Proceedings of the International Multiconference on
Computer Science and Information Technology pp. 717–722
ISBN 978-83-60810-27-9
ISSN 1896-7094

Region-based Measures for Evaluation of Color Image Segmentation

Andreea Iancu, Bogdan Popescu, Marius Brezovan and Eugen Ganea
University of Craiova
Software Engineering Department
Craiova, Bd. Decebal 107, Romania
{andreea.iancu, bogdan.popescu}@itsix.com, {brezovan_marius, ganea_eugen}@software.ucv.ro

***Abstract*—The present paper is aimed to compare the efficiency of a new segmentation method with several existing approaches. The paper addresses the problem of image segmentation evaluation from the error measurement point of view.We are introducing a new method of salient object recognition with very good results relative to other already known object detection methods. We developed a simple evaluation framework in order to compare the results of our method with other segmentation methods. The experimental results offer a complete basis for parallel analysis with respect to the precision of our algorithm, rather than the individual efficiency.**

I. Introduction

The problem of segmentation is an important research field and many segmentation methods have been proposed in the literature so far ([1],[3],[4],[7]). The aim of image segmentation is the domain-independent partition of the image into a set of regions which are visually distinct and uniform with respect to some property, such as grey level, texture or color.

Image segmentation is one of the most important operations performed on acquired images. Image segmentation evaluation [2] focuses on two main properties: objectivity and generality. Objectivity means that all the test images in the benchmark should have an unambiguous ground-truth segmentation so that the evaluation can be conducted objectively. Generality means that the test images in the benchmark should have a large variety so that the evaluation results can be extended to other images and applications.

The main objective of this paper is to emphasize the very good results of image segmentation obtained by our segmentation technique, $Graph-Based\ Salient\ Object\ Detection$, and to compare them with other existing methods. The algorithms that we use for comparison are: $Normalized\ Cuts$, $Efficient\ Graph-Based\ Image\ Segmentation$ ($Local\ Variation$) and $Mean\ Shift$. All of them are complex and well known algorithms, with very good results in this area and building the knowledge based on their results represents a solid reference.

The experiments were completed using the images and ground-truth segmentations in the Berkeley segmentation dataset [8]. Since the ground-truth segmentation may not be well and uniquely defined, each test image in the Berkeley benchmark is manually segmented by a group of people.

The segmentation accuracy is measured taken into consideration the global consistency error and the local consistency error. We will provide comparative results that reflect a well-balanced behavior of the algorithm we propose.

The paper is organized as follows. In Section III we briefly present previous studies in the domain of image segmentation and the segmentation method we propose. The methodology of performance evaluation is presented in Section IV. The experimental results are presented in Section V. Section VI concludes the paper and outlines the main directions of the future work.

II. Related Work

Image segmentation evaluation is an open subject in today's image processing field. The goal of existing studies is to establish the accuracy of each individual approach and find new improvement methods. The segmentation methods require ground truth image segmentations as reference. The main drawback of providing such reference is represented by the resources that are needed. However, after analyzing the differences between the image under study and the ground truth segmentation, a performance proof is obtained.

Region-based segmentation methods can be broadly classified as either model-based [13] or visual feature-based [14] approaches. A distinct category of region-based segmentation methods that is relevant to our approach is represented by graph-based segmentation methods. Most graph-based segmentation methods attempt to search a certain structures in the associated edge weighted graph constructed on the image pixels, such as minimum spanning tree [3], or minimum cut [15].

Berkeley image segmentation benchmark is the reference that we use for our study. Using the same input, the provided image dataset, we are developing a customized methodology in order to efficiently evaluate our algorithm.

The closest work to ours is [3], in which an image segmentation is produced by creating a forest of minimum spanning trees of the connected components of the associated weighted graph of the image. The novelty of our contribution concerns two main aspects: (a) in order to minimize the running time we construct a hexagonal structure based on the image pixels, that is used in both color-based and syntactic-based segmentation algorithms, and (b) we propose an efficient method for segmentation of color images based on spanning trees and both color and syntactic features of regions.

III. Segmentation Methods

We will compare four different segmentation techniques, the Mean Shift-Based segmentation algorithm [4], Efficient Graph-Based segmentation algorithm [3], Normalized Cuts segmentation algorithm [7] and our own region-based segmentation method. We have chosen Mean Shift-Based segmentation because it is generally effective and has become widely-used in the vision community. The Efficient Graph-Based segmentation algorithm was chosen as an interesting comparison to the Mean Shift. Its general approach is similar, however, it excludes the mean shift filtering step itself, thus partially addressing the question of whether the filtering step is useful. Due to its computational efficiency, Normalized Cuts represents a solid reference in our study. We use all these algorithms as terms of comparison for the evaluation we performed.

