

Original Research Paper

Unsupervised Cosegmentation Model based on Saliency Detection and Optimized Hue Saturation Value Features of Superpixels

*Rasha Shoitan and Mona M. Moussa

Department of Computer and Systems, Electronic Research Institute, Cairo, Egypt

Article history

Received: 14-06-2021

Revised: 05-07-2021

Accepted: 08-07-2021

Corresponding Author:

Rasha Shoitan

Department of Computer and Systems, Electronic Research Institute, Cairo, Egypt

Email: rasha.shoitan@eri.sci.eg

Abstract: Segmenting out the foreground from a single image remains a challenge in computer vision. Image co-segmentation has been used recently to alleviate the single image segmentation by exploiting the information of the common object to be segmented from a group of images having the same object. This research proposes an unsupervised co-segmentation technique based on saliency detection and optimized features of the histogram of Hue, Saturation and Value (HSV) of the superpixels. The proposed method is formulated as the conventional Markov Random Field (MRF) segmentation model with an added co segmentation constraint. First, an initial segmentation is extracted based on a saliency technique. Afterward, a Particle Swarm algorithm (PSO) is utilized to select, iteratively, some superpixels on the inner and outer boundary of the initial segmentation to be a foreground or a background according to the optimized energy function. PSO is guided by the HSV dominant colors of the image class and the superpixels around the foreground to decide if a certain superpixel is related to the foreground or the background. The proposed method is evaluated by two datasets: iCoseg and MSRC, along with comparisons to the results of using ten conventional methods based on the Intersection over Union (IoU) metric. The experimental results demonstrate that the proposed method can segment the object successfully and accurately more than the traditional co-segmentation methods, even with a cluttered background.

Keywords: Cosegmentation, Saliency, HSV, SLIC, MRF

Introduction

In an era where digital cameras are increasing rapidly, many images need to be effectively utilized by expert and artificial intelligent systems. Extracting the objects from these images is beneficial in understating and analyzing their contents. Indeed, segmenting the objects from an image is crucial for many multimedia and computer vision applications, such as skeletonization, image classification, action recognition, scene analysis and image retrieval (Merdassi *et al.*, 2020). However, it is hard to automatically segment the foreground object from a single image because of the lack of information for this object. Recently, researchers attempted to cope with this lack of information and increase the segmentation accuracy of the objects by exploiting the details of common objects that exist in a set of similar images. The process of segmenting out the common objects through simultaneously processing the collection of similar images is known as co-

segmentation. In the last years, different cosegmentation methods are suggested segmenting the common object. Some of these methods use the traditional segmentation models and add a certain foreground consistency constraint to segment the common object in a pair of images (Rother *et al.*, 2006; Hochbaum and Singh, 2009; Mukherjee and Dyer, 2011). Then, the idea is broadened to segment the common object in a group of images (Mukherjee and Peng, 2011; Meng *et al.*, 2016). However, these methods still suffer from the lack of segmentation accuracy. In this study, an unsupervised co-segmentation method based on the traditional MRF segmentation model and PSO is proposed to improve the segmentation accuracy of the segmented objects in a group of images. The main purpose of the proposed method is to use the PSO for the first time to solve the proposed MRF co-segmentation model. Also, a new foreground consistency term based on the histogram of HSV color space is proposed to improve the accuracy of the segmentation

results. The proposed algorithm utilizes a saliency detection technique (Borji *et al.*, 2019) to detect the initial foreground for each image. Afterward, the superpixels around the contour of the initial segmented foreground are determined. The PSO is responsible for achieving the best co-segmentation by flipping the state of the superpixels around the contour to a foreground or a background then evaluating the energy function. The energy function starts with the initial foreground to achieve a consistent segmentation through two fields; (1) within the image, (2) within the class (intra-class), which is the co-segmentation term. The proposed co-segmentation term in the energy function compares the dominant colors of the initial foregrounds from all the images in the same class and the dominant colors of the selected superpixel. Based on this comparison, a weight is added or subtracted from the energy function. According to the energy function value, PSO decides the best state of a superpixel to be a foreground or a background. The four main contributions in the proposed method are:

- 1) Exploiting the fast convergence and the good global search capability advantages of the PSO to optimize the energy function
- 2) Using HSV color domain instead of RGB to determine the dominant colors leads to better separation of the color information from the intensity
- 3) Facilitating the comparison between the dominant colors by converting them to a binary matrix
- 3) Searching for the superpixels on the foreground contour instead of searching in the whole image makes it easier to reach the best solution

Literature Review

Object co-segmentation was first introduced by Rother *et al.* (2006) to demonstrate that segmenting the common objects from a pair of images improves the segmentation accuracy of the object than using one single image. From that time, co-segmentation has drawn much attention and different methods are introduced (Merdassi *et al.*, 2020). Some of these methods describe the co-segmentation problem as an energy function of the conventional segmentation model, adding the foreground consistency constraint term. Therefore, the co-segmentation problem is solved by minimizing this energy function that enforces the foreground consistency (Meng *et al.*, 2016). Rother *et al.* (2006) formulate the energy minimization function as an MRF-based segmentation with the L_1 -norm of the histograms of the common foreground objects to force their consistency. This co-segmentation problem is solved by the Trust-Region Graph Cuts optimization method (TRGC). Mukherjee and Dyer (2011) used L_2 norm instead of L_1 for the foreground consistency term to relax

the energy minimization problem to linear programming and solve it using Pseudo Boolean optimization. Unfortunately, the existence of the histogram difference term in both types of research increases the complexity of the optimization problem. Therefore, Hochbaum and Singh (2009) simplify the energy minimization function by rewarding the foreground consistency term and optimizing it using the max-flow algorithm.

The previously mentioned methods are applied to pair images and have restricted applications. Thus, the co-segmentation methods have been generalized to be applied to a group of images to extract common objects for practical applications. Batra *et al.* (2010; 2011) have scaled the co-segmentation for multiple images by introducing an interactive co-segmentation method. This method depends on an action from the user to add scribbles on an image to discriminate between foreground and background then the Gaussian mixture model and the Grabcut are used to co-segment these images. Collins *et al.* (2012) propose an interactive co-segmentation framework based on a random walker model to add constraints information for foreground regions. However, the co-segmentation results are affected by the position and the size of these scribbles. Also, Lee *et al.* (2015) deduce that a random walker is insufficient to segment the images accurately because it depends on the single random walker. Therefore, they propose a graph-based system to analyze Multiple Random Walkers (MRW) motions and relations to improve co-segmentation performance.

Mukherjee and Peng (2011), the author has broadened the MRF co-segmentation methods to multiple images by modifying the histogram term and considering the foreground scale variations. Meng *et al.* (2016) address the problem of Co-segmentation of Multiple Groups of images (CMG) and formulate it as an energy minimization function consisting of three terms: The single image segmentation term, the single group term and the multiple group term, then the Expectation-Maximization algorithm (EM) is utilized to solve this optimization problem. On the other hand, Joulin *et al.* (2010) propose a co-segmentation method based on discriminative clustering to assign the labels jointly for the common foregrounds in a group of images. Chang *et al.* (2011) use co-saliency to detect the foreground locations and utilize this initial foreground as prior knowledge in the energy function. This energy function is solved using the graph cut technique.

