Superpixel Clustering for Detection of Binary Object Hierarchy Using Modernized Classical Clustering Methods

Mikhail Kharinov
St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)
St. Petersburg, Russia
khar@iias.spb.su

***Abstract*. The paper proposes the simplest non-algorithmic definition of superpixels as image elements, which itself determines the algorithm for their calculation. A system of three classical methods of image approaching by piecewise constant approximations by means of iterative clustering of image pixels is considered: Ward's clustering, split-and-merge method and K-means method. The modernization of these methods is suggested for reduction of the approximation error *E* (total squared error) to the achievable minimum values for a fixed cluster numbers *g* in the current approximation. Advanced versions of the classical methods for reducing of the approximation error *E* are combined in so-called standard model for detecting of binary hierarchy of objects in the image by means of iterative superpixel clustering. In this paper the advanced versions of mentioned methods are presented and the standard model of binary hierarchy of objects in the image is briefly described.**

***Keywords*: digital image, pixel clustering, piecewise constant approximation, total squared error, minimization, superpixel hierarchy, object detection**

I. Introduction

Modern low-level computer vision and object detection in an image is rather a highly desirable project than an established science about image elements and how to organize and classify these elements to effectively detect objects in a scene. In order to make object detection at the initial stage of image processing a science, first of all, there is a lack of formal definition of image elements, definition of the hierarchy of objects in the image, as well as a system of methods for their effective calculation.

This paper provides a brief overview of the mathematical model, which proposes the definition of the elements that make up the image. The model provides the detection of hierarchically structured objects using modernized methods of classical cluster analysis. Since the model is based on the classical cluster analysis, it is called the standard model.

II. Superpixels

Although there are no generally accepted definition for image elements, the term has already been coined for them. Wanted image elements are called "superpixels".

This work was performed within the framework of the theme 0060-2019-0011 (Fundamentals and technologies of big data for sociocyberphysical systems).

The subject of superpixels [1] becomes especially relevant due to the increase in image resolution and the need to reduce the computational complexity of image processing by replacing operations with pixels by operations with superpixels. Usually superpixels are conceptualized as elements of objects or enlarged pixels, in particular, as image segments matching to the boundaries between objects. Grouping pixels is a very good method to reduce computational complexity at the initial stage of image processing. But this does not neglect the problem of justifying how to use the enlarged pixels in the best way.

Superpixels are defined as the maximal sets of pixels that implies preliminary construction of the initial $1, 2, ..., g_1$ series of optimal piecewise constant image approximations in bottom up merging and top down splitting techniques.

The idea of superpixels and approximating of an image by a hierarchy of superpixels is illustrated in Fig. 1.

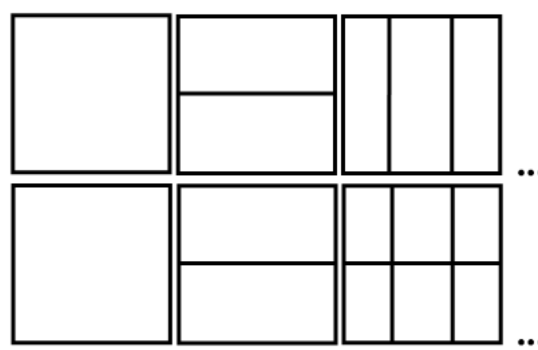

Fig. 1. The sequence of optimal image partitions (above) and the hierarchy of superpixel partitions (bottom).

If the optimal image approximations with $1, 2, 3, ..., g_1$ pixel clusters are determined by the sequence of partitions shown in the upper row of pictures, then the hierarchy of image approximations by superpixels is determined by the hierarchical sequence of partitions into 1, 2, 6..., s superpixels shown in the bottom row of pictures, where $g_1 \leq s \leq (g_1)!$and the exclamation mark "!" indicates a factorial.

Fig. 1 clearly explains that superpixels preserve the boundaries between pixel clusters that are disappeared in the current optimal approximation of the image. The effect of the disap-

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

pearance of boundaries between pixel clusters is caused by non-hierarchical optimal approximations. It just means that, starting from the optimal approximation, the computer does not "see" the sharp boundaries between the objects. Therefore, to detect objects, it is preferable to use superpixel clustering with a value of error E exceeding the minimum possible one for a given number of superpixels.

