A survey on the state-of-the-art superpixel segmentation

Isabela Borlido Barcelos*
Image and Multimedia Data Science Laboratory (IMScience)
Pontifical Catholic University of Minas Gerais (PUC Minas)
Belo Horizonte, Brazil
Email: isabela_borlido@hotmail.com

Alexandre X. Falcão
Laboratory of Image Data Science (LIDS)
Institute of Computing (IC),
University of Campinas (UNICAMP)
Campinas, Brazil
Email: afalcao@ic.unicamp.br

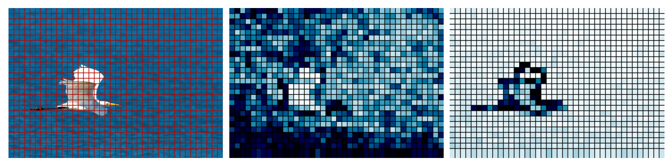
Silvio J. F. Guimarães
Image and Multimedia Data Science Laboratory (IMScience)
Pontifical Catholic University of Minas Gerais (PUC Minas)
Belo Horizonte, Brazil
Email: sjamil@pucminas.br

***Abstract*—In contrast to the rapid progress of superpixel segmentation, their methods are often compared only with classical approaches. Also, the previous superpixel methods categorizations are insufficient to cover the recent literature. In addition, although the inner color similarity usually underlies superpixel methods, both color homogeneity measures have important drawbacks. In this work, we fill these gaps by providing a new taxonomy for superpixel segmentation, a new color homogeneity measure, and an extensive comparison among 20 superpixel methods. Experiments show that the proposed measure, named *Similarity between Image and Reconstruction from Superpixels* (SIRS), is more robust to slight color variations than Explained Variation. Using SIRS and the commonly used superpixel metrics, we evaluated 20 superpixel segmentation methods and provided insights into the different approaches based on the clustering categories in our taxonomy.**

I. Introduction

Superpixel segmentation consists of partitioning images into several disjoint groups of connected pixels, named superpixels, according to a predetermined criterion (*e.g.*, color similarity). Such a procedure reduces workload, provides high-level semantic information, and enables accurate object delineation. Consequently, superpixel segmentation methods have been used in several applications [1]–[7]. In superpixel literature, regardless of the absence of consensus on the desired superpixels' properties, most authors agree that superpixels must be connected, adhere to the objects' borders, have a compact and regular shape, be computationally efficient, and produce a controllable number of superpixels [8], [9]. Nevertheless, since improving one property may lead to another's worsening, many superpixel methods fail to meet all of these criteria and try to manage this trade-off [10], [11].

Since superpixel methods can attend to different properties, the evaluation measures may vary depending on the optimized property. For instance, some measures only consider the objects' borders in the ground truth and ignore internal partitioning. However, this may not be sufficient since an image can consider different objects for different tasks. Other measures overcome ground-truth dependency by assessing color homogeneity [12], [13]. However, *Intra-cluster Variation* (ICV) [12] cannot be compared between images since it does not produce normalized values. In contrast, the *Explained Variation* (EV) [13] produces normalized values but may not accurately describe perceptually homogeneous regions in certain situations. Our proposed measure, the *Similarity between Image and Reconstruction from Superpixels* (SIRS), represents superpixels by their most frequent colors and measures color homogeneity as the image reconstruction error, accurately capturing color homogeneity, as shown in Figure 1. Superpixel papers usually compare their proposals to classical approaches with few comparisons with newer methods, hampering the determination of the true impact of their contributions. Furthermore, recent approaches have not been included in previous benchmarks [8], [9], [14]–[16]. This work fills this gap by evaluating 20 superpixel segmentation methods among the most recently proposed and commonly used ones. Our assessment covers compactness, delineation, color homogeneity, visual quality, and stability (omitted in this paper). The results provide valuable insights into the pros and cons of the methods, supporting the choice of the most suitable one for a given application. In summary, the contributions of this work are: (i) a comprehensive overview of the recent superpixel approaches; (ii) a taxonomy for superpixel methods with a less restrictive representation; (iii) a new color homogeneity measure for quantitative superpixel evaluation; (iv) an extensive assessment on various datasets.



(a) Segmentation (b) EV (c) SIRS

Fig. 1. Difference between color homogeneity measures in a grid segmentation with 1000 superpixels. Whiter values indicate higher scores.

This paper is organized as follows. Section II presents the mathematical image modeling used in this work. Subsequently, Section III presents the evaluation measure SIRS, and Section IV describes the superpixels taxonomy. In Section V we

*Work related to the M.Sc. dissertation entitled *A survey on the state-of-the-art superpixel segmentation*.

compare SIRS with the commonly used color homogeneity measure for superpixels and Section VI present an extensive evaluation with 20 superpixel methods. Finally, Section VII presents the conclusions and future works.