Saliency has been widely used in different ways to solve the co-segmentation problem. Jerripothula *et al.* (2014) utilize the Geometric Mean Saliency (GMS) technique to form a global saliency map by fusing the aligned saliency maps of a group of similar images, then segmenting out the single image using this global saliency

map. Because the process is repeated for each image, the GMS method suffers from excessive calculations for large scale datasets. Furthermore, Jerripothula *et al.* (2014) propose an image co-segmentation method based on Saliency Co-Fusion (SCF) to increase the robustness of the co-segmentation system. The authors apply different saliency detection techniques on each image to improve the joint object and get multiple saliency maps. These saliency maps are weighted and fused at the superpixel level and the resultant saliency map is used to implement single segmentation on each image. Afterward, Jerripothula *et al.* (2015) solve the GMS technique issue of using the Group Saliency Propagation (GSP) method for performing the co-segmentation. The basic idea of this method is to cluster similar images into groups and choose a key image that represents each group. This method reduces the transferred information between images and decreases the calculations (Lu *et al.*, 2019).

Moreover, Meng *et al.* (2012) propose a co-segmentation method using the salient information and the Shortest Path Algorithm (SPA) to segment similar objects with different colors. First, the local regions are extracted from each image then a digraph is built by the similar local regions and the saliency map. The problem is constructed as the shortest path problem and solved by a dynamic programming method. Other authors suggest the idea of joint processing between various visual tasks to support each other with useful information such as Dai *et al.* (2013) propose an energy function that combines the co-sketch and the co-segmentation for aligning the similar objects in a group of images to improve the segmentation results. However, Jerripothula *et al.* (2017) propose a framework that couples the co-skeletonization and the co-segmentation to use the interdependencies between them to support each other.

Faktor and Irani (2013) propose a co-segmentation method by composition. The algorithm measures the overlapping degree between the co-occurring regions and the initial segments, then constructs a co-segment map reflecting the score of overlapping for each pixel. Finally, the Grabcut is applied to this score map to obtain the segmentation output. Recently, Convolutional Neural Networks (CNNs) have been used in the co-segmentation. Kamranian *et al.* (2018) propose CNN-based Feature Visualization (CNN-FV) for detecting the foreground and use this information in the energy minimization function to improve the co-segmentation results. Gong *et al.* (2020) propose a co-segmentation technique that calculates the visual correlation between images based on the co-attention computation block and Siamese

network. Li and Liu (2021) propose a new IS-Triplet loss and merge it with conventional image segmentation loss for the co-segmentation application.

The Proposed Co-segmentation Method

This section presents a detailed description of the proposed co-segmentation method. As mentioned in the introduction section, the proposed technique formulated the problem as an MRF based segmentation model with a further term for co-segmentation, where the optimization is achieved using PSO. In the proposed method, a saliency object detection technique is used to extract an initial foreground in the images and then PSO selects superpixels around the contour of its initial foreground. Finally, the PSO optimizes the energy function and decides if each superpixel is related to the foreground or the background according to the co-segmentation information obtained from the images group. The detailed information of each process is discussed in the following subsections:

Salient Detection Method

Salient object detection is used in the proposed method to recognize and segment out the most visually attractive objects or regions from the background in an image. The advantage of using the salient object detection method in co-segmentation is that it concentrates only on distinctive and interesting objects.

Therefore, it is used in the proposed approach to extract an initial foreground to be used as prior information in the energy minimization function. A saliency detection method based on the deep learning method mentioned in (Borji *et al.*, 2019) is used in the proposed technique because of its ability to locate the most salient regions accurately.

Superpixels on the Contour

Superpixels are used in the proposed technique instead of pixel level for many advantages. First, superpixels contain additional information than pixels since each superpixel includes the pixels that have the same visual properties, so this superpixel has a perceptual meaning. Second, the superpixels arrange the pixels in a compact form, which is beneficial for large computational tasks. Third, superpixels abide by the image boundaries when clustering the pixels, which are essential in image segmentation (Achanta *et al.*, 2012; L  zoray *et al.*, 2017; Stutz *et al.*, 2018). In this research, the Simple Linear Iterative Clustering (SLIC) algorithm (Achanta *et al.*, 2012; Stutz *et al.*, 2018) is used as a preprocessing stage to over segment the images into many superpixels. The proposed technique exploits the adherence to the image boundaries of the superpixels and selects only the

superpixels around the initially segmented foreground. This step directs the PSO search for the correct places and reduces the time to search in the whole image, as shown in Fig. 1. Each particle in the PSO is represented as a vector; its length equal to the number of the selected superpixels (n) around the initial foreground contour, as shown in Eq. 1. Each particle is initialized by the status of the selected superpixel (foreground 'FG' or background 'BG'):

$$Particle = [FG\ BG\ BG\ \dots\ FG_n] \quad (1)$$

In each iteration, each particle selects one of the superpixels to flip its status (foreground/background) and check the effect on the energy minimization function. Afterward, PSO decides whether to flip one of the selected superpixels based on the calculated fitness function value. Then the superpixels on the boundary are updated to be used in the next iteration.

Dominant Color

The cosegmentation term in the energy minimization function is considered the most crucial part of any cosegmentation technique, as this term affects the accuracy of the segmentation and the complexity of the energy minimization function. Most previous studies use the RGB color domain and some features in the cosegmentation term to segment the common object (Rother *et al.*, 2006; Hochbaum and Singh, 2009; Mukherjee and Dyer, 2011). However, RGB color space is not recommended for color analysis because RGB cannot differentiate between color and luminance (intensity). Therefore, in the proposed algorithm, HSV color space discriminates between intensity and color and considers the color characteristic (Shaik *et al.*, 2015; Garcia-Lamont *et al.*, 2018). The HSV color space consists of three components which are the Hue (H), the Saturation (S) and the Value (V). The Hue term represents the pure color (yellow, orange, blue, etc.). Saturation term measures the degree of mixing the Hue with white color. However, the Value term measures the intensity or the brightness of the color, as shown in Fig. 2 (Bourne, 2010).

For determining the dominant colors in the proposed technique, all the images from the same class are segmented using the salient object detection technique. The segmented foregrounds are then converted to HSV color space. Subsequently, the 2-D histogram is drawn between the Hue and the Saturation terms for each segmented foreground. The vertical axis of the histogram represents the Hue, while the horizontal axis represents the Saturation. The histogram presents the

color frequency distribution for each image foreground on the Hue and Saturation terms. As shown in Fig. 3, a group of images from the same class is used to get the salient regions and then a histogram is created using the Hue and Saturation values of these regions. It can be noticed that there are a lot of yellow points in the histograms of the images which reflect that these locations have non-zero elements (colors). Also, it can be observed that most of the yellow points are concentrated on the top right corner of the histogram, which means that the color is red (Fig. 2). The histograms generated from all segmented foregrounds are summed and a new one is created. This summed histogram represents all the colors of the images in this class. The dominant color is determined from the summed histogram by arranging its values (colors) in a descending order, where the first color is the most frequent color in the summed histogram (the highest number of pixels). According to the shape of the histogram (wide or narrow), a percentage from the arranged colors is considered dominant. The histogram shape is determined based on a ratio of non-zero elements in the histogram. This ratio is calculated as:

$$Ratio = \frac{\#of\ nonzeros\ in\ summed\ histogram}{\#\ of\ histogram\ bins} \quad (2)$$

If the ratio is small, the histogram is narrow and a high percentage of the colors is taken as dominant colors. However, if the ratio is high, the histogram is wide and a low percentage from the colors is taken as dominant colors. Thus, dominant colors are determined to help PSO choose the state of a given super pixel (foreground or background) by comparing the super pixel colors and the dominant colors of the image class. In the proposed method, this comparison is facilitated by converting the histogram matrix that contains the dominant colors to a binary matrix according to the following equation:

$$binary_matrix(i, j) = \begin{cases} 1 & \text{if } histogram(i, j) > 0 \\ 0 & \text{else} \end{cases} \quad (3)$$

The 1's in the binary matrix reflects the dominant colors that are represented using yellow points, as shown in Fig. 3.