A. Graph-Based Salient Object Detection

We present an efficient segmentation method that uses color and some geometric features of an image to process it and create a reliable result [10]. The used color space is RGB because of the color consistency and its computational efficiency.

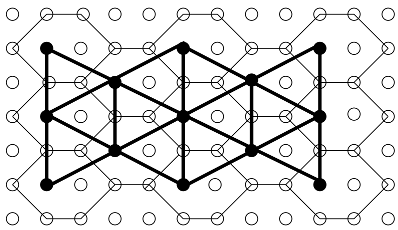

Fig. 1. The grid-graph constructed on the hexagonal structure of an image

What is particular at this approach is the basic usage of hexagonal structure instead of color pixels. In this way we can represent the structure as a grid-graph $G = (V, E)$ where each hexagon h in the structure has a corresponding vertex $v \in V$, as presented in Figure 1. Each hexagon has six neighbors and each neighborhood connection is represented by an edge in the set E of the graph. For each hexagon on the structure two important attributes are associated: the dominant color and the coordinates of the gravity center. Basically, each hexagonal cell contains eight pixels: six from the frontier and two from the middle.

Image segmentation is realized in two distinct steps. The first step represents a pre-segmentation step when only color information is used to determine an initial segmentation. The second step represents a syntactic-based segmentation step when both color and geometric properties of regions are used.

The first step of the segmentation algorithm uses a color-based region model and will produce a forest of maximum spanning trees based on a modified form of the Kruskal's algorithm. In this case the evidence for a boundary between two adjacent regions is based on the difference between the internal contrast and the external contrast between the regions. The color-based segmentation algorithm builds a maximal spanning tree for each salient region of the input image.

The second step of the segmentation algorithm uses a new graph, which has a vertex for each connected component determined by the color-based segmentation algorithm. In this case the region model contains in addition some geometric properties of regions such as the area of the region and the region boundary. The final segmentation step produces a forest of minimum spanning trees based on a modified form of the Borůvka's algorithm. Each determined minimum spanning tree represents a final salient region determined by the segmentation algorithm.

B. Efficient Graph-Based Image Segmentation

Efficient Graph-Based image segmentation [3], is an efficient method of performing image segmentation. The basic principle is to directly process the data points of the image, using a variation of single linkage clustering without any additional filtering. A minimum spanning tree of the data points is used to perform traditional single linkage clustering from which any edges with length greater than a given threshold are removed [6].

Let $G = (V, E)$ be a fully connected graph, with m edges $\{e_i\}$ and n vertices. Each vertex is a pixel, x, represented in the feature space. The final segmentation will be $S = (C_1, ..., C_r)$, where C_i is a cluster of data points. The algorithm [3] can be shortly presented as follows:

1) Sort $E = (e_1, ..., e_m)$ such that $|e_t| \le |e'_t| \forall t < t'$
2) Let S^0=$(\{x_1\}, ..., \{x_n\})$ in other words each initial cluster contains exactly one vertex.
3) For $t = 1, ..., m$
    a) Let x_i and x_j be the vertices connected by e_t.
    b) Let $C^{t-1}_{x_i}$ be the connected component containing point x_i on iteration $t-1$ and $l_i = max_{mst} C^{t-1}_{x_i}$ be the longest edge in the minimum spanning tree of $C^{t-1}_{x_i}$. Likewise for l_j.
    c) Merge $C^{t-1}_{x_i}$ and $C^{t-1}_{x_j}$ if:

$$|e_t| < min\left\{ l_i + \frac{k}{C^{t-1}_{x_i}}, l_j + \frac{k}{C^{t-1}_{x_j}} \right\} \qquad (1)$$

4) $S = S^m$.

C. Normalized Cuts

Normalized Cuts method models an image using a graph $G = (V, E)$, where V is a set of vertices corresponding to image pixels and E is a set of edges connecting neighboring pixels. The edge weight $w(u, v)$ describes the affinity between two vertices u and v based on different metrics like proximity and intensity similarity. The algorithm segments an image into two segments that correspond to a graph cut (A, B), where A and B are the vertices in the two resulting subgraphs.

