Interactive Co-segmentation Using Histogram Matching and Bipartite Graph Construction

Harsh Bhandari

Interactive Co-segmentation Using Histogram Matching and Bipartite Graph Construction

Dissertation submitted in partial fulfillment of the requirements for the degree of

Master of Technology
in
Computer Science

by

Harsh Bhandari
[Roll No: CS-1502]

under the guidance of

Dr. Bhabhatosh Chanda
Professor
Electronics and Communication Sciences Unit

Indian Statistical Institute
Kolkata-700108, India
July 2017

To my family and my supervisor

Acknowledgements

I would like to show my highest gratitude to my advisor, *Prof. Bhabatosh Chanda*, Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, for his guidance and continuous support and encouragement. He has literally taught me how to do good research, and motivated me with great insights and innovative ideas.

My deepest thanks to all the teachers of Indian Statistical Institute, for their valuable suggestions and discussions which added an important dimension to my research work.

Finally, I am very much thankful to my parents and family for their everlasting supports.

Last but not the least, I would like to thank all of my friends for their help and support. I thank all those, whom I have missed out from the above list.

Harsh Bhandari
Indian Statistical Institute
Kolkata - 700108 , India.

Abstract

Co-segmentation is defined as jointly partitioning multiple images having same or similar objects of interest into foreground and complementary part is mapped as background.

In this thesis a new interactive co-segmentation method using a global energy function and a local smooth energy function with the help of histogram matching is being proposed. The global scribbled energy takes the help of histograms of the regions in the image to be co-segmented and the user scribbled images to estimate the probability of each region belonging either to foreground or background region. The local smooth energy function helps in estimating the probability of regions having similar colour appearance.

To further improve the quality of the segmentation, bipartite graph is constructed using the segments. The algorithm has been implemented on iCoseg and MSRC benchmark data sets and the experimental results show significant good results compared to many state-of-the-art unsupervised co-segmentation and supervised interactive co-segmentation methods.

Keywords: *Co-segmentation, histogram matching, Bhattacharya Distance, Bipartite Graph.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

With the steady growth of computing power and rapid decline in cost of memory, and with the development of high-end digital cameras and ever increasing access to internet, digital acquisition of visual information has become increasingly popular in recent years. Users can easily capture more and more images and share them on internet using social networks like Facebook and Twitter or using WhatsApp. Users generally have several related images of same objects, events or places which the researchers want to exploit for many tasks such as constructing a 3D model of a particular object or developing image retrieval applications [30]. In such tasks it is imperative to extract the foreground objects from all images in a group of related images. The idea of co-segmentation, first introduced in[23], refers to simultaneously segmenting two or more images, where the same (or similar) objects appear with different (or unrelated)

backgrounds, performing segmentation of similar regions (objects) in all the images using suitable attributes of the foreground is a challenging task. The co-segmentation problem has attracted much focus in the last decade, most of the co-segmentation approaches [13, 18, 20, 21, 28, 29, 31] are motivated by Markov Random Field(MRF) based energy functions, generally solved by optimization techniques such as linear programming [20]. This co-segmentation of foreground objects from multiple related images is the goal of this algorithm.

1.2 Our Contribution

We present a new framework to perform image co-segmentation problem by first computing the probability of each pixel to be in foreground or background region by applying the Bhattacharya Distance on segments of image. Next, to further increase the quality of the segmentation we construct a bipartite graph by using the probability computed using the Bhattacharya Distance and using the size of the segments we compute the matching coefficient which further reduces the misclassification of segments.

We have provided mathematical analysis as well as simulation in details for our algorithm. We have compared performance of our algorithm with the existing algorithms with respect to accuracy of classification of each pixel into foreground or background. Time complexity of our algorithm is $O(\sum_{i=1}^{n}(N(R^i))^2 * L$ where, $N(R^i)$ is number of segments in image I^i and L = number of bins in the histogram of a segment.

1.3 Organization

The rest of the thesis is organized as follows.