II. Background

Let an *image* $\mathcal{I}$ be defined as a pair $(\mathbf{I}, I)$ in which $\mathbf{I} \subset \mathbb{Z}^2$ is the set of *pixels' coordinates* whose *colors* is a vector mapped by $I(p) \in \mathbb{R}^m$, given $m \in \mathbb{N}^*$. Note that, when $m = 1$, $\mathcal{I}$ is *grayscale* and it is *colored* otherwise. We may compute the ℓ*-norm* of $I(p) = \langle I_1(p), \ldots, I_m(p)\rangle$ of the colors of the pixel p by $\|I(p)\|_\ell = \left(\sum_{j=1}^m |I_j(p)|^\ell\right)^{1/\ell}$, given $\ell \in \mathbb{N}^*$. By setting $\ell = 1$ and $\ell = 2$, the ℓ-norm is equivalent to the *Manhattan* and *Euclidean* distances, respectively.

If a set $X \subseteq \mathbf{I}$ of pixels is provided, one may calculate its *mean color* $\mu(X) \in \mathbb{R}^m$ by $\mu(X) = \frac{\sum_{x\in X} I(x)}{|X|}$, where $|X|$ denotes its size. Furthermore, we may *segment* X into $k \in \mathbb{N}^*$ subsets by a function $\mathbf{S}(X, k) \in \mathbb{P} \setminus \emptyset$, being $\mathbb{P}$ the *power set*, resulting in a *partition* (or *grouping*) $\{X_1, \ldots, X_K\}$ such that $\bigcup_{i=1}^k X_i = X$, $\bigcap_{i=1}^k X_i = \emptyset$, and $k \leq |X|$. We may extend such concepts for describing the *segmentation* $S \in \mathbf{S}(\mathbf{I}, k)$ of an image $\mathcal{I}$, in which every S_i is a *region* or *superpixel*.

III. A color homogeneity measure for superpixels

Since the average color used in the ICV and EV is insufficient to describe the colors of a superpixel, a new descriptor is necessary. Supposing that a small set of colors can represent a visually homogeneous superpixel, we can assume that a heterogeneous superpixel cannot be well represented in the same way. Therefore, the heterogeneity of a superpixel may be related to the quality of its description from a few colors. Thus, this descriptor's error in recovering the superpixel content from the original image may be related to its color homogeneity. Based on this idea, in this work, we assess the quality of the superpixel segmentation by its ability to reconstruct the original image. More formally, let $\mathcal{R} = (\mathbf{I}, R)$ be a *reconstructed image* of $\mathcal{I}$ in which every pixel $p \in \mathbf{I}$ has its reconstructed (or predicted) color $R(p) \in \mathbb{R}^m$. Such reconstruction is ideal when $R \equiv I$. If a segmentation S is provided, the popular approach is to assign $R(p) = \mu(S_i)$ for all $p \in S_i$ and every $S_i \in S$.

A. RGB Bucket Descriptor

We argue that the color information of any superpixel can be represented by a minimal set of colors due to its homogeneity property. In order to build the palette of the most relevant colors in each superpixel $S_i \in S$, we exploit the RGB space, represented as a cube in $[0,1]^3$. By merging the white and black color vertices, the vertices correspond to the colors with maximum intensity in some channels. Therefore, considering an image $\mathcal{I} = (\mathbf{I}, I)$ and the reconstructed image $\mathcal{R} = (\mathbf{I}, R)$, I and R map to normalized RGB colors.

First, let $G^{S_i} \in \mathbf{S}(S_i, 7)$ represent the set of 7 disjoint groups related to each of the cube's vertices, whose colors are $V = \{c_1, ..., c_8\}$, in which $c_l \in [0,1]^3$ for $1 \leq l \leq 7$. One may divide the RGB space according to the vertices of its cube representation and merge the white and black vertices to represent gray levels. Therefore, V corresponds to all possible combinations of RGB color channels. Let $x = \langle x_i\rangle_{i=1}^m$ a vector that indicates the color channels with maximum intensity in $I(p)$ such that $x_i = \mathbb{1}(I_i(p) = \|I(p)\|_\infty)$. We populate each $G_l^{S_i} \in G^{S_i}$ by assigning every $p \in S_i$ to its most similar group using a mapping function $M(p)$ (Equation 1).

TABLE I
FEATURES CATEGORIES IN TAXONOMY

Feature category	Explanation
Pixel-level	Raw data resources in images — *e.g.*, pixel color, position, and depth
Mid-level	features that can be computed based on a set of pixels, smaller than the entire image — *e.g.*, patch-based feature, path-based feature, gradient, or boundary
High-level	features that combine pixel properties and high-level information. The high-level information cannot be extracted from a small set of pixels. They are given directly by the user or predicted by other models — *e.g.*, saliency map, semantic features, texture, or a desired object geometry

$$M(p) = \operatorname*{argmin}_{c_i \in V} \{\|x - c_i\|_1\} \tag{1}$$

Although $G_l^{S_i}$ contains pixels similar to c_l, they may present significantly distinct luminosities (*i.e.*, color shades), which can be suppressed if the mean color is desired. Thus, we split it into $\lambda \in \mathbb{N}^*$ subgroups (or *buckets*), denoted by $\widehat{G}_l^{S_i} \in \mathbf{S}(G_l^{S_i}, \lambda)$. Without abuse of notation, we insert every $p \in G_l^{S_i}$ into its respective group $\widehat{G}_{l,b}^{S_i}$ given $b = \lfloor \|I(p)\|_\infty \lambda \rfloor$.

