UNSUPERVISED COSEGMENTATION BASED ON GLOBAL CLUSTERING AND SALIENCY

Lucas Lattari[*], *Anselmo Montenegro*[†], *Cristina Vasconcelos*

Instituto de Computação
Universidade Federal Fluminense
Niterói, RJ, Brazil

ABSTRACT

This paper introduces a new method for unsupervised cosegmentation. Our method combines saliency information with a Global Clustering step, which reveals parts of the objects by detecting similar subregions across image collections, based on a low dimensional descriptor that includes color, texture and positional features. The saliency information is used to yield a classification of the global clusters into foreground and background and also classify regions not detected as global clusters into potential background or foreground. These four types of regions are the input seeds for a Graph Cuts procedure that computes the final cosegmentation. The Graph Cuts result can also be used to compute a refined version of the saliency information which enables us to define an iterative cosegmentation pipeline. Our framework produces remarkable results in comparison with state-of-the-art works, even in challenging datasets with illumination variance, occluded objects and identical background.

Index Terms— Cosaliency, Cosegmentation, Graph Cuts, Salient Object Detection, Markov Random Fields.

1. INTRODUCTION

In the past 40 years, several variations of the image segmentation problem were investigated by the computer vision community. This paper investigates the Image Cosegmentation (IC) problem, which can be understood as the task of simultaneously segmenting a set of images by separating the foreground from the background in each image. The foreground is usually considered as the set of objects of interest with equivalent visual attributes. This problem was introduced by Rother et al. [1] in a restricted scenario, where only two images were cosegmented and a nearly identical foreground lies in front of a distinct background. Since then, the cosegmentation problem was further explored in different ways [2, 3, 4].

Our proposal is classified as unsupervised, since it does not receive any cue indicating which the objects of interest, prior to the Cosegmentation, are and it does not depend on any learning steps based on pairs of input images for which a ground truth is given.

Vicente et al. [2] introduce the concept of ”objectness” into their Cosegmentation framework, which detects the foreground of the collection by building hundreds of features extracted from a given set of ground-truth images. The segmentation is finally obtained via a Random Forest regressor of these features. Chai et al. [3] proposed a Cosegmentation approach based on two steps. The first step uses a 1076-dimensional descriptor to segment each image individually with a Grabcut [5], using a seed window at the center of the figure. The segmented regions are separated into object or background and used as input to train a standard linear support vector machine (SVM) [6]. Kim et al. [7] formulate a distributed cosegmentation approach based on a hierarchical intra-inter image segmentation. They initially group weakly supervised data into global homogeneous cluster, followed by a constrained affinity matrix that shows how to correlate the segmentation layers in order to obtain the final result. The intra-inter image segmentation approach of Kim et al. is similar in philosophy to our local-Global Clustering steps as the experiments have shown.

Many Cosegmentation approaches do not adequately take into account the level of similarity that may exist between the objects of interest and the background, i.e., many collections may have nearly identical color features shared by the object and the background. On the other hand, in some images, the object may not be highlighted, causing the failure of standard cosaliency methods [8]. Methods based purely on saliency may not present satisfactory results when dealing with noisy images, partially occluded objects or images with multiple objects of interest. Our hypothesis is that to solve these issues, it is crucial to combine many different attributes such as color, texture patches, bidimensional positioning and saliency, to obtain the desired results.

As in Yu et al. [4], we propose a Cosegmentation Markov Random Field framework that incorporates Cosaliency in our model. Nevertheless, in our method, we combined the cosaliency strategy with a Global Clustering approach that groups similar image regions across the collection, based on their color, texture and position. In these experiments, we show that the Global Clustering step is able to reveal new

[*] Thanks to CAPES agency for funding.

[†] Thanks to FAPERJ agency for funding under the project E-26/110.345/2014.

978-1-4799-8339-1/15/$31.00 ©2015 IEEE

objects containing information about the objects of interest, which are not captured by pure saliency-based methods. By combining saliency information with Global Clustering, we develop a new model that can reveal the foreground and background of the image set in a more robust way than previous methods as, for instance, [4].