- In Chapter 2 we have discussed previous work related to Co-segmentation.
- There is a brief description of required background in the Chapter 3.
- We have elaborated our proposed algorithm in Chapter 4.
- In Chapter 5 we have provided detail analysis of the algorithm.
- In Chapter 6, Section 6.1 contains performance of the algorithm visualized by simulation and Section 6.2 contains the performance evaluation of our proposed algorithm
- Conclusion and future direction research in this field is described in Chapter 7.

Chapter 2

Related Work

The identification of similar objects in more than one image is a fundamental problem and has relied on construction of models [7, 33]. Most co-segmentation methods are derived from single-image segmentation methods by adding similar foreground constraints in the MRF based optimization framework. Early co-segmentation approaches [13, 19, 20, 23] and [28] only used a pair of images as input making an assumption of sharing a common foreground object. Similar to image-segmentation, co-segmentation can be classified into two groups: unsupervised and interactive co-segmentation.

2.1 Unsupervised Co-segmentation

Many unsupervised approaches [5, 9, 11, 15, 16, 21, 22, 24, 25, 27, 31] have recently been developed to co-segment multiple images and have achieved more accurate results

than any classic single-image segmentation algorithm. In [23] image co-segmentation method is introduced by combining MRF framework and global constraints with foreground histogram matching. Houchbaum and Singh [13] proposed a max-flow algorithm by modifying the histogram matching and using clustering for co-segmentation. Joulin [15] proposed a combination of normalized cuts and kernel methods to design a discriminative clustering co-segmentation framework. Recently, they extended their framework to multi-class co-segmentation [16]. Inspired by single-image interactive segmentation methods [3, 4, 12] several co-segmentation algorithms have been proposed in recent years. But these methods fail to perform well when the foreground and background are similar in images as it is difficult to find common objects automatically. The interactive co-segmentation methods alleviate these problems by using interactive scribble.

2.2 Interactive Co-segmentation

Interactive co-segmentation algorithm [1,2] added user scribbles in some input images to build two Gaussian Mixture Models (GMM) one for each of foreground and background classes. A graph cut algorithm is then used to co-segment these images. In contrast to co-segmentation approaches within MRF, some researchers [5] proposed an image co-segmentation method using random walker algorithm based on normalized Euclidean distance of pixel intensity. However the random walk optimization will make the co-segmentation results sensitive to the quantities and positions of the user scribbles [26]. In [8] another approach was proposed where global scribbled energy function using Gaussian Mixture Model and a local smoothness function using spline

regression have been designed to improve the image co-segmentation performance.

This study formulates the interactive co-segmentation problem in terms of Gibbs energy optimization followed by generating bipartite graphs of regions of images complementing the existing MRF segmentation framework. This improves the accuracy of co-segmentation of complex images having foreground objects with variations in colour and texture. Higher order energy optimization [14,17,34] has been widely used in many fields: computer vision and image processing like image denoising and single image segmentation. Bipartite graphs are constructed using unlabelled segments of the image as one set of vertices and the scribbled foreground and background segments as another set of vertices. Each unlabelled vertex is connected to the other set obtained from scribbled region with a weighted edge. This strategy makes the framework effective in realistic scenario as shown in Fig.2.1, parts of foreground and background regions are similar in colour where the proposed algorithm efficiently captures the foreground objects after bipartite graph formulation.