We name *RGB Bucket Descriptor* (RBD) the descriptor $\text{RBD}(S_i) = \{c_1, ..., c_\alpha\}$, in which $c_i \in [0,1]^3$, resultant from the selection of the $\alpha \in \mathbb{N}^*$ most relevant colors within G^{S_i} by some predetermined criterion. In this work, $\text{RBD}(S_i)$ selects the average color $\mu(G_{l,b}^{S_i})$ of the most populated buckets, irrespective of l (*i.e.*, its vertex-based group). Although inaccurate for heterogeneous sets of pixels, the refinement for generating $G_{l,b}^{S_i}$ leads to a better approximation of the most predominant colors by the mean operator. On the other hand, by promoting such grouping, colors with visually indistinguishable differences are assigned to the same bucket, reducing the probability of selecting slight variations of the most frequent color.

B. Similarity between Image and Reconstruction from Superpixels

Given $\text{RBD}(S_i) = \{c_1, \ldots, c_\alpha\}$, one could generate a proper approximation of the original texture by the correct ordering, but such task is challenging. Conversely, we propose evaluating the best reconstruction possible from the most relevant colors for measuring the color variation description of S_i. Thus, we build $\mathcal{R}$ such that $R(p) = \text{argmin}_{c_j \in \text{RBD}(S_i)} \{\|I(p) - c_j\|_1\}$.

After generating $\mathcal{R}$ from S, we may compute the *Mean Exponential Error* (MEE), shown in Equation 2 between it and the original image $\mathcal{I}$ for weighting each error accordingly:

$$\mathrm{MEE}(S) = \frac{1}{|\mathbf{I}|} \sum_{S_i \in S} \sum_{p \in S_i} \|R(p) - I(p)\|_1^{2-\psi} \tag{2}$$

in which $\psi = \max\left\{\|c_l - c_j\|_1\right\}$ and $c_l, c_j \in \mathrm{RBD}(S_i)$. If a superpixel requires a palette of highly discrepant colors, the error impact should be greater since it describes a complex pattern. Conversely, if the relevant colors are similar and, thus, represent a more uniform texture, such impact must be small. Finally, we may define the *Similarity between Image and Reconstruction from Superpixels* (SIRS), in Equation 3, by a Gaussian distribution centered at MEE(S):

$$\mathrm{SIRS}(S) = \exp^{-\frac{\mathrm{MEE}(S)}{\sigma^2}} \tag{3}$$

in which σ^2 is a parameter that controls the importance of small error variations. In SIRS, the higher the value, the better the color homogeneity in S, represented within $[0, 1]$.

IV. Taxonomy for Superpixel Segmentation

A suitable categorization must satisfy the following statements: (i) the categories must be abstract enough to encompass (*i.e.*, be valid for) all methods; and (ii) the categories must be distinct from each other to allow comparing and merging strategies. For a superpixel method, one may identify at most three steps: (i) initial processing, (ii) main processing, and (iii) final processing. The former performs a pre-processing whose output is used in the main processing for the superpixel computation strategy. Finally, cluster refinement may be performed in the final processing step to ensure connectivity or for fine-tuning. Superpixel methods may also use several features, such as boundary maps, semantic features, affinity maps, and saliency maps. Therefore, in addition to the processing steps, our taxonomy includes a feature classification based on their abstraction level. As a superpixel method can use more than one feature (*e.g.*, pixel colors and boundary map), our feature classification considers the used feature with the highest abstraction level among the categories in Table I.

We reviewed 45 superpixel methods and categorized their processing steps. For superpixel segmentation, neural architectures are not restricted to the main processing, being also used to extract features or for superpixel refinement. Although superpixel segmentation is a pixel labeling task, some neural architectures are trained for other tasks, such as image reconstruction or inpainting. Therefore, we categorize the neural networks in superpixel methods based on their architectures and training tasks. An extensive review of classical and recent superpixel methods along with the proposed taxonomy applied to the analyzed methods were omitted in this paper and can be seen in the dissertation text.

V. Evaluating Color Homogeneity Measure

In this Section, we present the experiments for our proposed evaluation measure, SIRS, in *Birds* [17], *Sky* [18], and *Extended Complex Scene Saliency Dataset* (ECSSD) [19] datasets using five superpixel methods with different properties. The *Birds* [17] consists of 150 images of Birds whose thin elongated legs are difficult to segment and, thus, may compromise the color description. The *Sky* [18] has 60 images with large homogeneous regions with subtle luminosity variations. Finally, the *Extended Complex Scene Saliency Dataset* (ECSSD) [19] has 1000 images with objects and backgrounds with complex textures. Specifically, DISF [20] and SH [21] are recent superpixel methods focused on object delineation, while IBIS [22] and SLIC [23] present high compactness with fair delineation. Finally, we consider a grid-based segmentation (GRID), representing maximum compactness but poor delineation. The implementation of SIRS is available online [3].