Our contributions are summarized as follows:

1. The proposal of a new object-background model that combines intra and inter image clustering and saliency information, that is able to reveal which parts of the image are background and foreground, relying on descriptors with a reduced number of features.

2. The introduction of a Global Clustering approach that identifies and groups object and background regions across the image collection into Gaussian components with similar features.

3. A system that extends the original Object Salient Detection of [9] to image collections. It uses such model in a Cosegmentation pipeline with a fine segmentation step yielded by a Graph Cut approach. In our pipeline, it is possible to reuse the final Cosegmentation result as a refined saliency map which can be used as input to a new cosegmentation iteration, resulting in an iterative method.

2. PROPOSED METHOD

Consider an image collection $I = \{I_1, \dots, I_n\}$ and a family of features $F_i(e)$, where e is a subregion of an image I_i. We model the Image Cosegmentation problem as the determination of the simultaneous segmentation of I, where a segmentation is a partition of I_i into two groups $O(I_i)$ and $B(I_i)$, such that $O(I_i) \cup B(I_i) = I_i$ and $O(I_i) \cap B(I_i) = \emptyset$. $O(I_i)$ represents the object of interest that co-occurs in I_i, where the notion of interest in each image is given by a function $f(I_i) : I \rightarrow \mathbb{R}$. Here, the function of interest is defined in terms of saliency maps, but other aspects could be used to measure what is considered an object of interest.

The conceptual model of the method is subdivided into four steps named as: *Local Clustering*, *Global Clustering*, *Object Cosegmentation* and *Cosegmentation Refinement*. Its scheme is presented by the flowchart in Figure 1.

One of the foundations of the image cosegmentation problem is the existence of similar looking objects across the image collection I. The image regions defined by such objects tend to share nearly identical features. For each I_i, it is necessary to extract features F_i that will be used during the whole process of cosegmentation. We consider three types of features based on color, texture patches and spatial localization.

The Local Clustering substage is responsible for constructing a set of clusters for each I_i. Well known techniques such as K-means [10] and mixture of Gaussians [11] compute good solutions for individual images. From these initial intra-clusters, we aim to identify in the next stage those which are similar among the entire image collection, obtaining sets of inter-image clusters. The Local Clustering stage is further subdivided into Feature Extraction and Image Content Clustering subtasks. The Feature Extraction produces $F_i(e)$ features, where e is a subregion of I associated to pixels $p \in I_i$ for each image $I_i, i = \{1, \dots, n\}$ separately. The feature vector $F_i(e) = \{col(e), tex(e), pos(e)\} \in \mathbb{R}^d$ is composed by three major components, describing, respectively: color, texture and bidimensional position.

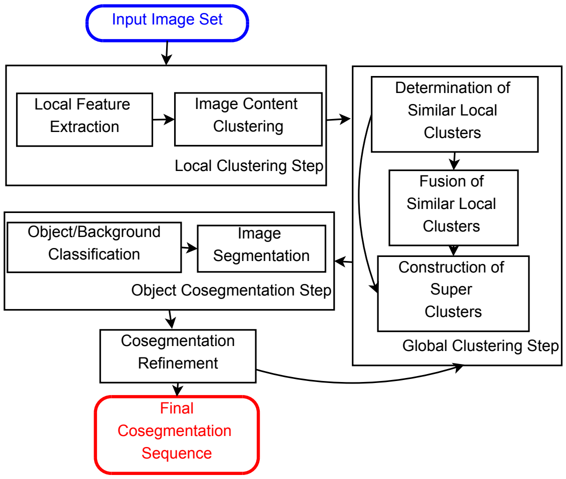
Input Image Set
Local Feature Extraction
Image Content Clustering
Local Clustering Step
Determination of Similar Local Clusters
Fusion of Similar Local Clusters
Construction of Super Clusters
Global Clustering Step
Object/Background Classification
Image Segmentation
Object Cosegmentation Step
Cosegmentation Refinement
Final Cosegmentation Sequence

Fig. 1: A flowchart that depicts the proposed method.