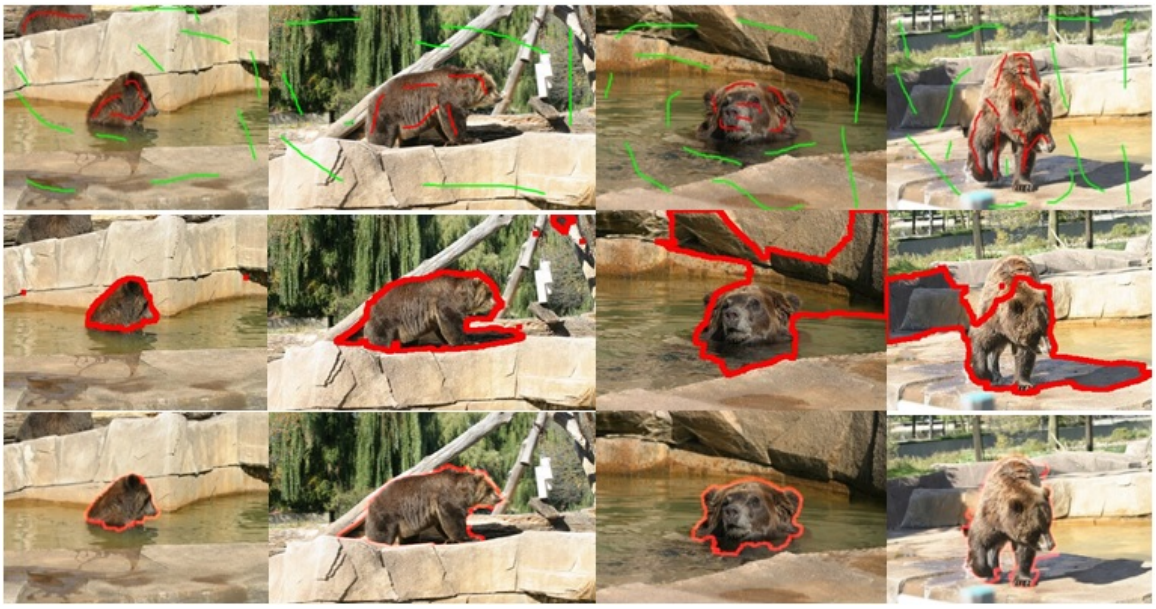

Figure 2.1: Co-segmentation results. Left: input images Alaskan Brown Bear with scribbled group. Middle: results of co-segmentation by composition [9]. Right: result of the proposed approach.

Chapter 3

Overview of the proposed approach

Compared to the existing image co-segmentation methods, the proposed algorithm offers the following contributions.

1. The proposed interactive co-segmentation algorithm can be divided into two phases:
    - probability estimation of each pixel to be in foreground/background region
    - bipartite graph generation
2. A new segment classification method is presented using the histograms of the segments of the image based on a global scribbled unary energy function and a local pairwise smooth energy function subject to prior information. This results in classification of each pixel into foreground/background regions and helps in constructing bipartite graph for final co-segmentation results.
3. A bipartite graph construction method is proposed using the scribbled fore-

ground/background regions as well as the regions of original image which need to be properly classified.

The workflow of the proposed interactive co-segmentation framework using global and local energy function along with bipartite graph construction is shown in Fig.3.1

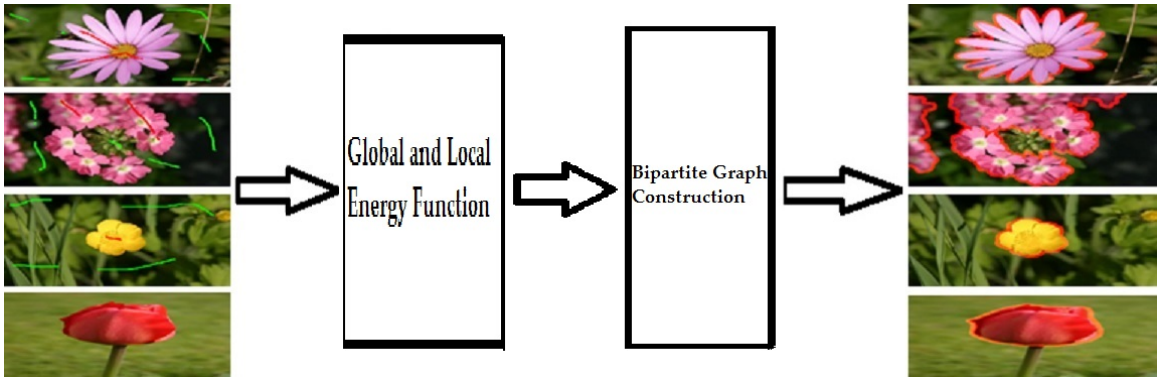

Figure 3.1: The Left images are scribbled images indicating red foreground region and green background region along with one no scribble image since the proposed algorithm can segment some of the images without scribble. The right side images are result of co-segmenting four images in the Flower class in the MSRC dataset[32].