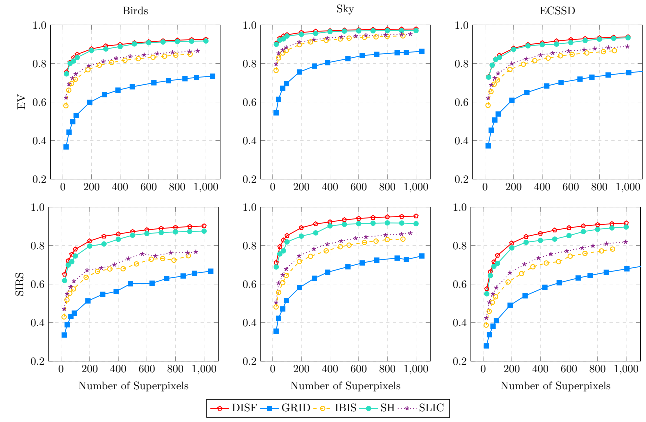


Fig. 2. Results for Birds, Sky, and ECSSD for EV and SIRS.

A. Quantitative results

As one can see in Figure 2, both SIRS and EV distinguish methods that maximize delineation (*i.e.*, DISF and SH) with those opting for more compact superpixels (*i.e.*, GRID, SLIC, and IBIS). However, EV presents a lesser spread than SIRS, as exemplified in the distance between IBIS' and SH's curves. Moreover, EV tends to result in significantly higher values, especially in contexts where superpixels are increasingly heterogeneous. For example, GRID obtains a score over 0.5 on the Sky dataset with only 25 superpixels. Conversely, SIRS offers a more meticulous discrepancy even with methods with similar performance, like DISF and SH. Also, due to its penalization, SIRS exhibits a more coherent range of values when few superpixels are generated — *i.e.*, in a more heterogeneous segmentation. In the same example, GRID scored less than 0.4 on the same dataset.

B. Qualitative results

Figure 3 presents a visual comparison between SIRS and EV in images with large homogeneous (Figure 3(a)-(c)) or texturized (Figure 3(d)-(i)) regions. As shown in Figure 3(a)-(c), while EV has a higher penalty even in smooth color transitions (the sky in Figure 3(a)-(c)), SIRS is robust to such changes. Concerning more textured backgrounds, SIRS also demonstrates robustness in simpler textures (Figures 3(d)-(f)). However, more complex textures (Figures 3(g)-(i)) may receive a significant penalty in SIRS, but are generally softer

[3] https://github.com/IsabelaBB/SIRS-superpixels

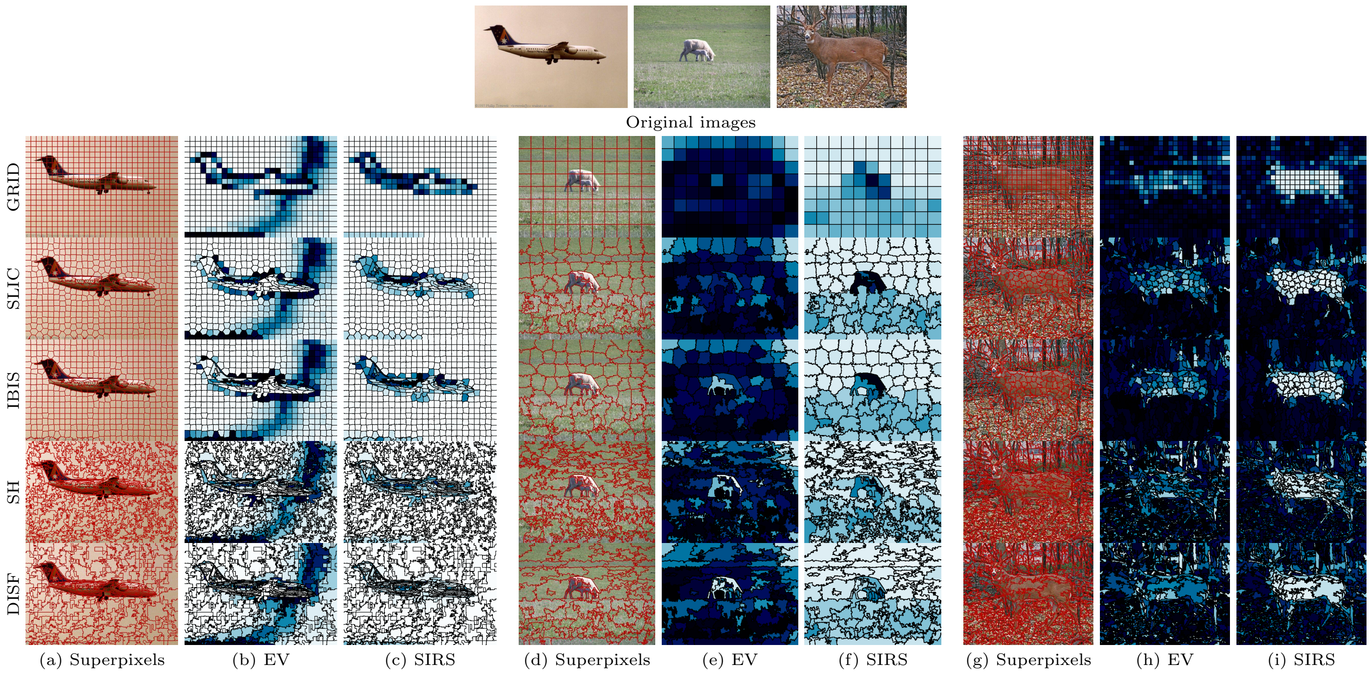


Fig. 3. Qualitative evaluation comparison with images from Sky (a)-(c) and ECSSD (d)-(i) with 100 and 500 superpixels with EV and SIRS evaluations. Low to high homogeneity scores are represented as dark blue to white colors.

than EV. Moreover, our measure tends to be more correlated to delineation than EV, given the penalizations in regions with high color variance — often at the object borders. As the delineation performance decreases, the color variation captured tends to be more heterogeneous (*i.e.*, RBD generates a more diverse palette), leading to a more drastic penalization. We argue that such robustness is directly linked to the accurate color selection from RBD, correctly describing superpixel homogeneity.

