHILD [23], though any other descriptor could be used instead.

The main part of our algorithm are the subsequent iterations. In each iteration we train a discriminative classifier in a one-vs-all fashion, i.e. for each label a binary classifier is trained to distinguish between the superpixels belonging to that label (positive examples) and all the others (negative examples). Similarly to [19] we find that for our purposes Support Vector Machines (SVM) [24] work sufficiently well. We use these classifiers to re-classify all the superpixels to obtain a probability vector of labels for each superpixel. At the end of each iteration, we assign each superpixel to the highest probability in the label vector.

Although there are as many labels as there are superpixels in the beginning, which increases the computational cost of each training phase, the number of labels quickly declines in subsequent iterations. This is because many labels, especially those at the beginning when each label has a small number of superpixels, happen not to output the highest probability for any vector and are thus automatically removed from the label pool in all subsequent iterations.

Lastly, as in [17], [19], we adopt the idea of enforcing spatial consistency of the segmentation using an MRF. Before assigning the superpixels to their new labels, we perform regularization by penalizing neighbouring superpixels that are very similar, but have different labels. See Figure 1 for an example of superpixels, the pairwise potentials in the Markov random field, and the final regularized segmentation. A more detailed description of each step is presented followed by quantitative results and comparison in Section 3.

2 METHODS

The task of segmenting an image can be formulated as assigning a label l_i to each pixel. The number of labels K, however, is unknown a priori. An iterative approach can therefore be applied that starts with labelling every pixel with its own label and then gradually reduces the number of labels. Our approach is a two-stage approach composed of a pre-segmentation stage and followed by an iterative segmentation stage described below.

2.1 Pre-segmentation

Many regions of the image are visually similar and will likely have been assigned the same label in the final

segmentation. This means that such neighbouring pixels can be grouped, thus pre-segmenting the image. The image is over-segmented into a few hundred coherent groups of pixels previously defined as superpixels. This can be thought of as jump-starting the iterative merging of regions, since merging the pixels in the beginning with a hundred superpixels instead of a million pixels reduces the computational cost of any graph-based method that operate on the (super) pixel level, at the expense of having slightly less-refined boundaries. The algorithm used in this preliminary step is called Simple Linear Iterative Clustering (SLIC) [22]. It is a simple sped-up version of k-means that clusters the pixels simultaneously in a five-dimensional CIELAB and coordinate space (L, a, b, x, y). Since our method is agnostic of the choice of the superpixel pre-segmentation algorithm, we also experiment with MS [4] and FH [2] as a pre-segmentation, and present quantitative differences in Section 3.

2.2 Super-pixel description and classification

After oversegmenting the image, the next step is to use a descriptor to obtain discriminative features for each superpixel. The feature descriptor used in our algorithm is COLOR moments augmented Cumulative Histogram-based Image Local Descriptor (COLOR-CHILD) [23]. The color part contains the first, second and third image moments of all three color channels, whereas the texture part includes the information obtained from the first and second-order derivatives. The color and texture features together comprise the D = 57-dimensional descriptor (9 color dimensions and 48-bin quantized histogram) $f_i \in \mathbb{R}^{57\times 1}$ for the i^{th} superpixel.

These descriptors can readily be used to learn a classifier for each label. For example, assume we have M superpixels labelled by K labels. A classifier such as a one-versus-all support vector machine (SVM) [24] can be learned for each class from the features extracted from the superpixels labelled by the same label, resulting in K SVMs. The input to the SVM for label k is thus:

$$X = [f_1, \dots, f_M]^T, \quad y = [\delta_{l_1,k}, \dots, \delta_{l_M,k}]^T, \qquad (1)$$

where δ_{ij} is the Kronecker delta function, which is 1 if $i = j$ and 0 otherwise. In other words, all the superpixels that are currently labelled as k are the positive examples and all the rest are the negative examples, corresponding to the classic one-vs-all fashion for training multi-class SVMs.

Each SVM is calibrated by a Platt calibration such that it outputs a probability of observing label l_i, given its feature descriptor f_i. These classifiers are then applied back to each superpixel so that for the next iteration each superpixel is assigned a class label of the SVM with the maximum probability:

$$l_i = \operatorname*{argmax}_j p(l_j|f_i)$$

where $j \in \{1, ..., K\}$. In this way, the superpixels are relabelled. But an independent classification of the pixels will likely result in a noisy segmentation and some kind of regularization should be enforced to penalize the neighbouring superpixels that do not belong to the same class.