Chapter 4

Proposed Algorithm

In this paper, a novel algorithm has been introduced that classify each pixel into foreground/background regions.

Let S =$\{I^1, \cdots, I^n\}$ be a set of n images and let P be a subset of k images selected from S indicating foreground and background scribbles where k$\ll$n. Let p^i_j denotes a pixel of image I^i with index j and $a^i_{j,l}$ be its probability for foreground/background region with l= 0 indicating the background region and l= 1 as the foreground region. To make the algorithm computationally efficient, each image I^i is divided into small segments $r^i_m \ \epsilon \ R^i$ by using an over-segmentation method such as mean-shift [6] or efficient graph [10] method and N(R^i) be the total number of segments in image I^i i.e. $1 \leq$ m $\leq$ N(R^i). Let $b^i_{m,l}$ be the probability of segment r^i_m belonging to region l ϵ {0, 1}. To compute, $b^i_{m,l}$, two energy functions are designed: a unary second order global scribbled energy function E_{global} and a pairwise second order local smooth function E_{local}. Let S^0 be the set of scribbled background segments and S^1 be the set

of scribbled foreground segments of set P.

4.1 Global Scribbled function

How to effectively utilize the user scribbles is the key for interactive co-segmentation. The global energy function is responsible for providing the probability of segment $b^i_{m,l}$ to foreground/background region based on the prior probability computed as below:

Let $c^i_{m,l}$ be the initial estimate of segment r^i_m belonging to region l where

$$c^i_{m,l} = 1 - e^i_{m,l} \tag{4.1}$$

$$\begin{aligned} e^i_{m,0} &= \frac{\min_{s0\in S^0}(-\log(D(h^i_m, h^{s0})))}{\min_{s0\in S^0}(-\log(D(h^i_m, h^{s0}))) + \min_{s1\in S^1}(-\log(D(h^i_m, h^{s1})))} \\ e^i_{m,1} &= \frac{\min_{s1\in S^1}(-\log(D(h^i_m, h^{s1})))}{\min_{s0\in S^0}(-\log(D(h^i_m, h^{s0}))) + \min_{s1\in S^1}(-\log(D(h^i_m, h^{s1})))} \end{aligned} \tag{4.2}$$

Note that $D(x, y)$ stands for Bhattacharya Distance defined as

$$D(x, y) = \sum_{s=1}^{s=L} \sqrt{h^i_x(s).h^i_y(s)}$$

where h^i_q = normalized histogram of region r^i_q in image I^i and L = number of bins in a histogram.

$D(x, y)$ states the closeness of two distribution x and y. $D(x, y) = 1$ indicates the two distributions are identical and $D(x, y) = 0$ indicates that the distributions have no relation.

Thus the unary global energy function can be defined as

$$E_{global} = \sum_{m=1}^{N(R^i)} (b^i_{m,l} - c^i_{m,l})^2 \tag{4.3}$$

Eq.4.3 can be solved by converting the equation into matrix form thus,

$$E_{global} = (\vec{b^i_l} - \vec{c^i_l})^T * I_{N(R^i)\times N(R^i)} * (\vec{b^i_l} - \vec{c^i_l}) \tag{4.4}$$

where $\vec{b^i_l}$ and $\vec{c^i_l}$ are vectors of size $N(R^i)$ and I is an identity matrix of size $N(R^i) \times N(R^i)$.

E_{global} is defined by satisfying a constraint that each segment tends to have the probability $b^i_{m,l}$ of segment r^i_m close to $c^i_{m,l}$ estimated through the Bhattacharya Distance.