VI. Evaluating superpixel methods

We selected a total of 20 superpixel methods, where 13 are recent proposals not evaluated in previous benchmarks: DISF [20], RSS [24], ODISF [25], IBIS [22], DRW [26], DAL-HERS [27], LNSNet [28], ISF [29], SNIC [30], SH [21], GMMSP [31], LSC [32], and SCALP [33]. We also selected the 6 methods recommended in [8]: SLIC [23], SEEDS [34], ERS [35], ETPS [36], CRS [37], and ERGC [38]. Finally, a grid segmentation (GRID) was used as a baseline. We evaluated according to: (i) boundary adherence using Boundary Recall (BR) [39] and Undersegmentation Error (UE) [14]; (ii) compactness with Compactness index (CO) [10]; and (iii) color homogeneity with Explained Variation (EV) [13] and SIRS. We also analyzed stability but we omitted these results due to lack of space. As the SIRS assessment in Section V, we selected Birds, Sky, and ECSSD datasets as they contain different challenging aspects. We also choose the Insects dataset [17], composed of 130 images of spiders, insects, and other invertebrates, whose images have more homogeneous backgrounds than Birds.

A. Quantitative evaluation

As one may see in Figure 4, the differences between the best methods in BR and UE delineation are minor and most methods have low leakage due to low UE scores. Concerning CO (third row in Figure 4), methods with higher compactness (GRID, CRS, ETPS, SCALP, SNIC, SLIC, and IBIS) have a parameter to control compactness. When evaluating the color homogeneity (fourth and fifth rows in Figure 4) with EV and SIRS, the results of the former were generally too high and closer to each other compared to the second one. However, in both, the higher color homogeneity was achieved by DISF, followed by SH, ISF, LSC, RSS, GMMSP, and SCALP.

Most path-based clustering methods had similar performance in object delineation, compactness, and homogeneity. Conversely, neighborhood-based clustering approaches had more variate results. Although methods with path-based clustering had optimal delineation, their superpixels have low compactness. Among these, DISF and ERS showed the best results in delineation and homogeneity. In contrast, LSC and SCALP, both neighborhood-based clustering approaches, produced more compact superpixels. Clustering methods based on contour optimization also had variable results, with IBIS achieving the best ones and CRS and SEEDS producing the worst results. Contour evolution-based methods had higher compactness but worse object delineation and color homogeneity.

Concerning clustering with a dynamic center update, DRW, and SNIC use strategies to adapt the number of generated superpixels to the image content. Despite their similarities, DRW and SNIC use different features and optimization func-