The segmentation cost is defined by:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \qquad (2)$$

where $cut(A,B) = \sum_{u \in A, v \in B} w(u,v)$ is the cut cost of (A,B) and $assoc(A,V) = \sum_{u \in A, v \in V} w(u,v)$ is the association between A and V. The algorithm finds a graph cut (A,B) with a minimum cost in $Eq.(1)$. Since this is a NP-complete problem, a spectral graph algorithm was developed to find an approximate solution [7]. This algorithm can be recursively applied on the resulting subgraphs to get more segments. For this method, the most important parameter is the number of regions to be segmented. Normalized Cuts is an unbiased measure of dissociation between the subgraphs, and it has the property that minimizing normalized cuts leads directly to maximizing the normalized association relative to the total association within the sub-groups.

D. Mean Shift

The Mean Shift-Based segmentation technique [4] is one of many techniques dealing with "feature space analysis". Advantages of feature-space methods are the global representation of the original data and the excellent tolerance to noise [9]. The algorithm has two important steps: a mean shift filtering of the image data in feature space and a clustering process of the data points already filtered. During the filtering step, segments are processed using the kernel density estimation of the gradient. Details can be found in[4]. A uniform kernel for gradient estimation with radius vector $h = [h_s, h_s, h_r, h_r, h_r]$ is used. h_s is the radius of the spatial dimensions and h_r the radius of the color dimensions. Combining these two parameters, complex analysis can be performed while training on different subjects.

Mean shift filtering is only a preprocessing step. Another step is required in the segmentation process: clustering of the filtered data points $\{x'\}$. During filtering, each data point in the feature space is replaced by its corresponding mode. This suggests a single linkage clustering that converts the filtered points into a segmentation.

Another paper that describes the clustering is [5]. A region adjacency graph (RAG) is created to hierarchically cluster the modes. Also, edge information from an edge detector is combined with the color information to better guide the clustering. This is the method used in the available EDISON system, also described in [5]. The EDISON system is the implementation we use in our evaluation system.

IV. Region-based performance evaluation

We present comparative results of segmentation performance for our region based segmentation method and the three alternative segmentation methods mentioned above.

Our evaluation measure is mainly related to the consistency between segmentations. We use segmentation error measures that provide an objective analysis of the segmentation algorithms.

A potential problem for a measure of consistency between segmentations is that there is no unique human segmentation of an image, since each human perceives the scene differently. In this situation you could declare the segmentations inconsistent. However, if one segmentation is a refinement of the other, then the error should be small. Therefore, the measures are designed to be tolerant to refinement. Some other aspects to be taken into account are that error measure should not depend on the pixelation level and should be tolerant to noise along region boundaries.[12]

We used two metrics in order to provide an objective comparison between the four segmentation methods and the human segmentation. The two error measures are described below. We applied the measures to the Berkeley segmentation database and the segmentation results of the four algorithms.

In order to describe the segmentation errors, we considered two different segmentations S_1 and S_2 and calculated a value in the range $[0..1]$ where 0 represents no error. For a given p_i we considered segments S_1 and S_2 that contain the pixel. If one segment is a proper subset of the other, then the pixel lies in an area of refinement, and the local error should be zero. Otherwise, the two regions overlap in a inconsistent manner and we should calculate the corresponding error. We use $\backslash$to denote the set difference and $|x|$ for cardinality of set x.If $R(S, p_i)$ is the set of pixels corresponding to the region in segmentation S that contains pixel p_i, the local refinement error is defined as:

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) n R(S_2, p_i)|}{|R(S_1, p_i)|} \qquad (3)$$

This local error measure is not symmetric. It encodes a measure of refinement in one direction only: $E(S_1, S_2, p_i)$ is zero precisely when S_1 is a refinement of S_2 at pixel p_i, but not vice versa. Considering this local refinement error in each direction at each pixel, there are two methods to combine the values into an error measure for the entire image. We apply two error measures as follows: $Global\ Consistency\ Error$ (GCE) that forces all local refinements to be in the same direction and $Local\ Consistency\ Error$ (LCE) that allows refinement in different directions in different parts of the image.[11] For a given n as the number of pixels we have:

$$GCE(S_1, S_2) = \frac{1}{n} min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\} \qquad (4)$$

$$LCE(S_1, S_2) = \frac{1}{n} \sum_i min\{E(S_1, S_2, p_i), E(S_2, S_1, p_i)\} \qquad (5)$$

We will evaluate the performance of our algorithm on the Berkeley Segmentation Database (BSD) [8]. We will refer the characteristics of the error metrics previously defined by Martin et al. [2], explore potential problems with these metrics in order to evaluate the quality of each segmentation and to characterize its performance over a range of parameter values.