Energy Minimization Function

In the proposed method, the common objects are segmented out using the conventional MRF-based segmentation model with a constraint that reflects the foreground consistency. This energy function is formulated as:

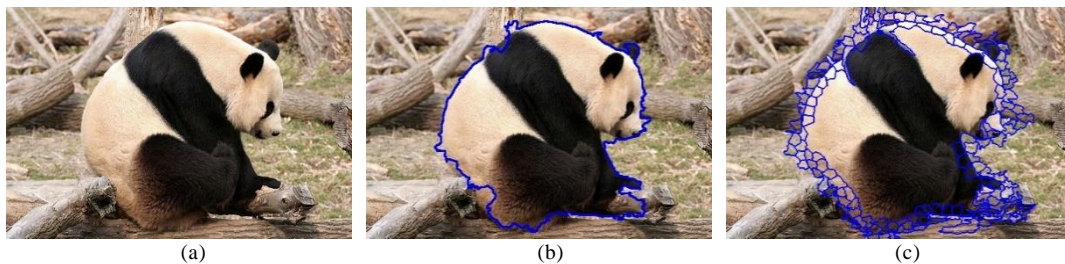


Fig. 1: (a) The original image to be segmented, (b) the segmented foreground at a certain iteration of PSO and (c) the superpixels on the contour of the foreground

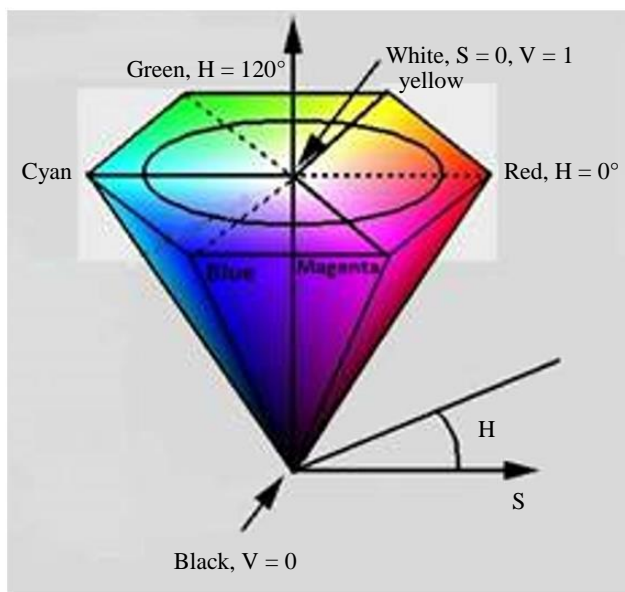


Fig. 2: HSV color model (Bourne, 2010)

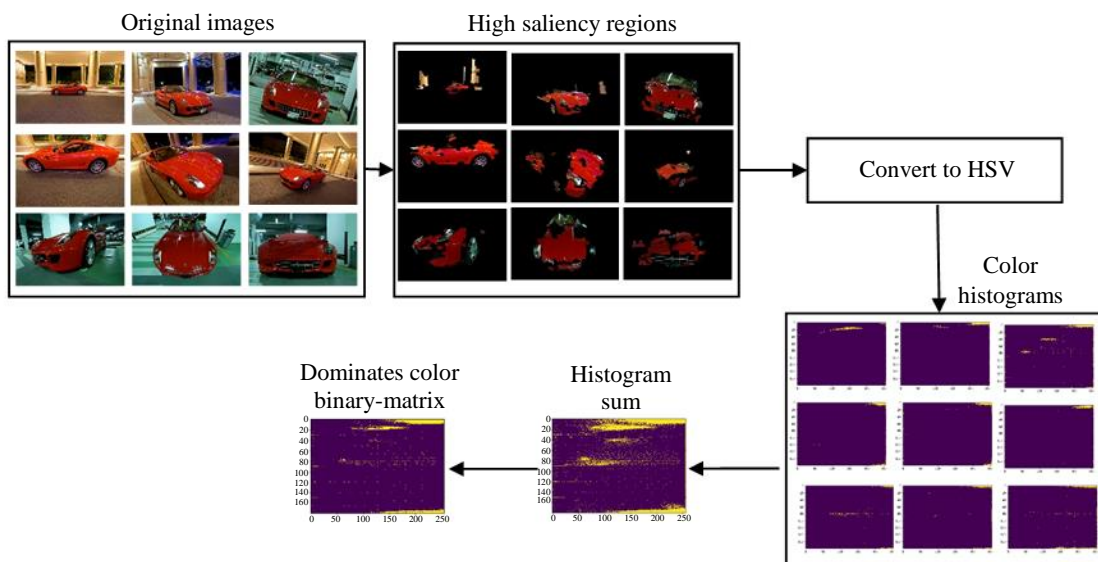


Fig. 3: Dominant color determination

$$E = E_{MRF} + E_{Coseg} \quad (4)$$

where, E_{MRF} is the segmentation term for a single image that assures the smoothness and differentiates between foreground and background in an image, while the E_{Coseg} is the cosegmentation term that enforces the consistency between the foregrounds in a group of images of the same class (intra-class) (Popova *et al.*, 2018).

The MRF term is responsible for assigning a binary label for each pixel to be a foreground or background according to the following equation:

$$E_{MRF} = E_u + E_p \quad (5)$$

E_u is the unary potential term encoding the probability that a pixel is relevant to foreground or background while E_p is the pairwise potential or smoothness term that reward assigning the same label to adjacent pixels with the same color feature.

In the proposed method, the cosegmentation term E_{coseg} is constructed according to a comparison between the superpixels selected by the PSO and the image group based on the dominant colors. Relative to their degree of similarity, a weight is added or subtracted from the energy function to help PSO to decide if the superpixel is a foreground or background. This similarity is measured by a Similarity Ratio (SR) which is defined as the intersection between the binary matrix of the class and the binary matrix of the selected superpixel as in the following equation:

$$SR = \frac{\text{binary_matrix}(\text{the class}) \cap \text{binary_matrix}(\text{selected superpixel})}{\text{binary_matrix}(\text{selected superpixel})} \quad (6)$$

A higher value of SR means that the dominant colors of superpixel are similar to the dominant colors of the class. However, a lower value means that superpixel colors are far distinct from the dominant colors of the class. Based on the SR value and the state of the superpixel (Foreground 'FG' or Background 'BG') suggested by PSO, a weight is added or subtracted from the fitness function. The following equation summarizes the weight calculation method:

$$\text{weight} = \begin{cases} -1 & \text{if } SR > th_1 \text{ and superpixel to be FG} \\ +1 & \text{if } SR > th_1 \text{ and superpixel to be BG} \\ -1 & \text{if } SR < th_2 \text{ and superpixel to be BG} \\ +1 & \text{if } SR < th_2 \text{ and superpixel to be FG} \\ 0 & \text{Otherwise} \end{cases} \quad (7)$$

where th_1 and th_2 are determined empirically to decide if the SR value is considered high or low. In the first two cases, SR has a high value, so if PSO suggests

flipping the superpixel to foreground, then rewards this suggestion; otherwise, penalties it. However, in the second pair of cases, SR has a small value, so if PSO suggests flipping the superpixel to the background, then rewards this suggestion otherwise penalties it. In the last case, the SR value is between th_1 and th_2 , which means that the superpixel's colors are neutral (neither dominant nor distinct), so the weight is zero. Thus, the final energy minimization equation is:

$$E = E_{MRF} + \text{weight} \quad (8)$$

A block diagram of the proposed algorithm is presented in Fig. 4 and the proposed algorithm steps are given below:

- Step 1:** Determine the saliency region (initial foreground) of the image to be segmented.
- Step 2:** Determine the saliency region of all the classes.
- Step 3:** SLIC is applied to the image class to generate superpixels for each image.
- Step 4:** Convert the segmented foregrounds from RGB to HSV colour space and then generate the histogram for each converted foreground.
- Step 5:** Sum the histograms of all the foregrounds in one histogram then determine the dominant colour as mentioned in the dominant colour subsection.
- Step 6:** Determine the superpixels on the foreground contour for the image to be segmented.
- Step 7:** For each iteration, the selected superpixels around the contour, the foreground and the dominant colour of the image class are used as an input for the PSO.
 - Step 7.1:** Calculate the HSV histogram of the selected superpixel.
 - Step 7.2:** Determine the dominant colours of the superpixel as mentioned in the dominant colour subsection.
 - Step 7.3:** Calculate the SR between the dominant colours of the image class in step 5 and the dominant colour of the selected superpixel in step 7.2 using Eq. 6
 - Step 7.4:** Determine the value of the weight which will be added to the energy minimization function using the SR value as in equation 7
 - Step 7.5:** PSO evaluates the energy function, then if a global best is reached then PSO flips the superpixel status.
 - Step 7.6:** Update the foreground and determine the superpixels around the updated foreground
- Step 8:** If the termination criterion is not met go to step 7

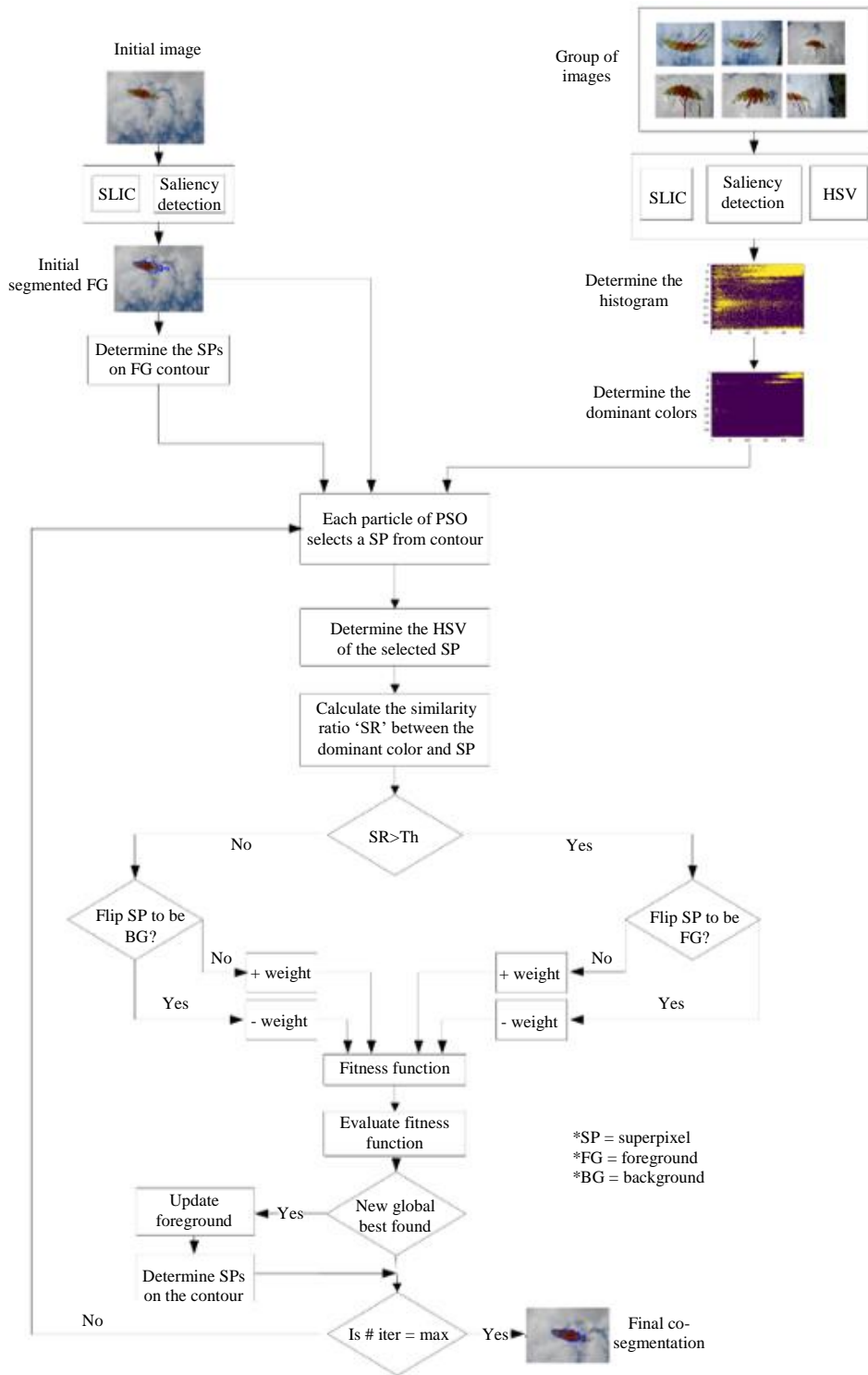


Fig. 4: Proposed co-segmentation algorithm flowchart

Experimental Results and Analysis

The performance of the proposed approach is evaluated by conducting different experiments on two

commonly benchmark cosegmentation datasets which are iCoseg (Batra *et al.*, 2010) and MSRC (Shotton *et al.*, 2006). The iCoseg dataset has 643 images. These images are divided into 38 classes. However, the MSRC dataset is

composed of 233 images of 8 object classes. In these datasets, the images in each class have the same object with different scales, viewpoints and illumination.

The parameters of PSO are initialized as maximum number of iterations = 100, number of particles = 5, lower decision bound = 0, upper decision bound = 10, personal acceleration coefficient1 = 1.4962, social acceleration coefficient 2 = 1.4962, maximum velocity = 6 and inertia coefficient = 0.2.