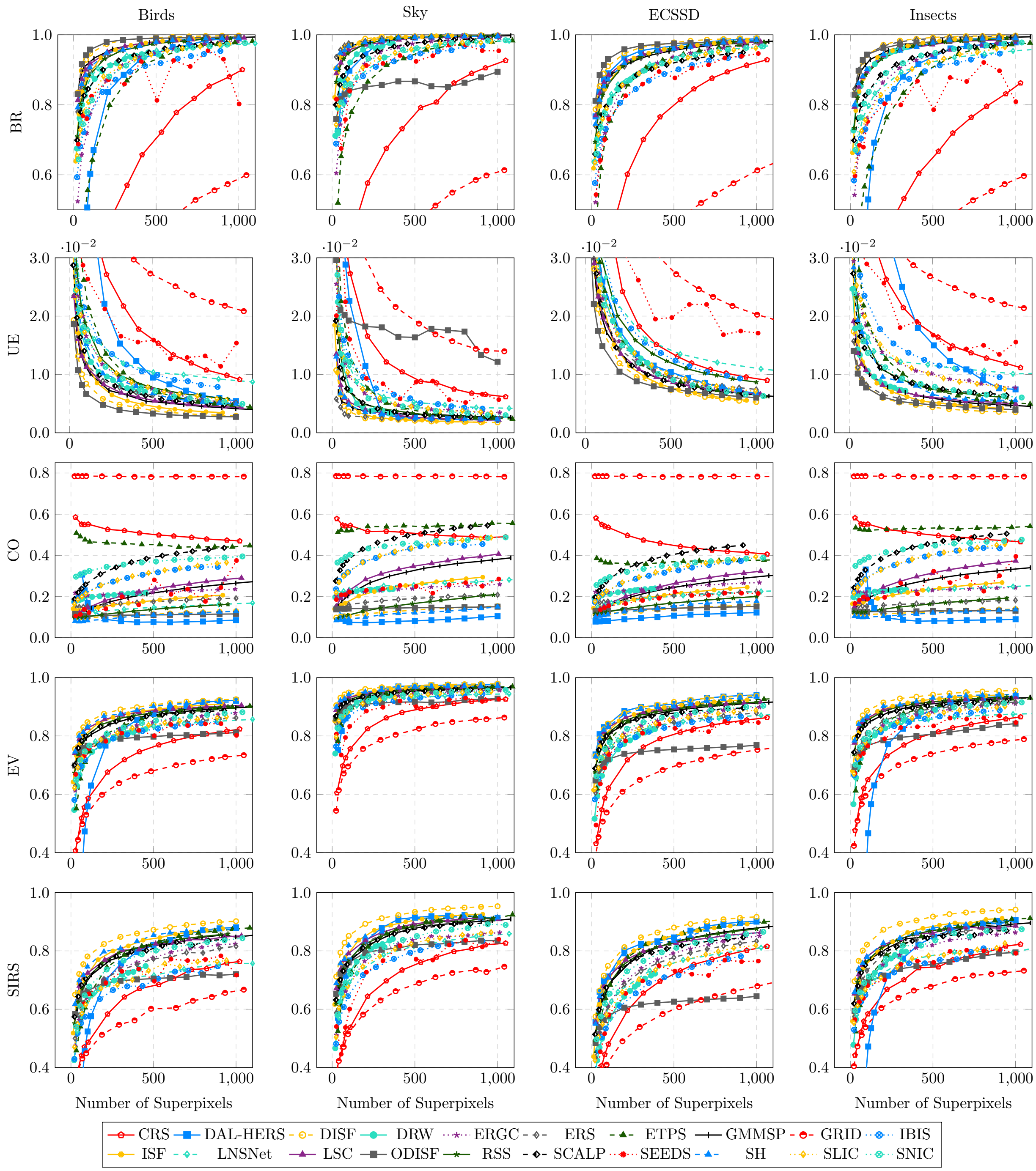


Fig. 4. Results for Birds, Sky, ECSSD, and Insects for BR, UE, CO, EV, and SIRS.

tions, which explains their different results. DRW has better delineation and fewer superpixels, while SNIC generates more compact and homogeneous superpixels. In our analysis, both hierarchical approaches, SH and DAL-HERS, had low compactness and high color homogeneity. However, SH had competitive delineation in contrast with worse results with DAL-HERS. Finally, GMMSP and LNSNet showed excellent delineation with BR, but LNSNet had heterogeneous superpixels with moderate compactness and more leakage. GMMSP achieved great results in all evaluated measures.

B. *Qualitative results*

Figure 5 presents superpixel segmentations, where the superpixel boundaries are shown in red. Concerning the path-based clustering methods, RSS produces elongated and thin superpixels at the strong image boundaries (*i.e.*, low compactity), while ISF produces compact superpixels in homogeneous regions but its high sensitivity to color variations generates

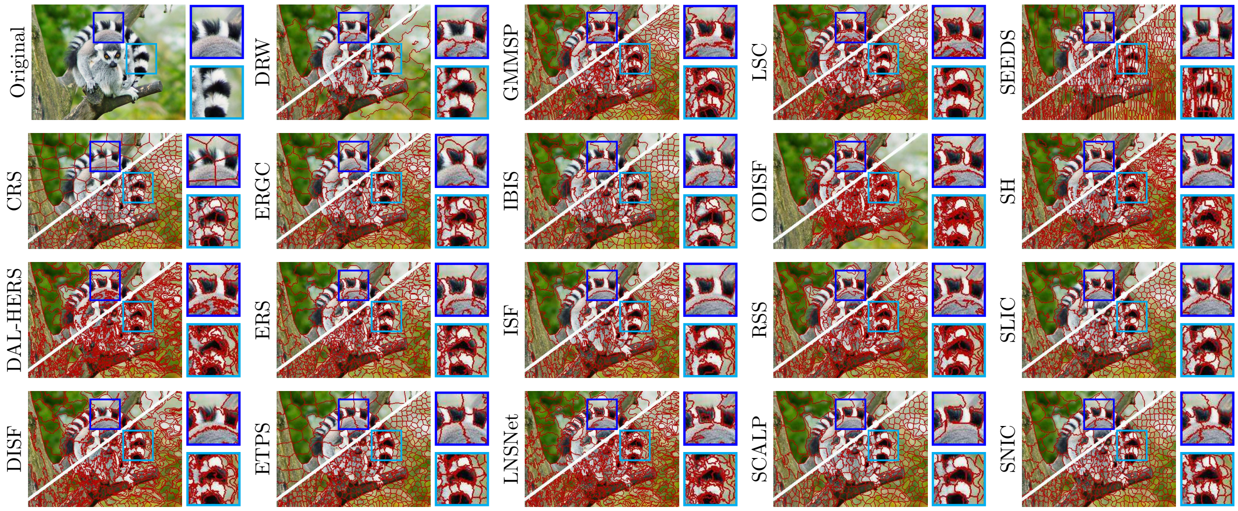


Fig. 5. Segmentation comparison with an image from ECSSD with 100 and 700 superpixels.

non-smooth superpixels in less homogeneous regions. Both DISF and ODISF maintain excellent boundary adherence and generate larger superpixels in homogeneous regions, but their superpixels are neither compact nor smooth. Similarly, ERGC has good boundary adherence, its superpixels do not have a high variance in size and they have smooth contours, but for fewer superpixels, the delineation reduces significantly.