2.3 Regularization of segmentation

To enforce regularization, we apply a Markov Random Field (MRF) on the superpixels. MRFs are first-order graphical models commonly used as a way to encode spatial dependencies present between neighbouring pixels in an image. They have found applications in image restoration, stereo vision, and segmentation [14], [16], [17], [19]. Our approach uses MRFs to take advantage of the structural information present in an image, that would otherwise be unused. Each superpixel is a variable, with dependencies between superpixels that share a boundary.

The particular type of MRF applied here and the corresponding procedure of energy minimization are described in Kristan et al. [17]. Let λ_{ij} be the similarity of superpixels i and j defined as

$$\lambda_{ij} = \frac{1}{Z_i} exp(-||\mathbf{C}_i - \mathbf{C}_j||^2)$$

where $\mathbf{C}_i = [r_i, g_i, b_i]^T$ denotes the intensity of the RGB values of the i-th superpixel, $||.||$ denotes the usual l_2 norm, and Z_i is a normalization constant ensuring the weights sum to 1, i.e., $\sum_{j\in N_i} \lambda_{ij} = 1$, where N_i is the neighbourhood of the i-th superpixel, Briefly, the energy function corresponding to the MRF is the following:

$$E \propto \sum_{i=1}^{M} \log p(l_i|f_i) - \frac{1}{2}(D_{KL}(\pi_i, \pi_{N_i}) + D_{KL}(\mathbf{p}_i, \mathbf{p}_{N_i})), \qquad (2)$$

where π_i denotes i's prior probability distribution over the labels, π_{N_i} is a weighted sum of the priors of i's neighbours $\pi_{N_i} = \sum_{j\in N_i, j\neq i} \lambda_{ij}\pi_j$, and $D_{KL}(p, q) = \sum_x p(x) log\left(\frac{p(x)}{q(x)}\right)$ is the Kullback-Leibler divergence. The variables $\mathbf{p}_i$ and $\mathbf{p}_{N_i}$ are the corresponding posteriors and the neighbourhood averages, similarly to the priors. This particular formulation of the MRF [17] treats the priors as well as posteriors as random variables and enforces an MRF on the prior and an MRF on the posterior.

Given the visual likelihoods for all superpixels estimated by the SVMs, i.e., $p(l_i|f_i)$, the posteriors $\mathbf{p}_i$ are computed over all superpixels by minimizing the energy function from (2) by the iterative approach from [17].

2.4 Iterative re-labeling and class reduction

Once the posteriors over the superpixels are computed, each superpixel is assigned the label with the maximum probability. The classes that do not receive any superpixels in a given iteration are removed from

the candidate classes. Then the remaining SVMs are relearned from the relabelled superpixels. The process of superpixel re-classification, unsupported class removal and SVM-relearning is repeated until convergence. The iterative procedure is summarized in Algorithm 1.

Algorithm 1 Unsupervised MRF-based segmentation

```
Input: Image I
(f_i, l_i) ← color-child(spix(I)) ∀i
while ∃i : l_i ≠ l_{i−1} do
    for each label l do
        train a SVM to compute p(l_i|f_i)
        minimize the energy function (2)
    for each superpixel i do
        //Assign each i-th superpixel to the MAP estimate of the label
        l_i = argmax_j p(l_j|f_i)
return labeled image I using final labels l
```

3 Results

We have used the following parameters in our evaluation: we used an RBF kernel for the SVM with $\gamma = 0.001$ and regularization constant $C = 1.0$. These parameters were kept fixed in all experiments.

Our algorithm was evaluated on a standard data set [8] consisting of 100 color images which contain a single object of interest that usually occupies a majority of the image (Figure 2). The ground truth consists of three manual foreground-background segmentations provided by three human annotators. The task is to correctly infer the foreground region from the background pixels. The performance is measured by the F measure:

$$F = \frac{2PR}{P+R}, \tag{3}$$

where P and R denote the precision and recall which measure the fraction of the segment that contains the foreground, and the fraction of the foreground contained by the segment, respectively. In addition to computing the F measure for each segment and reporting the best value as F_{single}, we also compute it for each combination of segments and report the highest value as F_{multi}:

$$F_{single} = \max\{F_s\} \;\; s \in S \tag{4}$$

$$F_{multi} = \max\{F_s\} \;\; s \in 2^{|S|} \tag{5}$$

where S is the set of segments comprising the final segmentation, $2^{|X|}$ denotes the power set of X, and F_x denotes the F-measure of a segment x.