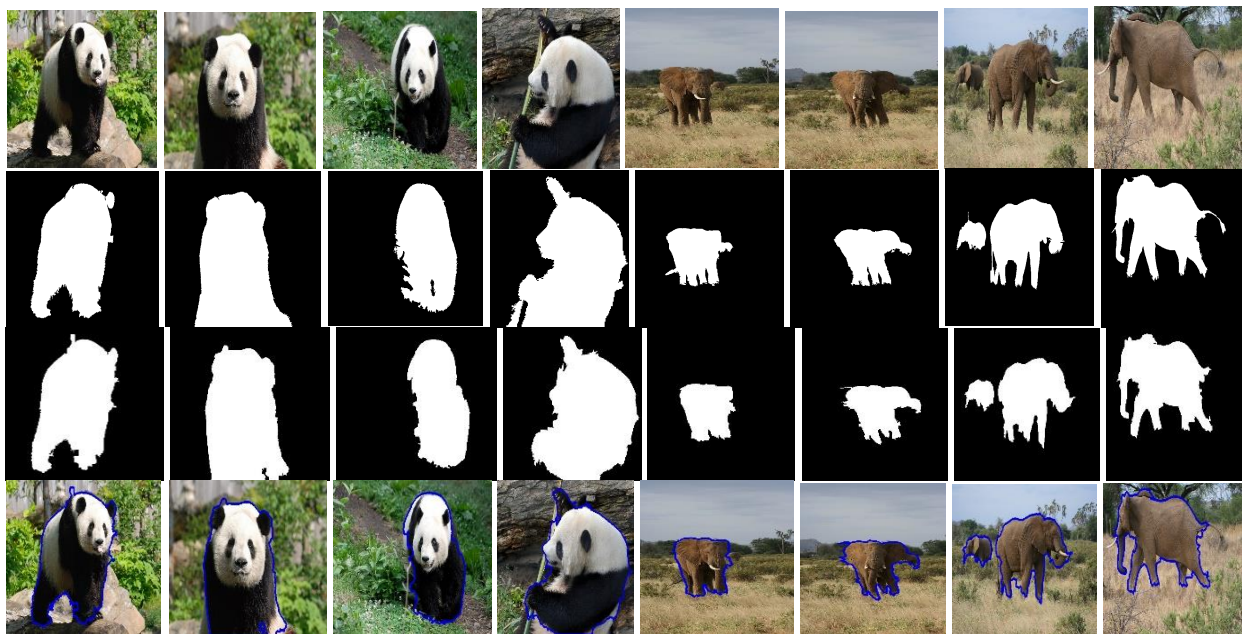
The performance of the proposed method is evaluated and compared with different previously proposed techniques (Meng *et al.*, 2012; 2016; Dai *et al.*, 2013; Faktor and Irani, 2013; Jerripothula *et al.*, 2014; 2016; 2017; Jerripothula *et al.*, 2015; Lee *et al.*, 2015; Kamranian *et al.*, 2018) using IoU metric. This metric is calculated as the intersection area of the segmented foreground and the ground truth to their union. The results of this comparison on iCoseg and MSRC in terms of IoU are tabulated in Table 1 and Table 2, respectively. In both tables, the average of the IoU is calculated for each class because each dataset has a massive number of images. It will be noticed that some results of methods (Meng *et al.*, 2016; Kamranian *et al.*, 2018) are missed; this is because the results are gotten from their original paper. However, the results of the other methods are obtained from (Lu *et al.*, 2019).

It can be perceived from the results presented in Table 1 that the proposed method outperforms the three saliency methods which are GMS, SCF and GSP. Although the SCF has applied different saliency techniques for each image then fuse between them to improve the performance of the cosegmentation technique, the proposed method succeeds in improving the results without the need for

extra processes. Also, it can be noticed from Table 1 that the proposed method gives the best performance for different challenging classes over the ten conventional techniques, such as ‘Red sox players’, ‘Stonehenge 1’, ‘Liverpool FC players’, ‘Ferrari’, ‘Taj Mahal’, ‘Elephants’, ‘Airshows-planes’, ‘Airshows-Huntsville’, ‘Gymantics-2’, ‘Skating-3’, ‘Soccer players’ and ‘Track and field’. On the other hand, the proposed method results for the rest of the classes are still competitive. Overall, the proposed technique achieves the highest success occurrences, which are 14 classes out of 38 classes.

Meanwhile, the best-performing conventional method, CMG, achieves 7 out of 38 classes. Also, the average IoU is calculated for all the classes for each technique and the results are presented in the last row of Table 1. It can be found out that the proposed method accomplishes the highest average of IoU which is 0.7536, over the other conventional techniques on the iCoseg dataset.

For MSRC dataset results in Table 2, it can be observed that none of the techniques performed well for all the classes and the results are a little bit lower than iCoseg. The first reason behind this is that the MSRC is considered more complicated than iCoseg where the same class can have different individuals with varied colors and shapes of the same category (as the animal class has cows, cheeps, horses, etc.). The second reason is that the MSRC ground truth is not accurate, as shown in the second and sixth rows of Fig. 6. Despite the problems in using MSRC dataset, the proposed method attains the highest success occurrences, which are 4 classes out of 8 classes compared to the other conventional methods. Also, the average IoU for the proposed method is still competitive.



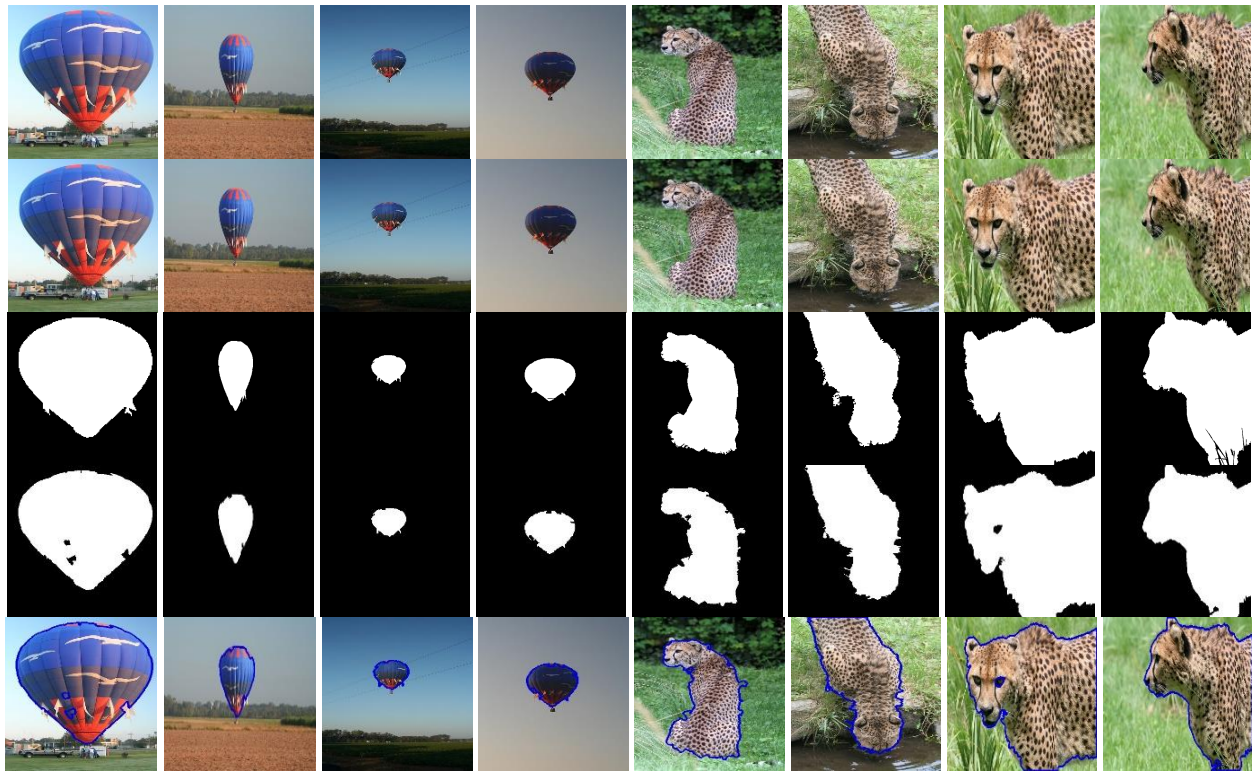
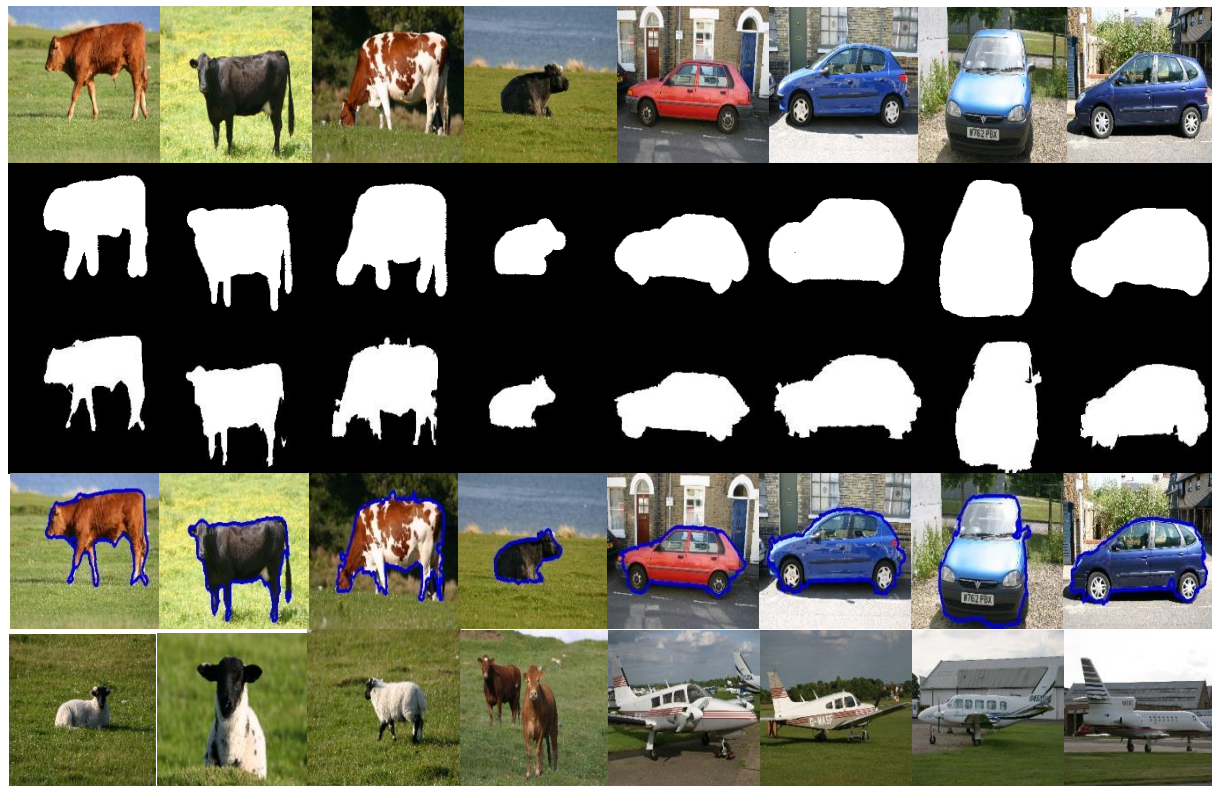