The Image Content Clustering is responsible for analyzing each I_i and grouping pixels of similar features into a set of Local Clusters $LC_i = \{LC_1, \dots, LC_i^m\}$, where m is the maximum number of clusters. After the construction of each individual cluster LC_i^s of similar pixels, their descriptors in a d-dimensional space $\mathbb{R}^d$ are extracted.

The next substage is the *Global Clustering* that, in a bottom-up perspective, attempts to identify the existence of groups of similar clusters across different images, creating *Global Clusters* (GC). Two Local Clusters LC_i^s and LC_j^t, with respective indices s and t, in distinct images I_i and I_j, are fused into a single Global Cluster GC_k, if they are considered similar, that is, $dist(LC_i^s, LC_j^t) < \epsilon_{global}$. Hence, $dist$ is a distance function defined on the Local Cluster descriptors given by the Bhattacharyya Distance [12]. The ϵ_{global} constant is the minimum distance between all pairs LC_i^s and LC_i^t that belong to the same I_i. The Global Clustering stage detects which Local Clusters are similar across the collection, determining which are to be fused, so that super clusters are generated. Each GC_k exists across the collection I.

After the Global Clustering phase, several Local Clusters may be left ungrouped. To overcome this problem, a top-down approach computes descriptors for the set GC_s already created. The similarity notion $reevalDist$ is used to revisit the remaining Local Clusters in a reevaluation loop. Their descriptors are then compared to the Global Clusters descriptors, that is, a Local Cluster LC_i^s will belong to a Global Cluster GC_k if $reevalDist(LC_i^s, GC_k) < \epsilon_{global}$. The function $reevalDist$ also uses the Bhattacharyya Distance [12].

The *Object Cosegmentation* task receives the set of all Global Clusters GC, where each GC_k is then classified into two categories: *object* or *background*. The classification procedure analyzes each I_i where a certain Global Cluster GC_k occurs, searching for cues that indicate whether it is an object or a background. After a Global Cluster is classified as an object or background, it is used as a seed model for the fine-grained cosegmentation using Graph Cuts.

The Object/Background classification procedure depends on the definition of saliency information as defined in Cheng et al. [9]. Consider that, for each I_i, a salient image S_i is computed. Each pixel $p \in GC_k$ is evaluated so that each GC_k is classified as object or background by using a voting scheme. If p belongs to the salient region of S_i, then $S(p) = 1$. Otherwise, p is part of the non salient area, and $S(p) = 0$. The total summation of salient pixels of GC_k is denoted as $|S(GC_k) = 1|$. Contrariwise, the total number of not salient pixels $p \in GC_k$ is defined as $|S(GC_k) = 0|$. Finally, for a collection I, if $|S(GC_k) = 1| > |S(GC_k) = 0|$, then GC_k is object. Otherwise, it is classified as background.

Even after the conclusion of the object/background classification procedure, there may be regions not classified. Here we do not ignore these cases, since many Local Clusters can be relevant for the computation of the final result. Thereby, each Local Cluster $LC_i \notin GC$ is classified as *probable object* or *probable background*, based on their saliency maps [9]. This is done in a similar manner to the previous classification procedure. That is, if $|S(LC_i) = 1| > |S(LC_i) = 0|$, then LC_i is considered a probably object region, otherwise as probably background.

In the Image Segmentation subtask, the classified regions associated to the Global Clusters and Local Clusters are used as seeds for a Graph Cuts energy minimization framework [13]. In our method, we replaced the typical user interaction seeds by those computed from the model inferred from the classified global and local clusters.