The current public version of the Berkeley Segmentation Database is composed of 300 color images. The images have a size of 481×321 pixels, and are divided into two sets, a training set containing 200 images that can be used to tune the parameters of a segmentation algorithm, and a testing set containing the remaining 100 images on which the final performance evaluations should be carried out.

We built a custom benchmark framework, that processes the Berkeley dataset, converts it to our proprietary format and preforms parallel analysis. Additionally, we adapted the other mentioned algorithms to the same evaluation format for unitary purposes.

The human segmented images provide the ground truth boundaries. Therefore, any boundary marked by a human subject is considered to be valid. Since there are multiple segmentations of each image by different subjects, it is the collection of these human-marked boundaries that constitutes the ground truth. Based on the output of the previously presented algorithms for a set of images, we will determine how well the ground truth boundaries are approximated.

In order to determine an algorithm's efficiency by comparing it to the ground truth boundaries, a threshold of the boundary map is needed.

We are providing an additional evaluation based on histogram representation of the error density characteristic for each algorithm.

V. Experimental Results

Our study of segmentation quality is based on experimental results and uses the Berkeley segmentation dataset provided at [8].

A. GCE and LCE Metrics

In order to proper evaluate the segmentation method we propose, we first need to better understand how the GCE and LCE error metrics work. Given two extreme cases:an under-segmented image, where every pixel has the same label (i.e. the segmentation contains only one region spanning the whole image), and a completely over-segmented image in which every pixel has a different label.

From the definitions of the GCE and LCE we can see that both measures evaluate to 0 on both of these extreme situations regardless of what segmentation they are being compared to. The reason for this can be found in the tolerance of these measures to refinement. Any segmentations is a refinement of the completely under-segmented image, while the completely over-segmented image is a refinement of any other segmentation.

In order to have a better analysis result and a more complete description for the errors we considered, we have performed 10 different tests for each subject per algorithm - Fig. 2.

More precisely, by varying several key parameters, we have obtained 10 distinct points that define the errors for each approach. For $Normalized\ Cuts$ [7] we have modified the $number\ of\ segments$ in the range of $\{5, 10, 12, 15, 20, 25, 30, 40, 50, 70\}$. The variable parameter

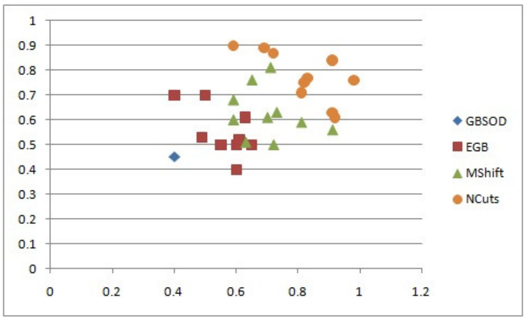

Fig. 2. Average GCE vs. LCE for Berkeley test images

for $Efficient\ Graph - Based\ Image\ Segmentation$ [3] was the scale of observation, k , in range {100, 200, 300, 400, 500, 600, 700, 800, 900, 1000}. For Mean-Shift [4] we have made 10 combinations from $Spatial\ Bandwidth$ {8, 16} and $Range\ Bandwidth$ {4, 7, 10, 13, 16}.

We calculated the GCE and LCE average values for the 100 test images provided by Berkeley. Figure 2 illustrates the GCE vs. LCE graphic representation.

In the resulting diagram (Fig. 2) we can see that the GCE vs. LCE error metric for our proposed method, denoted $GBSOD - Graph\ Based\ Salient\ Object\ Detection$ is situated below the values for the other algorithms indicating a better performance result, a smaller average error and a balanced algorithm. Analyzing the set of results for each parameter per algorithm, it's easy to distinguish which algorithm is generating better results; the smaller the error it is, the better is the accuracy of the respective algorithm.

B. Histogram based evaluation

We elaborated a histogram-based evaluation mechanism aimed to compare the segmentation results for the studied algorithms via the errors metrics.

In order to achieve this, we considered the human segmentation as the ground-truth segmentation and compared each algorithm with it, measuring the error metrics GCE and LCE.

For each algorithm we analyzed the 100 test images from Berkley and calculated the corresponding GCE and LCE. The histograms presented below illustrate this approach (Fig.3 - Fig. 10).