III. Example of Superpixels

Fig.2 presents the optimal and hierarchical superpixel approximations for standard "Lena" image.

In the left column the optimal image approximations with the number of tones from one to nine are demonstrated. The corresponding superpixel approximations of the image with 1, 2, 4, 7, 11, 16, 18, 24 and 28 tones are placed side by side in the right column. Note that the first and the second rows show the pairs of the same approximations as in Fig. 1. Under careful examination, it is not difficult to notice the differences in the approximations being compared, which are effortlessly manifested numerically.

Fig.3 describes the total sequence of 216 optimal approximations calculated for source integer pixel values without an initial pixel enlargement.

The upper graph in Fig.3 shows an increase in the number of clusters $s(g) \sim \sqrt{E(g)}$ in the superpixel image approximation accompanied with an increase in the number of clusters g in the current optimal approximation and a concomitant increment of the number g of reproducible optimal approximations. The dashed line on the graph corresponds to the case of a hierarchy of optimal approximations, when the superpixel approximations coincide with the optimal ones. A significant deviation of the curve $s(g)$ from the dashed straight line indicates for the current optimal approximation that there is a lack of data about sharp objects in optimal approximations with fewer clusters (Fig.1). The bottom graph in Fig.3 shows the dependence of the standard deviation σ on the number of clusters counted along the abscissa on a logarithmic scale. It expresses the approaching of a non-hierarchical sequence of optimal image approximations by the hierarchy of superpixel approximations. The dependence of the standard deviation σ on the number of clusters $s(g)$ in the superpixel approximation is shown by a black curve, and the dependence $\sigma(g)$ on the number of clusters g in the current optimal image approximation is shown in gray.

The hierarchy of superpixels is much more convenient for computing objects than a non-hierarchical set of clusters of optimal approximations. At the same time, the sequence of optimal approximations has a remarkable advantage, expressed in the fact that it is described by the convex dependence of the approximation errors E_g on the cluster numbers g:

$$E_g \leq \frac{E_{g-1} + E_{g+1}}{2}, g = 2, 3, ...,$$

which implies proper ordinate $\sigma \to E \sim \sigma^2$ transformation for the bottom graph in Fig.3.

Fig. 2. Sequences of optimal (left) and superpixel (right) approximations of "Lena" image.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

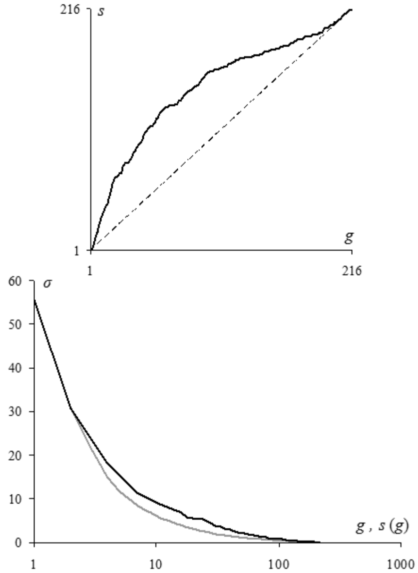

Fig. 3. The numbers of superpixels s depending on the number of clusters g (above). The curves $\sigma(g)$ for optimal (above) and $\sigma(s)$ for hierarchical superpixel image approximations (bottom).

IV. System of Three Clustering Methods

To get both advantages of optimal and hierarchical image approximations, it is quite attractive to construct a binary hierarchy of approximations, which is described by the convex dependence of the approximation error E on the number of clusters g. To do so, it seems reasonable to start with a sufficient number of superpixels and construct a target hierarchy that satisfies the convexity condition. In this case, the corresponding resulted curve will pass between the black and gray curves in the bottom graph of Fig.3.

At first glance, the optimal approximations for the above grayscale image, can be obtained using the multithreshold Otsu method [2], [3]. However, due to the exponential growth in the complexity of calculations, it is impossible to obtain even two dozens optimal approximations using this method. The problem is solved (Fig.2, Fig.3) through the combined use of another several algorithms of E minimization. These turned out to be three classical methods of cluster analysis, namely, Ward's pixel clustering, split-and-merge method, and K-means method [4]. An attractive circumstance is that these methods are suitable not only for grayscale, but also for color images. But, on the other hand, for the most effective application to images, they require some modernization.