Finally, after the Object Cosegmentation, we noticed that it is possible to reuse these segmented images and restart the method, iterating many times as necessary. This iterative approach generally positively impacts the accuracy of the method. We denote it as *Cosegmentation Refinement*. This is done by replacing the original salient images with our cosegmented images, produced in the first iterations of the presented approach. More specifically, after finishing the method computation for the first time, the algorithm is restarted in the Global Clustering stage, but their original salient images are replaced by the cosegmentation obtained in the previous iteration. Our experiments have shown that a very small number of interactions is sufficient to produce quite good results. In some simpler cases, even one interaction is sufficient.

Table 1: Cosegmentation results obtained with the iCoseg dataset by the proposed method. Bold numbers highlight the best method for each collection.

Class	Ours	[2] (Single)	[2] (All)	[19]	[20]	[21]
Alaskan Bear	**93.6**	79.0	90.0	60.4	58.2	74.8
Baseball	**97.4**	84.5	90.9	74.6	69.9	73.0
Stonehenge	**95.2**	84.2	63.3	83.3	61.1	56.6
Stonehenge 2	**90.6**	88.9	88.8	79.7	66.9	86.0
Liverpool	**94.6**	87.4	87.5	83.2	70.6	76.4
Ferrari	**92.8**	84.8	89.9	71.8	77.7	85.0
Taj Mahal	88.9	80.7	**91.1**	82.2	79.6	73.7
Elephant	**95.1**	75.4	43.1	74.3	62.3	70.1
Panda	91.7	87.8	**92.7**	79.5	80.0	84.0
Kite	80.3	89.3	**90.3**	81.5	87.0	87.0
Kite Panda	**95.8**	80.2	90.2	87.7	70.7	73.2
Gymnastics	**97.3**	82.1	91.7	72.2	83.4	90.9
Skating	**89.3**	78.4	77.5	73.4	69.9	82.1
Balloon	**98.4**	79.5	90.1	97.5	89.3	85.2
Statue	**98.0**	92.9	93.8	91.5	89.3	90.6
Bear	**97.8**	78.2	95.3	83.5	87.3	74.0

3. EXPERIMENTS

In this section, we discuss practical aspects of our experiments and present its performance compared to unsupervised state-of-the-art approaches in well known databases: the iCoseg [14], the MSRC [15] and the Weizmann horses [16]. The measure used is the accuracy, which is the percentage of pixels correctly classified as object or background. Our accuracy reported is the average of accuracies after three tests for each collection with images selected randomly.

The implementation of all clustering tasks was done in the Matlab toolbox. We used L*a*b* colors, Gabor texture features [17] and bidimensional pixel position to represent feature descriptors in the Local Cluster stage, encompassing a 77-dimensional feature vector, for images with standard resolution. We also used C++ and OpenCV for GrabCut implementation and Cheng et al. [9] code solution for saliency maps computation. Our Local Clustering task is initialized by a K-means [10]. Later, a GMM is computed by the Expectation-Maximization algorithm [18] with maximum number of Local Clusters $K = 20$.

The iCoseg dataset was introduced in [14] for an interactive supervised Cosegmentation framework. It is a challenging database considering that their objects vary considerably in terms of viewpoint and illumination. For a fair comparison with [2] and other works, we evaluate experiments in similar condition to theirs. Thereby, we used the same subset of 16 image classes and the same number of images as used by them. Table 1 summarizes our results.

Our method outperforms the compared methods in 13 out of 16 image collections. Figure 2 presents sample segmentations obtained by our approach. In these groups, the foreground regions of most images have homogeneous color and texture, which is a strong reason why our method yields very good accuracy results. The importance of Global Clustering is significant, since it groups similar foreground and background regions among the collections and identifies object regions that are not detected by methods only based on saliency techniques. The adoption of Global Clustering in consonance with saliency maps improves the overall results, complementing these methods.