Concerning neighborhood-based methods, SLIC and SCALP have highly compact superpixels with good boundary adherence. For fewer superpixels, SLIC compactness and delineation decrease. In contrast, LSC only produces compact superpixels in homogeneous regions but has smooth contours in simple textured areas. Similar to SLIC, SNIC produces more compact and well-defined superpixels, while DRW generates fewer, less compact, and more adherent superpixels. In contrast, superpixels in SEEDS are non-compact and non-smooth, which may lead to small leakage regions. CRS generates highly compact superpixels but with low adherence to the image boundaries. Similarly, ETPS and IBIS produce smooth and compact superpixels, but the compactness and smoothness in IBIS may vary depending on the region's homogeneity. In both ETPS and IBIS, the adherence to contours is significantly reduced with fewer superpixels.

Conversely, hierarchical methods (SH and DAL-HERS) have excellent delineation, but with non-smooth superpixels in highly textured areas and elongated and thin ones at some of the prominent image boundaries. DAL-HERS also generates several tiny superpixels, resulting in visibly poor segmentation. In contrast, LNSNet produces compact superpixels in homogeneous regions, but its sensitivity to color variations implies non-smoothness and its delineation may fail in regions with strong boundaries, causing small leaks. Methods with others' clustering strategies had excellent delineation. For instance, with graph-based clustering, ERS produces superpixels with low smoothness while GMMSP, with a data distribution-based clustering, produces smooth and compact superpixels, especially in homogeneous regions.

VII. Conclusion

In this work, we extensively review the recent literature on superpixel segmentation and propose a taxonomy for them. We also propose a new evaluation measure called SIRS to assess color homogeneity in superpixels. Compared to the EV, SIRS is more robust to less perceptual color variations. We also evaluate 20 superpixel methods. Our results demonstrate that path-based (ISF, DISF, ODISF, and ERGC) and hierarchical methods (SH and DAL-HERS) usually have the best delineation but low compactness. In contrast, most boundary evolution clustering methods (SEEDS, IBIS, CRS, and ETPS) have the highest compactness and the worst delineation. Also, neighborhood-based (SLIC, LSC, and SCALP) and dynamic center update clustering (SNIC and DRW) methods usually produce compact superpixels with moderate delineation. Concerning other clustering categories, graph-based (ERS) and data distribution-based clustering (GMMSP) obtain moderate compactness while a convolutional network for lifelong learning (LNSNet) had no compactness. The delineation is moderate for ERS and excellent for the others although LNSNet had more leakage. This study is part of the tutorial *”Superpixel segmentation: from theory to applications”*, in SIBGRAPI 2023. It also has an international conference paper [40], and a journal under review in ACM Surveys. Another conference paper [41] was published in parallel but is not part of the dissertation.

Acknowledgment

The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (Grant COFECUB 88887.191730/2018-00, Grant PROAP 88887.842889/2023-00 – PUC/MG and Finance Code 001), the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Grants 303808/2018-7, 407242/2021-0, 306573/2022-9) and Fundação de Apoio à Pesquisa do Estado de Minas Gerais – FAPEMIG (Grant APQ-01079-23 and PPM-00006-18).

References

[1] Y. Liang, Y. Zhang, Y. Wu, S. Tu, and C. Liu, "Robust video object segmentation via propagating seams and matching superpixels," *IEEE Access*, vol. 8, pp. 53 766–53 776, 2020.

[2] I. Borlido Barcelos, F. Belém, P. Miranda, A. X. Falcão, Z. K. G. do Patrocínio, and S. J. F. Guimarães, "Towards interactive image segmentation by dynamic and iterative spanning forest," in *Discrete Geometry and Mathematical Morphology*, J. Lindblad, F. Malmberg, and N. Sladoje, Eds. Cham: Springer International Publishing, 2021, pp. 351–364.

[3] W. Zhao, Y. Fu, X. Wei, and H. Wang, "An improved image semantic segmentation method based on superpixels and conditional random fields," *Applied Sciences*, vol. 8, no. 5, p. 837, 2018.

[4] L. Ren, L. Zhao, and Y. Wang, "A superpixel-based dual window rx for hyperspectral anomaly detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 7, pp. 1233–1237, 2019.

[5] J. Zhang, J. Chen, Q. Wang, and S. Chen, "Spatiotemporal saliency detection based on maximum consistency superpixels merging for video analysis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 606–614, 2019.

[6] X. Zhou, Y. Wang, Q. Zhu, C. Xiao, and X. Lu, "Ssg: superpixel segmentation and grabcut-based salient object segmentation," *The Visual Computer*, vol. 35, no. 3, pp. 385–398, 2019.

[7] P. Sellars, A. I. Aviles-Rivero, and C.-B. Schönlieb, "Superpixel contracted graph-based learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 4180–4193, 2020.

[8] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Computer Vision and Image Understanding*, vol. 166, pp. 1–27, 2018.

[9] M. Wang, X. Liu, Y. Gao, X. Ma, and N. Q. Soomro, "Superpixel segmentation: A benchmark," *Signal Processing: Image Communication*, vol. 56, pp. 28–39, 2017.

[10] A. Schick, M. Fischer, and R. Stiefelhagen, "Measuring and evaluating the compactness of superpixels," in *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. IEEE, 2012, pp. 930–934.