Original Ward's clustering [5], [6] provides the generation of hierarchical image approximations described by the convex dependence of the approximation error E on the cluster number g. To reduce computational complexity, some seed segments or clusters of close pixels are commonly instead of the original pixels [6]. However, due to high computational complexity, original Ward's method is rarely used. But its main drawback is the unstable clustering result, which varies depending on the number of enlarged pixel and on the heuristic enlarged pixel themselves.

In order not to deal with the case-dependent convex sequence $E(g)$ and, at the same time, speed up the processing, Ward's pixel clustering is performed within the parts of the image. Since the computational complexity of the Ward's method is quadratically depends on the number N of pixels in the image, it decreases just as fast when the image is processed by parts. According to our estimate [7], for recursively repeated processing by parts, the computational complexity is drop as $N^2 \rightarrow N^{\frac{4}{3}} \rightarrow N^{\frac{16}{15}} \rightarrow N^{\frac{256}{255}} \rightarrow etc$, depending on the recursion step.

In a single processing by parts, first the image pixels are divided into g_0 clusters, and for each cluster its own approximation hierarchy is constructed in the merge mode as for separate image. Then the hierarchies are combined into one, by reordering of the cluster mergings and hierarchy constructing is completed by the original Ward's method.

If the provided input partition of pixels into g_0 clusters coincides with the optimal one, then the above procedure will lead to a convex dependence $E(g)$. Otherwise, the convexity may be violated at g_0 value. To avoid violation of the convexity property at g_0 clusters, it is enough to process the input approximation by so called CI-method that is modernized split-and-merge reducing of the approximation error E without changing the number of clusters.

CI-method [7] locally supports the convexity condition and provides for a given image approximation a real minimization of the approximation error E under an unchanged cluster number g_0. It corrects the convexity violation at the cluster number g_0 and ensures the separability of Ward's clustering by the image parts. In this case, Ward's clustering by the image parts is simply recalculated into the conventional Ward's clustering without modifying of pixel clusters obtained for the image parts.

The commonly used K-means method, developed and adapted a long time ago for calculations by means of arithmometer, is replaced in the standard model by the significantly modernized K-meanless method [8], which, presupposes the preliminary approximation error E reduction to the vicinity of the minimum achievable values by recursive Ward's method and CI method.

Unlike K-means, versions of K-meanless method for estimating of the error increment either use the values of the target functional E [8] or the exact formula [9]. Thus, the modernization K-means→K-meanless consists, first of all, in a more accurate analytical criterion for reclassifying the sets of pixels from cluster to cluster, as well as, in minimizing E by reclassification not only individual pixels, but pixel sets from the entire hierarchy of the cluster parts of various sizes.

The modernized three methods for minimizing of the approximation error E all and sundry are distinguished from conventional ones by the use of the entire hierarchy of image clusters. In the case of CI-method and K-method for improving of the quality of a separate image approximation, the entire hierarchy of clusters is corrected through the iterative alternate application of three mentioned methods.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

V. Binary Object Hierarchies

In the standard model, all pixels of the image are divided among themselves into pixels of various objects, which are relegated, for example, to background or to other types of objects. It is supposed that objects make up a binary hierarchy produced by the image partition into g_0 basic objects while other objects are represented as parts or unions of basic objects.

A characteristic feature of the standard model is that the basic objects and the binary hierarchy of objects, germinated by them, are detected ambiguously. Depending on the number of objects g_0, as well as a concrete task, it is set during the process of tuning up of the software system for the best approximation of target objects by means of a minimum number of pixel clusters. For a given number of basic objects g_0, the optimal approximation of the image in g_0 colors with the minimum achievable approximation error $E \sim \sigma^2$ or the standard deviation σ is considered as the best.

VI. Standard Model of Object Detection

The declared standard model for detecting objects in a digital image actually combines two equivalent models, namely, a meaningful one, which describes the detection of objects by means of sets of pixels of an image matrix, and a computational model, in which the considered pixel clusters are described by networked graphs that support high-speed construction and transformation of pixel clusters, as well as storing and calculating of the required features without repeating the calculations once performed.