4.2 Local Smooth function

The local smooth energy function considers the smoothness of segments i.e. it provides interactive constraint that all segments in the image have same probability of belonging to foreground/background if having similar colour appearance. To compute the local energy function, a graph $G^i(V^{(i)}, E^{(i)})$ is constructed for image I^i where $V^{(i)} = b^i_{1,l}, \ldots, b^i_{N(R^i),l}$ and $E^{(i)} = (b^i_{1,l}, b^i_{1,l}), (b^i_{1,l}, b^i_{2,l}), \ldots, (b^i_{N(R^i),l}, b^i_{N(R^i),l})$ i.e. each vertex is connected to all other vertices including itself with weight $w^i_{p,q}$ where

$0 \leq w^i_{p,q} \leq 1$, p and q are the segments of the image

$$w^i_{p,q} = e^{\sum_{s=1}^{s=L} \sqrt{h^i_p(s) * h^i_q(s)} - 1}$$

E_{local} defines a constraint that all segments having similar colour appearance will have same probability of getting classified into foreground/background.Thus the pairwise local function can de defined as

$$E_{local} = \sum_{r,s=1}^{N(R^i)} w^i_{r,s} * (b^i_{r,l} - b^i_{s,l})^2 \tag{4.5}$$

Eq.4.5 can be solved by constructing Laplacian Le of graph $G^i(V^{(i)}, E^{(i)})$, which is a positive semi-definite matrix of size $N(R^i) \times N(R^i)$ represented as:

$$Le = \begin{bmatrix} \sum_{r=1}^{N(R^i)} w^i_{1,r} & 0 & \cdots & 0 \\ 0 & \sum_{r=1}^{N(R^i)} w^i_{2,r} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{r=1}^{N(R^i)} w^i_{N(R^i),r} \end{bmatrix} - \begin{bmatrix} w^i_{1,1} & w^i_{1,2} & \cdots & w^i_{1,N(R^i)} \\ w^i_{2,1} & w^i_{2,2} & \cdots & w^i_{2,N(R^i)} \\ \vdots & \vdots & \ddots & \vdots \\ w^i_{N(R^i),1} & w^i_{N(R^i),2} & \cdots & w^i_{N(R^i),N(R^i)} \end{bmatrix}$$

$$E_{local} = (\vec{b^i_l})^T * Le_{N(R^i) \times N(R^i)} * (\vec{b^i_l}) \tag{4.6}$$

4.3 Total Energy Optimization

The total energy can be summarized as a sum of the global scribbled energy, E_{global} and local smooth energy, E_{local}.

$$E^i_{Total} = E_{global} + E_{local}$$

$$E^i_{Total} = (\vec{b^i_l} - \vec{c^i_l})^T * I_{N(R^i)\times N(R^i)} * (\vec{b^i_l} - \vec{c^i_l}) + (\vec{b^i_l})^T * Le_{N(R^i)\times N(R^i)} * (\vec{b^i_l}) \tag{4.7}$$

b^i_l being the parameter to be computed can be found by differentiating E^i_{Total} with respect to b^i_l. Thus

$$\frac{\partial E^i_{Total}}{\partial b^i_l} = 2 * I_{N(R^i)\times N(R^i)} * (\vec{b^i_l} - \vec{c^i_l}) + 2 * Le_{N(R^i)\times N(R^i)} * (\vec{b^i_l})$$

$$I_{N(R^i)\times N(R^i)} * (\vec{b^i_l} - \vec{c^i_l}) + Le_{N(R^i)\times N(R^i)} * (\vec{b^i_l}) = 0$$

$$b^i_l = \frac{c^i_l}{I_{N(R^i)\times N(R^i)} + Le_{N(R^i)\times N(R^i)}} \tag{4.8}$$