Fig. 5: The proposed cosegmentation technique results for some classes of iCoseg dataset Row1: Original images Row 2: Ground truth Row 3: Thresholded foreground Row 4: segmented foreground



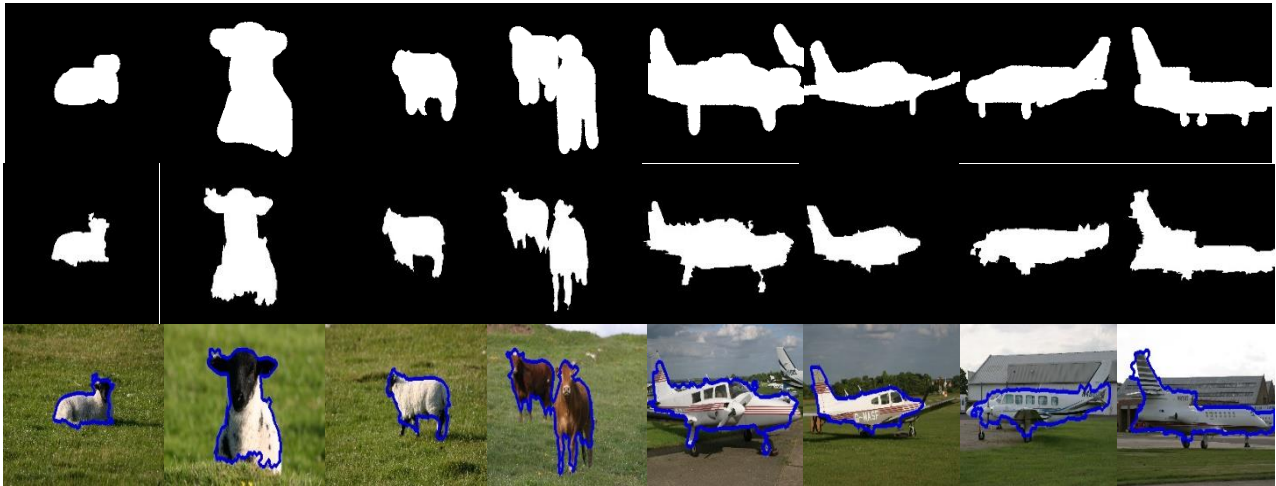


Fig. 6: The proposed cosegmentation technique results for some classes of MSRC dataset Row 1: Original images Row 2: Ground truth Row 3: The proposed thresholded foreground Row 4: The proposed segmented foreground

Table 1: The performance of the proposed technique compared to the conventional techniques based on the IoU metric on iCoseg