Finally, we assess the fragmentation of each method by counting the number of segments comprising the combined F-measure as follows:

$$F_{\text{frag}} = N - 1, \tag{6}$$

where N is the number of segments. Lower fragmentation, ideally zero, means that the object is represented by a single segment, whereas high F_{frag} implies oversegmentation.

To analyze the results of our method, we compare it to a number of state-of-the-art segmentation algorithms:

- Probabilistic Bottom-Up Aggregation and Cue Integration [8], denoted by PBACI, which gradually merges the pixels into successively larger regions by taking into account the intensity, geometry and texture.
- Segmentation by weighted aggregation [25], denoted by SWA, which determines the salient regions in the image and merges them into a hierarchical structure.
- Normalized cuts [26], denoted by N-cuts, which tackles the problem of segmentation by computing multiple minimum-normalized cuts on a pixel graph.
- Contour detection and hierarchical Image Segmentation [6], denoted by Gpb, which reduces the problem to contour detection and uses spectral clustering to combine local cues into a global framework.
- Mean shift [4], denoted by MS, which is a general mode-seeking algorithm on a non-parametric probability distribution, such as the color or intensity distribution.

Method	F_{single}	F_{multi}	F_{frag}
Our$_{MS}$	0.69 ± 0.01	$\mathbf{0.87} \pm 0.01$	0.45 ± 0.03
Our$_{FH}$	0.71 ± 0.01	0.85 ± 0.01	$\mathbf{0.43} \pm 0.03$
Our$_{SLIC}$	0.72 ± 0.01	0.84 ± 0.01	$\mathbf{0.40} \pm 0.03$
PBACI	$\mathbf{0.86} \pm 0.01$	0.87 ± 0.02	1.66 ± 0.30
SWA	0.76 ± 0.02	0.86 ± 0.01	2.71 ± 0.33
N-cuts	0.72 ± 0.02	0.84 ± 0.01	2.12 ± 0.17
Gpb	0.54 ± 0.01	$\mathbf{0.88} \pm 0.02$	7.20 ± 0.68
MS	0.57 ± 0.02	$\mathbf{0.88} \pm 0.01$	11.08 ± 0.96

Table 1. Results of single and multi-segment coverage on the dataset (95% confidence).

The results are given in Table 1, showing the average scores for all images in the data set. Since we experiment with different preprocessing superpixel segmentations, we denote using MS, FH and $SLIC$ with Our$_{MS}$, Our$_{FH}$ and Our$_{SLIC}$, respectively. The best results are achieved using $SLIC$ which has the lowest fragmentation and highest F measure for a single segment. Note that the all variants of our approach deliver the lowest fragmentation error. This means that our iterative approach consistently segments out the objects well regardless of the initial segmentation process.