The standard model, supporting the reversible computation [10], [11] to describe an ambiguous image as superposition of $g_0 = 1, 2, ...N$ object hierarchies, consists of:

- the concept of an object, binary hierarchy of objects, and the definition of superpixels, from which the system of necessary algorithms, model parameters, data structure and methods of accelerating of calculations are deduced;
- three above modernized method of classical cluster analysis as well as the method of filtering of the objects according to the established threshold [7];
- the data structure of algebraic multilayer network [7], in terms of which the definitions for objects and superpixels are reformulated and high-speed algorithms are performed.

In the standard model, the following settings are provided for detecting objects in the image:

- the number g_0 of basic objects identified with the pixel clusters of the optimal image approximation in g_0 colors, which is contained in the target hierarchy of image approximations;
- the number s of superpixels or the corresponding number g_1 of available optimal image approximations constituted of superpixels without any distortion;
- threshold parameter of heterogeneity $H_{threshold}$ for heterogeneity $\left|H \equiv \frac{dE}{dg}\right|$ itself, determined as the absolute value of the derivative of the approximation error E with respect to the number g of pixel clusters [7].

VII. Conclusion

Thus, within the framework of the standard model, the problem of binary hierarchy of objects detecting in an ambiguous image has been formulated and practically solved. Formally, the solution to this problem is expressed as an approximation of a "convex sequence" of optimal piecewise-constant image approximations by means of the same "convex binary hierarchical sequence" of image approximations, which contains an optimal approximation of an image in g_0 colors.

The peculiarity of just this paper is that for the invariant representation of the image regardless of the scale, as well as for the high-speed object detection without loss of accuracy, the definition of image elements (superpixels) is suggested, and an example of their calculation for the standard "Lena" image is given. In the future, it is planned to organize a database of optimal approximations and corresponding hierarchies of superpixels for "Lena", as well as for other standard images, and place the database in the public domain together with program texts and their executable modules used to generate superpixels and optimal image approximations from database.

The experience of experimental research of the standard model shows that it expands the area of effective application of classical cluster analysis by solving problems of processing images of various content [12].

References

[1] 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, 34(11), pp. 2274–2282, 2012.

[2] N. Otsu "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics 9(1), pp. 62–66, 1979.

[3] P.S. Liao , T.S. Chen, P.C. Chung "A fast algorithm for multilevel thresholding," J. Inf. Sci. Eng. 17(5), pp. 713–727, 2001.

[4] I.D. Mandel Cluster Analysis, Moscow: Finance and Statistics, 176 pp., 1988.

[5] J.H. Jr.Ward "Hierarchical grouping to optimize an objective function," Am. Stat. Assoc. 58(Issue 301), pp. 236–244, 1963.

[6] T. N. Tran, R. Wehrens, L. M. C. Buydens "SpaRef: a clustering algorithm for multispectral images," Analytica Chimica Acta 490 (1–2), pp. 303–312, 2003.

[7] M.V. Kharinov, A.N. Buslavsky "Object Detection in Color Image," In: Proc. of the 14th Intern. Conf. on Pattern Recognition and Information Processing (PRIP'2019), Minsk, 21–23 May, pp. 43–47, 2019, doi:10.13140/RG.2.2.28493.28640.

[8] S.D. Dvoenko "Meanless k-means as k-meanless clustering with the bi-partial approach," In: Proc. of the 12th Intern. Conf. on Pattern Recognition and Information Processing (PRIP'2014), Minsk, Belarus, pp. 50–54, 2014.

[9] M.V. Kharinov "Reclassification formula that provides to surpass K–means method," arXiv preprint, arXiv: 1209.6204v1, 10 pp., 2012.

[10] T. Toffoli "Reversible computing," In International Colloquium on Automata, Languages, and Programming, Springer Berlin Heidelberg. pp. 632–644, 1980.

[11] M.V. Kharinov "Reversible merging of structured clusters of pixels," Computer graphics and vision (Graphicon'2016), Proc. of the 26th intern. conf. on computer graphics and vision: Moscow, MSU, Sept. 19 - 23 2016, Nizhny Novgorod, pp. 298–302, 2016.

[12] I.G. Khanykov, V.A. Nenashev "The Application of the High-Speed Pixel Clustering Method in Combining Multi-Angle Images Obtained from Airborne Optical-Location Systems," 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), IEEE 2020, pp.1–8, doi:10.1109/WECONF48837.2020.9131157.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.