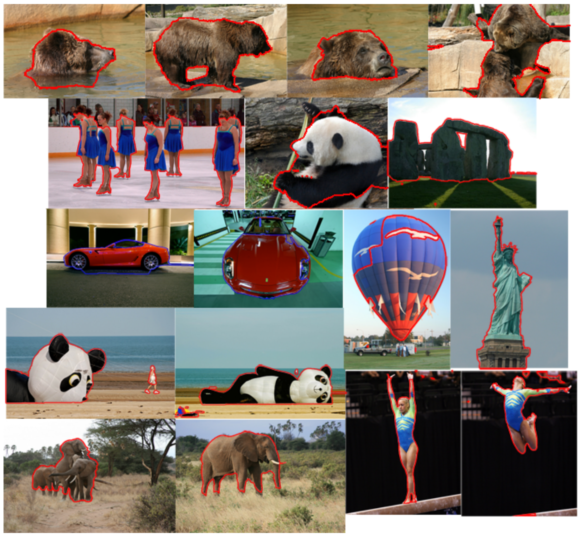

Fig. 2: Sample images of results obtained in the iCoseg dataset.

Our worst results were obtained by collections with salient images that do not adequately represent the foreground. In other words, even if many regions are global clustered perfectly, if these regions are not salient, then they will not be classified as object. That occurs in Taj Mahal and Kite collections. Parts of the background are included in their salient regions, which increases the probability of these being classified as object region. This is a current drawback of our method, which depends on saliency for object/background classification. However, this problem only occurs when the majority of the salient images in a given collection fails to detect the approximate correct regions corresponding to the objects of interest. In most cases, the saliency false positives and false negatives are overcome by the voting scheme.

Similarly to iCoseg, the MSRC dataset [15] was also introduced for supervised Cosegmentation. However, differently from iCoseg, the MSRC dataset contains images with object regions extremely variable, principally in their colors and texture features. That slightly reduces the accuracy of our results, as stated in Table 2. Nevertheless, our approach yields very competitive results in comparison with state-of-the-art methods, as depicted in Figure 3.

As it can be seen in the experiments, we achieve higher accuracy for many collections by considering the Global Clustering information. However, for some collections such as car, plane and bike, where the saliency is incorrectly computed for a large number of images, the accuracy is lower.

Table 2: Cosegmentation results obtained with the MSRC and Weizmann horses datasets by the proposed method. Bold numbers highlight the best method for each collection.

Class	Ours	[4]	[21]	[22]	[23]
Cars (Front)	85.7	83.6	87.7	65.9	**90.8**
Cars (Back)	73.8	74.5	85.1	52.4	**85.8**
Face	87.1	84.5	84.3	76.3	**87.3**
Cow	**92.3**	91.7	81.6	80.1	91.4
Horse	84.6	**87.6**	80.1	74.9	86.4
Cat	**87.6**	84.2	74.4	77.1	86.7
Plane	83.1	85.7	75.9	77.0	**87.7**
Bike	68.9	73.2	63.3	62.4	**76.8**

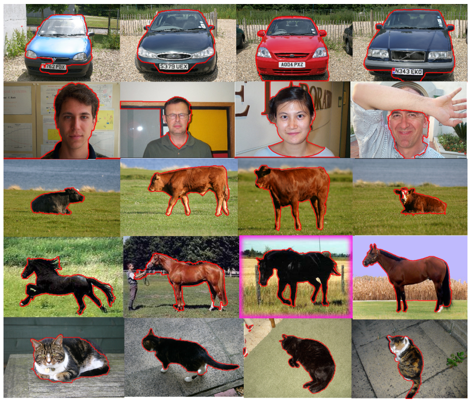

Fig. 3: Cosegmentation results obtained in the MSRC dataset by the proposed method.

4. CONCLUSION

We propose a fully unsupervised Image Cosegmentation model which is able to identify visually similar regions across several images and cosegment them in distinct classes. Our Global Clustering algorithm succeeds in detecting subregions that share similarities in their features among multi-images. Also, we create a robust approach to combine it with a saliency model, yielding better and comparable results with state-of-the-art algorithms.