[11] ——, "An evaluation of the compactness of superpixels," *Pattern Recognition Letters*, vol. 43, pp. 71–80, 2014.

[12] W. Benesova and M. Kottman, "Fast superpixel segmentation using morphological processing," in *Conference on Machine Vision and Machine Learning*, 2014, pp. 67–1.

[13] A. P. Moore, S. J. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–8.

[14] P. Neubert and P. Protzel, "Superpixel benchmark and comparison," in *Proc. Forum Bildverarbeitung*, vol. 6, 2012, pp. 1–12.

[15] D. Stutz, "Superpixel segmentation: An evaluation," in *German conference on pattern recognition*. Springer, 2015, pp. 555–562.

[16] B. Mathieu, A. Crouzil, and J. B. Puel, "Oversegmentation methods: a new evaluation," in *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 2017, pp. 185–193.

[17] L. A. C. Mansilla and P. A. V. Miranda, "Oriented image foresting transform segmentation: Connectivity constraints with adjustable width," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2016, pp. 289–296.

[18] E. B. Alexandre, A. S. Chowdhury, A. X. Falcao, and P. A. V. Miranda, "Ift-slic: A general framework for superpixel generation based on simple linear iterative clustering and image foresting transform," in *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, 2015, pp. 337–344.

[19] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended cssd," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 4, pp. 717–729, 2015.

[20] F. C. Belém, S. J. F. Guimarães, and A. X. Falcão, "Superpixel segmentation using dynamic and iterative spanning forest," *IEEE Signal Processing Letters*, vol. 27, pp. 1440–1444, 2020.

[21] X. Wei, Q. Yang, Y. Gong, N. Ahuja, and M.-H. Yang, "Superpixel hierarchy," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 4838–4849, 2018.

[22] S. Bobbia, R. Macwan, Y. Benezeth, K. Nakamura, R. Gomez, and J. Dubois, "Iterative boundaries implicit identification for superpixels segmentation: a real-time approach," *IEEE Access*, vol. 9, pp. 77 250–77 263, 2021.

[23] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[24] D. Chai, "Rooted spanning superpixels," *International Journal of Computer Vision*, vol. 128, no. 12, pp. 2962–2978, 2020.

[25] F. C. Belém, B. Perret, J. Cousty, S. J. Guimarães, and A. X. Falcão, "Towards a simple and efficient object-based superpixel delineation framework," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2021, pp. 346–353.

[26] X. Kang, L. Zhu, and A. Ming, "Dynamic random walk for superpixel segmentation," *IEEE Transactions on Image Processing*, vol. 29, pp. 3871–3884, 2020.

[27] H. Peng, A. I. Aviles-Rivero, and C.-B. Schönlieb, "Hers superpixels: Deep affinity learning for hierarchical entropy rate segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 217–226.

[28] L. Zhu, Q. She, B. Zhang, Y. Lu, Z. Lu, D. Li, and J. Hu, "Learning the superpixel in a non-iterative and lifelong manner," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1225–1234.

[29] J. E. Vargas-Muñoz, A. S. Chowdhury, E. B. Alexandre, F. L. Galvão, P. A. V. Miranda, and A. X. Falcão, "An iterative spanning forest framework for superpixel segmentation," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3477–3489, 2019.

[30] R. Achanta and S. Susstrunk, "Superpixels and polygons using simple non-iterative clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7 2017.

[31] Z. Ban, J. Liu, and L. Cao, "Superpixel segmentation using gaussian mixture model," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4105–4117, 2018.

[32] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6 2015.

[33] R. Giraud, V.-T. Ta, and N. Papadakis, "Robust superpixels using color and contour features along linear path," *Computer Vision and Image Understanding*, vol. 170, pp. 1–13, 2018.

[34] M. Van den Bergh, X. Boix, G. Roig, B. de Capitani, and L. Van Gool, "Seeds: Superpixels extracted via energy-driven sampling," in *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 13–26.

[35] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *CVPR 2011*, 2011, pp. 2097–2104.

[36] J. Yao, M. Boben, S. Fidler, and R. Urtasun, "Real-time coarse-to-fine topologically preserving segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2947–2955.

[37] C. Conrad, M. Mertz, and R. Mester, "Contour-relaxed superpixels," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 280–293.

[38] P. Buyssens, I. Gardin, and S. Ruan, "Eikonal based region growing for superpixels generation: Application to semi-supervised real time organ segmentation in ct images," *IRBM*, vol. 35, no. 1, pp. 20–26, 2014, biomedical image segmentation using variational and statistical approaches. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1959031813001437

[39] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 5, pp. 530–549, 2004.

[40] I. B. Barcelos, F. D. C. Belém, L. D. M. João, A. X. Falcão, and G. S. JF, "Improving color homogeneity measure in superpixel segmentation assessment," in *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, vol. 1. IEEE, 2022, pp. 79–84.

[41] F. Belém, I. Borlido, L. João, B. Perret, J. Cousty, S. J. F. Guimarães, and A. Falcão, "Fast and effective superpixel segmentation using accurate saliency estimation," in *Discrete Geometry and Mathematical Morphology*, É. Baudrier, B. Naegel, A. Krähenbühl, and M. Tajine, Eds. Cham: Springer International Publishing, 2022, pp. 261–273.