$I_{N(R^i)\times N(R^i)} + Le_{N(R^i)\times N(R^i)}$ can be written as $A_{N(R^i)\times N(R^i)}$ where

$$A = \begin{bmatrix} \sum_{r=1}^{N(R^i)} w^i_{1,r} & w^i_{1,2} & \cdots & w^i_{1,N(R^i)} \\ w^i_{2,1} & \sum_{r=1}^{N(R^i)} w^i_{2,r} & \cdots & w^i_{2,N(R^i)} \\ \vdots & \vdots & \ddots & \vdots \\ w^i_{N(R^i),1} & w^i_{N(R^i),2} & \cdots & \sum_{r=1}^{N(R^i)} w^i_{N(R^i),r} \end{bmatrix}$$

$$b^i_l = \frac{c^i_l}{A_{N(R^i)\times N(R^i)}} \tag{4.9}$$

Note that $a^i_{j,l} = b^i_{m^j,l}$, where m^j indicates the segment $r^i_{m^j}$ that pixel p^i_j belongs to.

4.4 Bipartite Graph Construction

The proposed energy function usually provides a good estimation for the foreground/background region in each image. But for images having complex textures and where the foreground and background regions have close colour appearance, such regions may get misclassified since the initial estimation may come close to 0.5.

To further increase the accuracy of the co-segmentation results, for each image I^i, a set X^i is constructed where the set provides segments r^i_r having $b^i_{r,1} \in \{0.5 - \epsilon_1, 0.5 + \epsilon_2\}$. All segments having $b^i_{r,1} > 0.5 + \epsilon_2$,

is classified into foreground region and those having $b^i_{r,1} < 0.5 - \epsilon_1$ is classified into background region. Set $Y^i_{r,l}$ is constructed which contains all adjacent segments of r^i_r classified into region l. To make better estimation of the segments in X^i, matching coefficient $d^i_{r,l}$ is computed as stated below:

$$d^i_{r,l} = \frac{N(r^i_{r,l}) * b^i_{r,l} + u^i_{r,l} + v^i_{r,l}}{N(r^i_{r,l}) + N(r^i_{tv,l}) + N(r^i_{tu,l})}$$

where $u^i_{r,l}$ indicates the highest similarity value between segment r^i_r and its adjacent segments in $Y^i_{r,l}$ and $v^i_{r,l}$ indicates the highest similarity value between segment r^i_r and S^l. $N(r^i_r)$ indicates number of pixels in segment r^i_r.

4.4.1 Computation of $v^i_{r,l}$

Let $G^i_{b,v}(X^i \bigcup S^l, E_v)$ be a complete bipartite graph with X^i and S^l be two disjoint set of vertices. E_v is the set of edges between X^i and S^l having weight $\varepsilon^i_{r,t,l}$ between segment $r^i_r \in X^i$ and segment $r_{tv,l} \in S^l$. Hence

$$\begin{aligned} \varepsilon^i_{r,tv,l} &= \frac{\sum_{j=1}^{L}((h^i_r(j)) - \frac{1}{L}) * ((h^i_{tv,l}(j)) - \frac{1}{L})}{\sqrt{(\sum_{j=1}^{L}(h^i_r(j))^2 - \frac{1}{L}) * (\sum_{j=1}^{L}(h^i_{tv,l}(j))^2 - \frac{1}{L})}} \\ v^i_{r,l} &= N(r^i_{tv,l}) * \max(\max_{r_{tv,l} \in S^l}(\varepsilon^i_{r,tv,l}, 0)) \end{aligned}$$

$N(r^i_{tv,l})$ indicates number of pixels in segment $r_{tv,l}$, which means a large segment will have large weight and $\varepsilon^i_{r,tv,l}$ considers the highest correlation that the segment can have with the scribbled foreground/background segments and thus improves the segmentation quality.