We believe that our model can be extended in many ways: for instance, to deal with the segmentation of object classes, instead of object and background regions only. Also, we want to improve our results by using machine learning algorithms to learn the relevance of individual features. Different weights can be associated to the features according to their level of co-occurrence in the object class of the image collection.

5. REFERENCES

[1] Carsten Rother, Tom Minka, Andrew Blake, and Vladimir Kolmogorov, “Cosegmentation of image pairs by histogram matching - incorporating a global constraint into MRFs,” in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, Washington, DC, USA, 2006, CVPR ’06, pp. 993–1000, IEEE Computer Society.

[2] Sara Vicente, Carsten Rother, and Vladimir Kolmogorov, “Object cosegmentation,” in *CVPR*, June 2011.

[3] Y. Chai, V. Lempitsky, and A. Zisserman, “Bicos: A bi-level co-segmentation method for image classification,” in *IEEE International Conference on Computer Vision*, 2011.

[4] Hongkai Yu, Min Xian, and Xiaojun Qi, “Unsupervised co-segmentation based on a new global GMM constraint in MRF,” in *2014 IEEE International Conference on Image Processing, ICIP 2014, Paris, France, October 27-30, 2014*, 2014, pp. 4412–4416.

[5] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004.

[6] Christopher J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, June 1998.

[7] Edward Kim, Hongsheng Li, and Xiaolei Huang, “A hierarchical image clustering cosegmentation framework.,” in *CVPR*. 2012, pp. 686–693, IEEE.

[8] Huazhu Fu, Xiaochun Cao, and Zhuowen Tu, “Cluster-based co-saliency detection.,” *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3766–3778, 2013.

[9] Ming-Ming Cheng, Niloy J. Mitra, Xiaolei Huang, Philip H. S. Torr, and Shi-Min Hu, “Global contrast based salient region detection,” *IEEE TPAMI*, 2014.

[10] J. A. Hartigan and M. A. Wong, “A K-means clustering algorithm,” *Applied Statistics*, vol. 28, pp. 100–108, 1979.

[11] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[12] T. Kailath, “The divergence and bhattacharyya distance measures in signal selection,” *Communication Technology, IEEE Transactions on*, vol. 15, no. 1, pp. 52–60, February 1967.

[13] Yuri Y. Boykov and Marie-Pierre Jolly, “Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images,” 2001.

[14] Dhruv Batra, Carnegie Mellon Univerity, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen, “iCoseg: Interactive co-segmentation with intelligent scribble guidance,” in *In CVPR*, 2010.

[15] J. Shotton, J. Winn, C. Rother, and A. Criminisi, “Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation,” in *In ECCV*, 2006, pp. 1–15.

[16] Eran Borenstein, “Combining top-down and bottom-up segmentation,” in *In Proceedings IEEE workshop on Perceptual Organization in Computer Vision, CVPR*, 2004, p. 46.

[17] Naotoshi Seo, “Texture segmentation with gabor filters,” Tech. Rep. ENEE731 Project, Department of Computer Science, University of Maryland, USA, November 2006.

[18] T. K. Moon, “The expectation-maximization algorithm,” *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[19] João Carreira and Christian Sminchisescu, “CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts,” *IEEE TPAMI*, vol. 34, no. 7, pp. 1312–1328, 2012.

[20] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother, “Cosegmentation revisited: Models and optimization,” in *Proceedings of the 11th European Conference on Computer Vision: Part II*, Berlin, Heidelberg, 2010, ECCV’10, pp. 465–479, Springer-Verlag.

[21] A. Joulin, F. Bach, and J. Ponce, “Discriminative clustering for image co-segmentation,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[22] Jose C. Rubio, “Unsupervised co-segmentation through region matching,” in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, 2012, CVPR ’12, pp. 749–756, IEEE Computer Society.

[23] Kai-Yueh Chang, Tyng-Luh Liu, and Shang-Hong Lai, “From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model.,” in *CVPR*. 2011, pp. 2129–2136, IEEE.