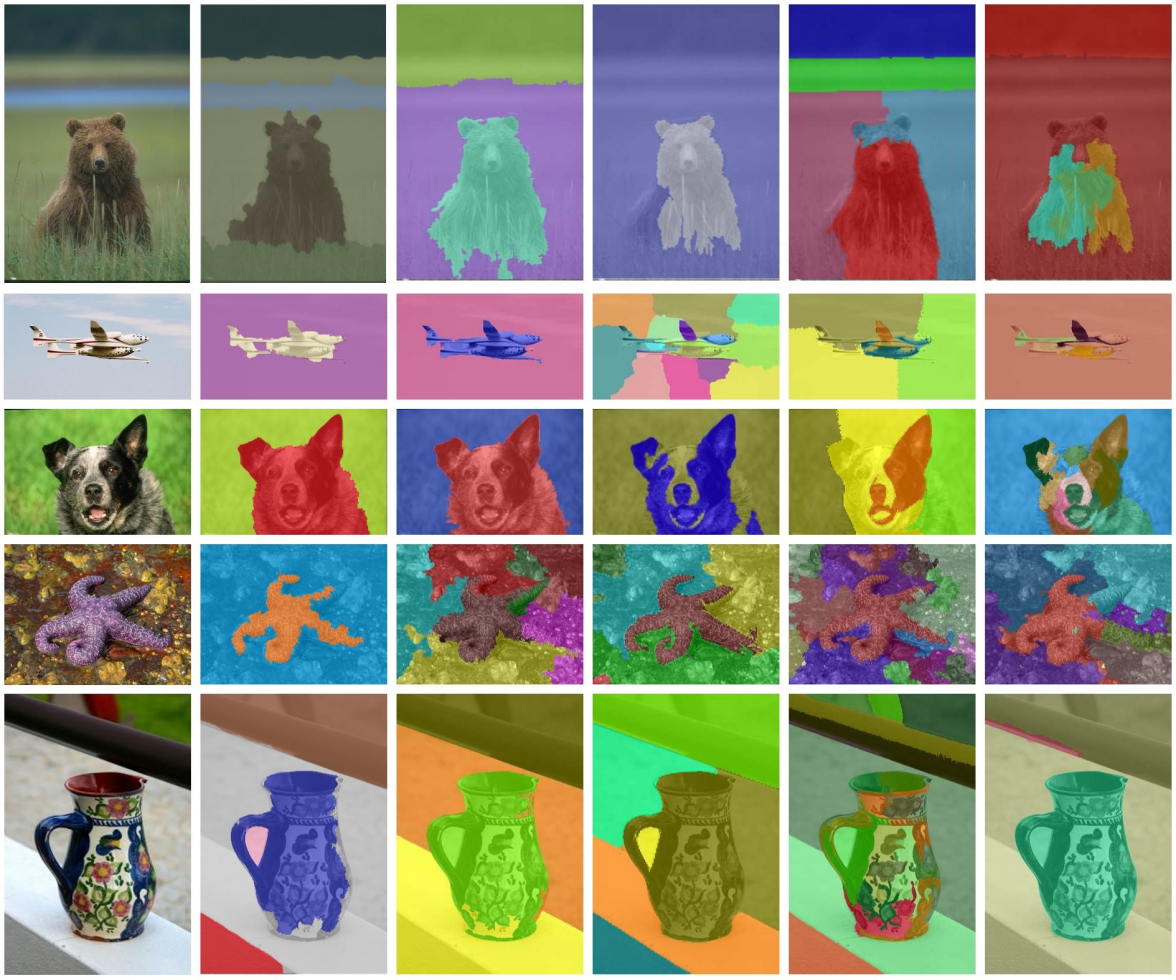

Figure 2. A few images from the dataset and different segmentations. From left to right: Original image, Our$_{SLIC}$, PBACI, SWA, Normalized cuts, and Mean shift.

Our algorithm delivers highly competitive results in both variants of the F measure. The advantage of our approach is very apparent in fragmentation, where it significantly outperforms the state-of-the-art, which means that it correctly identifies the object with an average of 1.4 segments regardless of the preprocessing step, whereas all other methods over-segment it.

It should be noted that there is an inverse relationship between the F measure, specifically F_{multi}, and the fragmentation. If a method has high fragmentation, meaning the foreground object is made up of several segments, it is natural to assume that they cover it better than a method that only produces one segment, but the ground truth has only one segment, which should be preferred. Therefore the advantage of our method is its correctly delineating the object in the image as being comprised of a single segment. This is because similar superpixels are identified as having the same label early in the iterative process and we are only left with a few segments.

Note that our method can artificially be made to favor improved multi-segment coverage at the cost of reduced one-segment coverage by increasing the F_{multi} which results in increase of F_{frag}. This can be achieved by forcing the SVMs to specialize to the superpixels belonging to their segment, which results in reduced merging of segments. For SVMs, this can be achieved by increasing the γ parameter, which enhances non-linearity and increases the specialization. A few examples of the (non-overspecialized) segmentation produced by our algorithm are shown in Figure 2.

4 Conclusion

An unsupervised iterative segmentation algorithm is proposed. The results show that the algorithm is comparable to the state-of-the-art in terms of precision and recall, and also outperforms the state-of-the-art by more often correctly identifying the segments belonging to a single object.