For a better description of the histogram based analysis, in Fig.3 and Fig.4 we have depicted the distribution of the the values of GCE respectively LCE for the 100 images processed using $GBSOD - Graph\ Based\ Salient\ Object\ Detection$ algorithm. It is very important that these value are more concentrated on smaller error values, which gives us the confidence that the presented method has good results. In Fig.5 and Fig.6 it can be seen the same analysis for $Normalized\ Cuts$, in Fig.7 and Fig.8 for $Efficient\ Graph - Based\ Image\ Segmentation$ and in Fig.9 and Fig.10 for $Mean - Shift$ algorithm.

Figures Fig.11 and Fig.12 are illustrating a comparison between all the four presented algorithms; it gives a good

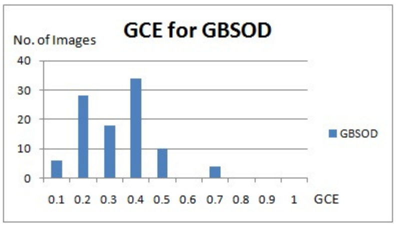

Fig. 3. GCE for Graph Based Salient Object Detection

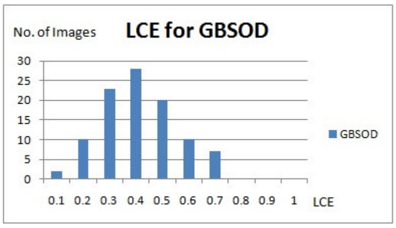

Fig. 4. LCE for Graph Based Salient Object Detection

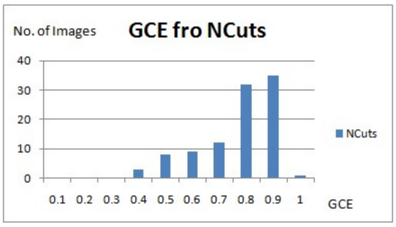

Fig. 5. GCE for Normalized Cuts

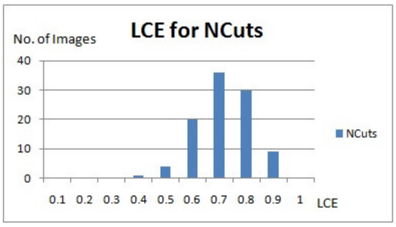

Fig. 6. LCE for Normalized Cuts

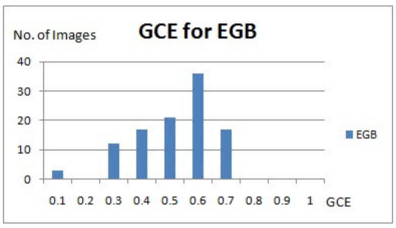

Fig. 7. GCE for Efficient Graph-Based Image Segmentation

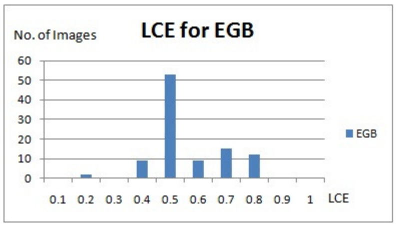

Fig. 8. LCE for Efficient Graph-Based Image Segmentation

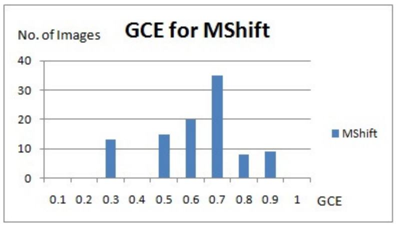

Fig. 9. GCE for Mean-Shift

perspective on the what error values generates each studied algorithm.

VI. Conclusion

In this paper we presented a new graph-method for image segmentation and extraction of visual objects. Starting from a survey of several segmentation strategies, we aimed at performing an image segmentation evaluation experiment.

Our segmentation method and other three segmentation methodologies were chosen for the experiment, and the complementary nature of the methods was demonstrated in the results. The study results offer a clear view of the effectiveness of each segmentation algorithm, trying in this way to offer a solid reference for future studies.

Future work will be carried out in the direction of integrating syntactic visual information into a semantic level of a semantic image processing and indexing system.

Acknowledgment

This work was supported by $CNCSIS - UEFISCSU$, project number $PNII - IDEI$ code/2008.

References

[1] K. Fu and J. Mui, "A survey on image segmentation. Pattern Recognition", 1981.