4.4.2 Computation of $u^i_{r,l}$

Let $G^i_{b,u}(r^i_r \bigcup Y^i_{r,l}, E_u)$ be a complete bipartite graph with r^i_r and $Y^i_{r,l}$ be two disjoint set of vertices. E_u is the set of edges between segment r^i_r and $Y^i_{r,l}$ having weight $\varepsilon^i_{r,tu,l}$ between segment r^i_r and segment $r_{tu,l} \in Y^i_{r,l}$. Hence, we can write

$$\begin{aligned}
\varepsilon^i_{r,tu,l} &= \frac{\sum_{j=1}^{L}((h^i_r(j)) - \frac{1}{L}) * ((h^i_{tu,l}(j)) - \frac{1}{L})}{\sqrt{(\sum_{j=1}^{L}(h^i_r(j))^2 - \frac{1}{L}) * (\sum_{j=1}^{L}(h^i_{tv,l}(j))^2 - \frac{1}{L})}} \\
u^i_{r,l} &= N(r^i_{tu,l}) * \max(\max_{r_{tu,l} \in S^l}(\varepsilon^i_{r,tu,l}, 0)) \\
d^i_{r,l} &= \frac{N(r^i_{r,l}) * b^i_{r,l} + u^i_{r,l} + v^i_{r,l}}{N(r^i_{r,l}) + N(r^i_{tv,l}) + N(r^i_{tu,l})} \\
a^i_{j,l} &= d^i_{r_j,l}
\end{aligned} \tag{4.10}$$

All pixels p^i_j with $a^i_{j,1} > 0.5$ are classified into foreground region and rest to background region.

Algorithm 1 Interactive Co-segmentation

Input: $S = \{I^1, \ldots, I^n\}$, set of n images,P = set of scribbled images,
Output: $a^i_{j,1}$

1: Perfrom Mean Shift oversegmentation on S
2: Select segments from P.
3: Classify scribbled segments into foreground/background based on red/green scribble respectively.
4: Compute $b^r_{m,l}$
5: Set ϵ_1 and ϵ_2 parameter
6: Construct Bipartite Graph
7: Compute $u^i_{r,l}$ and $v^i_{r,l}$
8: Set $d^i_{r,l}$ using $u^i_{r,l}$ and $v^i_{r,l}$
9: Set $a^i_{j,l}$
10: Select $a^i_{j,1} > 0.5$ as foreground pixel.
11: **return**

Chapter 5

Analysis Of The Algorithm

In this chapter, brief analysis of the proposed algorithm is made. We analyse *Global Scribble function, Local Smooth function and the effect of bipartite graph construction.* At the end we compare performance of our scheme with the existing schemes. We have verified the outcomes of the analysis in Section 6.1.

5.1 Analysis of Global Scribble Function

The unary global scribble function is responsible for the initial probability each segment of an image computed using the Bhattacharya Distance. Implementation of Mean Shift [6] with small value of spatial

bandwidth and range bandwidth provides a very good over-segmentation result. All the pixels in each segment of the image have close colour appearance to each other. Thus the expectation of the distribution of the pixel value is close to the average value of the segment.

Histogram of each segment is normalized to make it distribution of each segment. $D(p, q)$ checks the overlap of two histograms p and q with value 1 indicating that the histograms are identical and that of 0 indicating no relation among them. $-(\log(D(p, q)))$ thus represent the distance between histogram p and q where $0 \leq -(\log(D(p, q))) < \infty$ with value 0 indicating identical segments. For segment r_m^i we look for the segment with red/green scribbles representing foreground/background segments and label the segment to foreground or background region depending on the minimum distance between r_m^i and that of the scribbled segment.

Thus to compute the probability of each segment r_m^i belonging to foreground/background, we select a segment each from foreground and background scribbled segments for which the Bhattacharya Distance is minimum, computing the sum of two distances as a normalization factor to the distance between segment r_m^i and the segment from scribbled

segment sets providing minimum Bhattacharya Distance.

Thus segment r_m^i belongs to foreground region if its distance with that of a *red* scribbled segment is minimum compared to that of *green* scribbled segment as the probability which is $1 - e_{m,l}^i$ is higher for foreground then background. Hence we can write

$$c_{m,0}^i + c_{m,1}^i = 1 \tag{5.1}$$

Thus the global scribbled function designed using the Bhattacharya Distance provides a good estimation for each segment to part either with foreground or background region.

5.2 Analysis of Local Smoothness Function

The pairwise local smoothness function defines an interactive constraint on the image that all the segments of the image having similar colour appearance must have similar probability of belonging to foreground or background region maintaining the consistency of the segmentation labels thus making the co-segmentation results more effective.