Future work will involve considering other classifiers

that allow efficient learning and classification of high-dimensional features. As the current segment labelling is binary (hard assignment), using soft-labelling with a segment belonging to different labels simultaneously could also be explored. The pairwise MRF energy term, i.e. the edge weight between neighbouring superpixels, depends on the color similarity, but could also be extended to texture. We will also consider a hierarchical approach in which the segmentation presented in this work acts as a prior on pixel-level segmentation to further improve the segmentation quality, by having more refined segment borders. Lastly, saliency detection, the task of determining the important regions of an image, could benefit from our approach as a preliminary step.

REFERENCES

[1] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE TPAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[2] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[3] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 309–314, 2004.

[4] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE TPAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[5] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 705–718.

[6] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE TPAMI*, vol. 33, no. 5, pp. 898–916, 2011.

[7] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 212–225, 2008.

[8] S. Alpert, M. Galun, A. Brandt, and R. Basri, "Image segmentation by probabilistic bottom-up aggregation and cue integration," *IEEE TPAMI*, vol. 34, no. 2, pp. 315–327, 2012.

[9] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image co-segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1943–1950.

[10] M. Cho, S. Kwak, C. Schmid, and J. Ponce, "Unsupervised object discovery and localization in the wild: Part-based matching with bottom-up region proposals," *arXiv preprint arXiv:1501.06170*, 2015.

[11] H. Mobahi, S. R. Rao, A. Y. Yang, S. S. Sastry, and Y. Ma, "Segmentation of natural images by texture and boundary compression," *International journal of computer vision*, vol. 95, no. 1, pp. 86–98, 2011.

[12] Z. Ren and G. Shakhnarovich, "Image segmentation by cascaded region agglomeration," in *CVPR*. IEEE, 2013, pp. 2011–2018.

[13] C. Wang, L. Zhao, S. Liang, L. Zhang, J. Jia, and Y. Wei, "Object proposal by multi-branch hierarchical segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[14] J. Wu and A. C. Chung, "A segmentation model using compound markov random fields based on a boundary model," *Image Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 241–252, 2007.

[15] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 1, pp. 45–57, 2001.

[16] K. A. Tran, N. Q. Vo, T. T. Nguyen, and G. Lee, "Gaussian mixture model based on hidden markov random field for color image segmentation," in *Ubiquitous Information Technologies and Applications*. Springer, 2014, pp. 189–197.

[17] M. Kristan, V. Sulic Kenk, S. Kovacic, and J. Pers, "Fast image-based obstacle detection from unmanned surface vehicles," *Cybernetics, IEEE Transactions on*, 2015.

[18] P. Rantalankila, J. Kannala, and E. Rahtu, "Generating object segmentation proposals using global and local search," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 2417–2424.

[19] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *ICCV*. IEEE, 2009, pp. 670–677.

[20] Z. Li, X.-M. Wu, and S.-F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 789–796.

[21] X. Wang, H. Li, C.-E. Bichot, S. Masnou, and L. Chen, "A graph-cut approach to image segmentation using an affinity graph based on l0-sparse representation of features," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 4019–4023.

[22] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[23] S. H. Anamandra and V. Chandrasekaran, "Color child: A robust and computationally efficient color image local descriptor," in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*. IEEE, 2014, pp. 227–234.

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[25] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt, "Hierarchy and adaptivity in segmenting visual scenes," *Nature*, vol. 442, no. 7104, pp. 810–813, 2006.

[26] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *International journal of computer vision*, vol. 43, no. 1, pp. 7–27, 2001.

Aleksandar Dimitriev received his B.Sc. degree from the Faculty of Computer and Information Science, University of Ljubljana, where he is currently an M.Sc. student. His research interests include probabilistic machine learning and Bayesian statistics and their applications to computer vision, bioinformatics and other fields of computer science.

Matej Kristan received his Ph.D from the Faculty of Electrical Engineering, University of Ljubljana in 2008. He is an Assistant Professor at the ViCoS Laboratory at the Faculty of Computer and Information Science and at the Faculty of Electrical Engineering, University of Ljubljana. His research interests include probabilistic methods for computer vision with a focus on visual tracking, dynamic models, online learning, object detection and vision for mobile robotics.