Class	#image	GMS	GSP	SCF	CST	CSZ	MRW	SPA	COMP	CMG	CNN-FV	Prop.
Brown bear 1	19	0.6421	0.665	0.6004	0.4284	0.6357	0.6658	0.5239	0.664	0.9	0.760	0.8337
Brown bear 2	5	0.7218	0.752	0.7058	0.6326	0.7607	0.7622	0.4193	0.849	0.75	-	0.8048
Redsox players	25	0.6411	0.6902	0.7246	0.1276	0.6773	0.7099	0.2865	0.6993	0.69	0.61	0.7639
Stonehenge 1	5	0.6245	0.5936	0.6354	0.4404	0.5924	0.7232	0.0303	0.7375	0.81	0.734	0.8378
Stonehenge 2	18	0.7633	0.7871	0.7183	0.813	0.7416	0.7832	0.2159	0.6794	0.56	-	0.6742
Liverpol FC	33	0.4708	0.4581	0.5466	0.4366	0.549	0.5292	0.4491	0.4666	0.57	-	0.6128
Ferrari	11	0.6555	0.652	0.6691	0.6364	0.6376	0.7417	0.5273	0.6697	0.69	0.724	0.8085
Taj Mahal 1	5	0.5446	0.5062	0.5278	0.4656	0.5585	0.7449	0.5674	0.5737	0.5	0.667	0.7662
Taj Mahal 2	5	0.4284	0.4817	0.4792	0.3716	0.3592	0.4664	0.2716	0.3734	0.6	-	0.7982
Pyramids	10	0.7343	0.6342	0.5449	0.4061	0.5918	0.5849	0.2469	0.6271	0.69	-	0.4280
Elephants	15	0.6807	0.7241	0.6517	0.3087	0.7336	0.7034	0.4229	0.5106	0.8	0.743	0.8854
Goose	31	0.6681	0.6901	0.5231	0.3959	0.6962	0.7888	0.6157	0.8419	0.78	0.812	0.7717
Pandas 1	25	0.7367	0.7165	0.7122	0.7832	0.6219	0.6956	0.4283	0.6952	0.77	0.767	0.7546
Pandas 2	21	0.6412	0.6216	0.5749	0.6676	0.5064	0.7269	0.6974	0.2779	0.76	0.766	0.6990
Helicopter	12	0.7841	0.8034	0.7774	0.6908	0.8236	0.796	0.8839	0.7387	0.81	-	0.8260
Planes	39	0.5049	0.5148	0.5634	0.5791	0.5049	0.5713	0.2968	0.6089	0.59	-	0.6841
Huntsville	22	0.3915	0.4162	0.5237	0.2646	0.5172	0.7757	0.5476	0.5816	0.53	-	0.8579
Cheetah	33	0.7552	0.7647	0.7357	0.7719	0.7152	0.7984	0.5463	0.22	0.8	0.787	0.7666
Kite-1	18	0.7378	0.6915	0.8172	0.294	0.8142	0.8503	0.7758	0.8911	0.71	-	0.7129
Kite-2	10	0.5447	0.4645	0.5596	0.5347	0.5892	0.4687	0.239	0.6302	0.71	-	0.4609
Kite-3	7	0.7625	0.7151	0.8322	0.5681	0.7647	0.8823	0.4869	0.9591	0.75	-	0.7519
Kite-4	11	0.6831	0.6923	0.7048	0.3859	0.7234	0.8502	0.7359	0.6747	0.6	-	0.8225
Gymantics-1	6	0.8333	0.8396	0.8359	0.8693	0.8374	0.7852	0.7892	0.5174	0.89	-	0.8392
Gymantics-2	4	0.6487	0.6659	0.7657	0.3937	0.6757	0.7685	0.8086	0.7907	0.81	-	0.8593
Gymantics-3	6	0.6224	0.6203	0.8131	0.7512	0.7654	0.7581	0.803	0.7944	0.86	-	0.8300
Skating-1	11	0.6759	0.7007	0.5889	0.5888	0.6751	0.838	0.3554	0.5928	0.68	-	0.7065
Skating-2	12	0.8461	0.8113	0.8602	0.1177	0.9139	0.9292	0.8052	0.8445	0.85	-	0.8409
Skating-3	13	0.2233	0.3038	0.4509	0.1136	0.4158	0.1096	0.174	0.1584	0.53	-	0.7025
Soccer players-1	36	0.6552	0.648	0.668	0.6787	0.6619	0.6399	0.5813	0.5462	0.68	-	0.6987
Soccer players-2	16	0.504	0.5249	0.4684	0.4633	0.5062	0.5005	0.2847	0.4154	0.5	-	0.5802
Monks	17	0.7017	0.72	0.762	0.7048	0.7134	0.7627	0.5858	0.5802	0.77	-	0.7455
Hot Balloons	24	0.6549	0.6348	0.7516	0.5038	0.7421	0.8493	0.5536	0.9378	0.63	-	0.9300
Statue of liberty	41	0.7892	0.7772	0.7439	0.6764	0.8234	0.7741	0.8426	0.6997	0.72	0.865	0.8098
Christ the Redeemer	13	0.7275	0.6968	0.7122	0.6397	0.7813	0.8388	0.6897	0.7089	0.73	0.783	0.7336
Track and field	5	0.5066	0.5786	0.4897	0.3671	0.4193	0.5457	0.2839	0.3271	0.48	-	0.7902
Windmill	18	0.3145	0.317	0.4462	0.1441	0.3386	0.2723	0.4044	0.2521	0.58	-	0.5613
Kendo-1	30	0.7835	0.8297	0.8219	0.151	0.9022	0.921	0.6885	0.8051	0.91	-	0.8196
Kendo-2	11	0.8725	0.9127	0.8483	0.449	0.8922	0.9478	0.6732	0.8749	0.94	-	0.8690
Success Occurrences	1/38	1/38	0/38	1/38	0/38	0/38	6/38	1/38	5/38	7/38	-	14/38
Average		0.6441	0.6478	0.6620	0.4899	0.6626	0.7068	0.5142	0.6267	0.7087	-	0.7536

Table 2: The performance of the proposed technique compared to the conventional techniques based on the IoU metric on MSRC

Class	#image	GMS	GSP	SCF	CST	CSZ	MRW	SPA	COMP	CMG	CNN -FV	Prop.
Animal	23	0.6285	0.5826	0.6194	0.6291	0.6502	0.6606	0.6039	0.5863	-	-	0.7308
Tree	30	0.7567	0.7583	0.6889	0.7302	0.6798	0.7374	0.6623	0.6637	-	-	0.2890
Building	30	0.7762	0.7117	0.7099	0.7487	0.6834	0.6909	0.497	0.6282	-	-	0.5860
Plane	30	0.5418	0.5416	0.5507	0.4939	0.5309	0.5089	0.4992	0.4802	-	-	0.7304
Cow	30	0.7816	0.7681	0.7432	0.7623	0.7678	0.7399	0.7237	0.7105	-	0.842	0.7707
Face	30	0.6172	0.6303	0.6214	0.583	0.5921	0.4996	0.3592	0.5654	-	0.655	0.6648
Car	30	0.7126	0.6922	0.6672	0.6953	0.579	0.6782	0.6089	0.5681	-	0.712	0.7328
Bike	30	0.4282	0.4096	0.4691	0.5243	0.3178	0.528	0.4384	0.4906	-	-	0.3963
Success Occurrences	1/8	1/8	0/8	0/8	0/8	0/8	1/8	0/8	0/8	-	-	4/8
Average		0.6554	0.6368	0.6337	0.6459	0.6001	0.6304	0.5491	0.5866	-	-	0.6126

Visual results of the proposed method on different image classes of iCoseg and MSRC are shown in Fig. 5 and Fig. 6, respectively. Figure 5 shows the proposed segmentation result of four images classes from iCoseg dataset, which are ‘Panda’, ‘Elephant’, ‘Hot balloons’ and ‘Cheetah’. It can be seen from the figure that the proposed method segments out the common object accurately even with a cluttered background. On the other hand, Fig. 6 shows the results of the proposed method of four image classes which are ‘Cow’, ‘Car’, ‘Animal’ and ‘Plane’, from MSRC dataset. It can be seen from the figure, that the binary foreground of the proposed method is sharper and more accurate than the ground truth, which leads to the common object being localized and segmented accurately.

Conclusion

In this study, an unsupervised co-segmentation technique is proposed to extract the common objects from a group of images. The proposed method exploits the MRF based segmentation method and adds to it a co-segmentation term. PSO is used in the proposed co-segmentation technique, for the first time as far as the authors know, to optimize the energy function method. The saliency technique is used to create an initial segmentation for each image in the class. On the other hand, the HSV color system is used to determine the dominant colors for the segmented images. The superpixels around the initial foreground to be segmented are selected and assigned to the PSO particles. Selecting the superpixel around the foreground contour supports PSO to concentrate on the region of interest and reduce the computation time. The dominant colors of the image class and the dominant colors of the selected superpixels are compared. Their similarity ratio is added as a weight to the energy function for assisting the PSO to efficiently discriminate if a superpixel is related to the background or the foreground. The experiments have been conducted on two benchmark image co-segmentation datasets and have achieved high accuracy compared to other research in this field. The results prove that using the dominant colors for foreground consistency term and optimizing the function using PSO improves the co-segmentation results even with an interfered background.

Author’s Contributions

Rasha Shoitan: Contributed to define the research problem, formulation of methods and writing the paper.

Mona M. Moussa: Assisted in the implementation of methods and conduct of the experiment.

Ethics

The authors verify that the research paper is original and have unpublished data. The corresponding author confirms that the other authors have read and approved the manuscript and no ethical issues involved.