[2] 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", Proc. Int. Conf. Comp. Vis., vol. 2, pp. 416-425, 2001.

[3] P. Felzenszwalb and D. Huttenlocher, "Efficient Graph-Based Image Segmentation", Intl J. Computer Vision, vol. 59, no. 2, 2004.

[4] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, 2002.

[5] C. Christoudias, B. Georgescu, and P. Meer, "Synergism in Low Level Vision", Proc. Intl Conf. Pattern Recognition, vol. 4, pp. 150-156, 2002.

[6] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward Objective Evaluation of Image Segmentation Algorithms", IEEE Transactions on pattern analysis and machine inteligence, Vol. 29, No. 6, 2007.

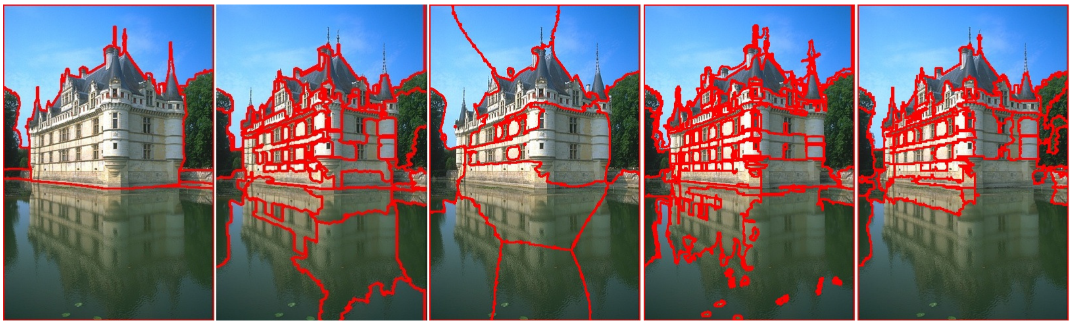

Fig. 13. Comparative segmentation results: Human Segmentation, Graph-based Salient Object Detection, Normalized Cuts, Efficient Graph-based, Mean-Shift

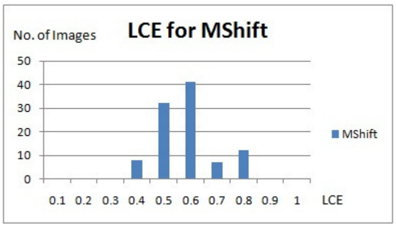

Fig. 10. LCE for Mean-Shift

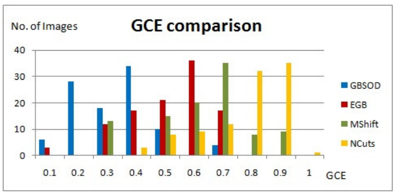

Fig. 11. GCE overall comparison

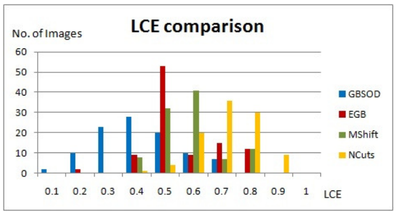

Fig. 12. LCE overall comparison

[7] J. Shi and J. Malik, “Normalized Cuts and Image Segmentation”, IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, No. 8, 2000.
[8] “Berkeley Segmentation and Boundary Detection Benchmark and Dataset”, 2003, http://www.cs.berkeley.edu/projects/vision/grouping/segbench.
[9] R.O. Duda, P.E. Hart, and D.G. Stork, “Pattern Classification”, John Wiley & Sons, New York, 2000.
[10] D. Burdescu, M. Brezovan, E. Ganea, and L. Stanescu, “A New Method for Segmentation of Images Represented in a HSV Color Space”, Lecture Notes in Computer Science, 5807, 606-617, 2009.
[11] D. Martin, C. Fowlkes, D. Tall, and J. Malik, “A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics”, ICCV Vancouver, 2001.
[12] R. Unnikrishnan, C. Pantofaru, and M. Hebert, “A Measure for Objective Evaluation of Image Segmentation Algorithms”,
[13] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Image segmentation using expectation-maximization and its application to image querying and classification”, IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(8),1026–1037, 2002.
[14] J. Fauqueur and N. Boujemaa, “Region-based image retrieval: Fast coarse segmentation and fine color description”, Journal of Visual Languages and Computing, 15(1), 69–95, 2004.
[15] J. Shi and J. Malik, “Normalized cuts and image segmentation”, Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 731–737, 1997.