References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11), 2274-2282.
<https://ieeexplore.ieee.org/abstract/document/6205760>
- Batra, D., Kowdle, A., Parikh, D., Luo, J., & Chen, T. (2010, June). Icoseg: interactive co-segmentation with intelligent scribble guidance. In *2010 IEEE Computer Society Conference On Computer Vision and Pattern Recognition (PP. 3169-3176)*. IEEE.
<https://ieeexplore.ieee.org/abstract/document/5540080>
- Batra, D., Kowdle, A., Parikh, D., Luo, J., & Chen, T. (2011). Interactively co-segmenting topically related images with intelligent scribble guidance. *International Journal of Computer Vision*, 93(3), 273-292.
<https://link.springer.com/article/10.1007/S11263-010-0415-X>
- Borji, A., Cheng, M. M., Hou, Q., Jiang, H., & Li, J. (2019). Salient object detection: a survey. *Computational Visual Media*, 5(2), 117-150.
<https://link.springer.com/article/10.1007/S41095-019-0149-9>
- Bourne, R. (2010). *Fundamentals of digital imaging in medicine*. springer science & business media. ISBN-10: 1848820879.

- Chang, K. Y., Liu, T. L., & Lai, S. H. (2011, June). from saliency to co-segmentation: an efficient and fully unsupervised energy minimization model. In CVPR 2011 (PP. 2129-2136). IEEE. <https://ieeexplore.ieee.org/abstract/document/5995415/>
- Collins, M. D., Xu, J., Grady, L., & Singh, V. (2012, June). Random walks based multi-image segmentation: quasiconvexity results and gpu-based solutions. IN 2012 IEEE Conference on Computer Vision and Pattern Recognition (PP. 1656-1663). IEEE. [HTTPS://ieeexplore.ieee.org/abstract/document/6247859](https://ieeexplore.ieee.org/abstract/document/6247859)
- Dai, J., Wu, Y.N., Zhou, J., & Zhu, S., 2013. Cosegmentation and cosketch by unsupervised learning, NO. 1. https://www.cv-foundation.org/openaccess/content_iccv_2013/html/Dai_Cosegmentation_and_Cosketch_2013_ICCV_paper.html
- Faktor, A., & Irani, M. (2013). Co-segmentation by composition. in proceedings of the IEEE international conference on computer vision (PP. 1297-1304). https://openaccess.thecvf.com/content_iccv_2013/html/Faktor_Co-segmentation_by_Composition_2013_ICCV_paper.html
- Garcia-Lamont, F., Cervantes, J., López, A., & Rodriguez, L. (2018). Segmentation of images by color features: a survey. *Neurocomputing*, 292, 1-27. <https://doi.org/10.1016/j.neucom.2018.01.091>
- Gong, X., Liu, X., Li, Y., & Li, H. (2020). A novel co-attention computation block for deep learning based image co-segmentation. *Image and Vision Computing*, 101, 103973. <https://www.sciencedirect.com/science/article/abs/pii/S0262885620301050>
- Hochbaum, D. S., & Singh, V. (2009, September). An efficient algorithm for co-segmentation. IN 2009 IEEE 12th International Conference on Computer Vision (PP. 269-276). IEEE. <https://ieeexplore.ieee.org/abstract/document/5459261>
- Jerripothula, K. R., Cai, J., & Yuan, J. (2015, September). Group saliency propagation for large scale and quick image co-segmentation. IN 2015 IEEE International Conference on Image Processing (ICIP) (PP. 4639-4643). IEEE. <https://ieeexplore.ieee.org/abstract/document/7351686/>
- Jerripothula, K. R., Cai, J., & Yuan, J. (2016). Image co-segmentation via saliency co-fusion. *IEEE Transactions on Multimedia*, 18(9), 1896-1909. <https://ieeexplore.ieee.org/abstract/document/7484309>
- Jerripothula, K. R., Cai, J., Lu, J., & Yuan, J. (2017, July). Object co-skeletonization with co-segmentation. IN 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (PP. 3881-3889). IEEE. <https://ieeexplore.ieee.org/abstract/document/8099896>
- Jerripothula, K. R., Cai, J., Meng, F., & Yuan, J. (2014, October). Automatic image co-segmentation using geometric mean saliency. IN 2014 IEEE international conference on image processing (ICIP) (PP. 3277-3281). IEEE. <https://ieeexplore.ieee.org/abstract/document/7025663>
- Joulin, A., Bach, F., & Ponce, J. (2010, June). Discriminative clustering for image co-segmentation. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 1943-1950). IEEE. <https://ieeexplore.ieee.org/abstract/document/5539868>
- Kamranian, Z., Tombari, F., Nilchi, A. R. N., Monadjemi, A., & Navab, N. (2018). Co-segmentation via visualization. *Journal of Visual Communication and Image Representation*, 55, 201-214. <https://doi.org/10.1016/j.jvcir.2018.05.014>
- Lee, C., Jang, W. D., Sim, J. Y., & Kim, C. S. (2015). Multiple random walkers and their application to image cosegmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (PP. 3837-3845). https://openaccess.thecvf.com/content_cvpr_2015/html/Lee_Multiple_Random_Walkers_2015_CVPR_paper.html
- Lézoray, O., Meurie, C., Celebi, E., Lézoray, O., Meurie, C., Celebi, E., Processing, I., Vision, C., Lézoray, O., & Juin, B.M., (2017). Superpixels for Image Processing and Computer Vision. To cite this version: HAL Id: hal-01620616 Superpixels for Image. <https://hal.archives-ouvertes.fr/hal-01620616/document>
- Li, Z., & Liu, X. (2021). Improving image co-segmentation via deep metric learning. Arxiv Preprint Arxiv:2103.10670. <https://arxiv.org/abs/2103.10670>
- Lu, Z., Xu, H., & Liu, G. (2019). A survey of object co-segmentation. *IEEE Access*, 7, 62875-62893. <https://ieeexplore.ieee.org/abstract/document/8715354/>
- Meng, F., Cai, J., & Li, H. (2016). Cosegmentation of multiple image groups. *Computer Vision and Image Understanding*, 146, 67-76. <https://doi.org/10.1016/j.cviu.2016.02.004>
- Meng, F., Li, H., Liu, G., & Ngan, K. N. (2012). Object co-segmentation based on shortest path algorithm and saliency model. *IEEE Transactions on Multimedia*, 14(5), 1429-1441. <https://ieeexplore.ieee.org/abstract/document/6194335>
- Merdassi, H., Barhoumi, W., & Zagrouba, E. (2020). A comprehensive overview of relevant methods of image co-segmentation. *Expert Systems with Applications*, 140, 112901. <https://doi.org/10.1016/j.eswa.2019.112901>
- Mukherjee, L., & Peng, J., 2011. Scale Invariant cosegmentation for image groups *, p. 1881-188. <https://ieeexplore.ieee.org/abstract/document/5995420>

- Popova, A., Eerola, T., & Kälviäinen, H. (2018, September). Comparison of co-segmentation methods for wildlife photo-identification. In *International Conference on Advanced Concepts for Intelligent Vision Systems* (PP. 139-149). Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-030-01449-0_12
- Rother, C., Minka, T., Blake, A., & Kolmogorov, V. (2006, June). Cosegmentation of image pairs by histogram matching-incorporating a global constraint into MRFS. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (VOL. 1, PP. 993-1000)*. IEEE. <https://ieeexplore.ieee.org/abstract/document/1640859>
- Shaik, K. B., Ganesan, P., Kalist, V., Sathish, B. S., & Jenitha, J. M. M. (2015). Comparative study of skin color detection and segmentation in hsv and ycbcr color space. *Procedia Computer Science*, 57, 41-48. <https://www.sciencedirect.com/science/article/pii/S1877050915018918>
- Shotton, J., Winn, J., Rother, C., & Criminisi, A. (2006, May). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *European Conference on Computer vision* (PP. 1-15). Springer, Berlin, Heidelberg. https://link.springer.com/chapter/10.1007/11744023_1
- Stutz, D., Hermans, A., & Leibe, B. (2018). Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding*, 166, 1-27. <https://www.sciencedirect.com/science/article/abs/pii/S1077314217300589>