Implementation of the function is made by the construction of lapla-

cian of the graph constructed using the segments of the image with an edge of weight $w^i_{r,s}$ between r^i_r and r^i_s where

$$w^i_{r,s} = \mathrm{e}^{(D(h^i_{r,l}, h^i_{s,l}))-1}$$

indicating a similarity measure between the segments where $w^i_{r,s}$ close to 1 shows the segments are very identical to each other and 0 indicating segments with complete different colour appearance.

5.3 Analysis of bipartite graph construction

Use of histogram matching and computing the probability, provide a good result for segments which are either close to foreground scribble or that of background scribble. Thus for those group of images which have distinctive foreground and background region will yield good results. But for complex set of images where both foreground and background region have similar colour appearance, the proposed algorithm provides a rough and good estimation of the segments, but for segments which have close colour appearance with both foreground and background region, thus distance of the segment with both foreground and background region are close to each other. This will provide estimation

close to 0.5. To further improve the accuracy of the segmentation, a new segmentation algorithm using bipartite graph construction is designed by selecting segments having foreground probability in the range $\{0.5 - \epsilon_1, 0.5 + \epsilon_2\}$ where ϵ_1 and ϵ_2 are user defined parameters where $0 \leq \epsilon_1, \epsilon_2 \leq 0.5$. It is advised to set the parameters small so that more number of segments adjacent to those in $X^i \in I^i$ but not included in X^i can be obtained, as it will reduce the time computation and with higher values of the parameter the number of adjacent segments will get reduce finally leaving only the foreground/background scribble segments for matching coefficient computation.

For each segment $r_r^i \in X^i$, the algorithm finds the most similar foreground and background segment from the user scribbled segments by finding the highest correlated segments from foreground/background scribbled set.

Next, the algorithm finds the most similar foreground and background segment from those adjacent to r_r^i but not included in X^i. This will further increase the co-segmentation smoothness.

The matching parameter is used to estimate the segmentation quality which is based on region consistency assumption which encourages

all pixels belonging to a region to take same label. The more pixels of a segment have the same label, better is the segmentation quality on this segment.

In the bipartite graph construction for r_r^i, only quality of the segmentation is considered since pixels in the scribbled segments have already been classified into foreground/background region. Similarly, the adjacent segments which have been classified into foreground/background region plays a vital role in estimating the quality of the segment r_r^i. Number of pixels in the best related scribbled and adjacent segment can influence the matching coefficient. If segment r_r^i is closer to foreground, $u_{r,l}^i$ and $v_{r,l}^i$ is larger and $N(r_{tu,l})$ and $N(r_{tv,l})$ respectively plays influential role in determining matching coefficient.

The proposed algorithm considers the similarity between segment r_r^i and the scribbled foreground/background segments along with the foreground/background segments adjacent to it, which helps in establishing the quality of the segment by maximizing the matching coefficient.

The effectiveness of bipartite graph construction for improving the co-segmentation result is illustrated by running the panda image group from iCoseg dataset with and without the bipartite graph construction.

The algorithm without bipartite graph construction only includes the global scribble function and local smooth function. The panda group includes 20 images with common object panda and different complex background. The initial part of the algorithm provides an average accuracy of 86.30% and with the use of bipartite graph construction the accuracy increases to 96.42%. Therefore, co-segmentation with the help of bipartite graph construction outperforms that of using global and local energy function.

Qualitatively, using the global and local function, some of the images loses the foreground region and some provides redundant background as foreground region as shown in Fig.5.1.The second row in the figure shows that some foreground regions are lost and contain some redundant regions. The third row shows the result after implementing the whole algorithm which is much closer to the ground truth shown in final row. The problem lies with the fact that the initial algorithm does not work well in cases where the foreground and background are very similar, since this will lead to the probability of a segment getting classified into both foreground and background region close to each other or lack of enough foreground object information. To overcome this problem, bipartite graphs are constructed with the regions hav-

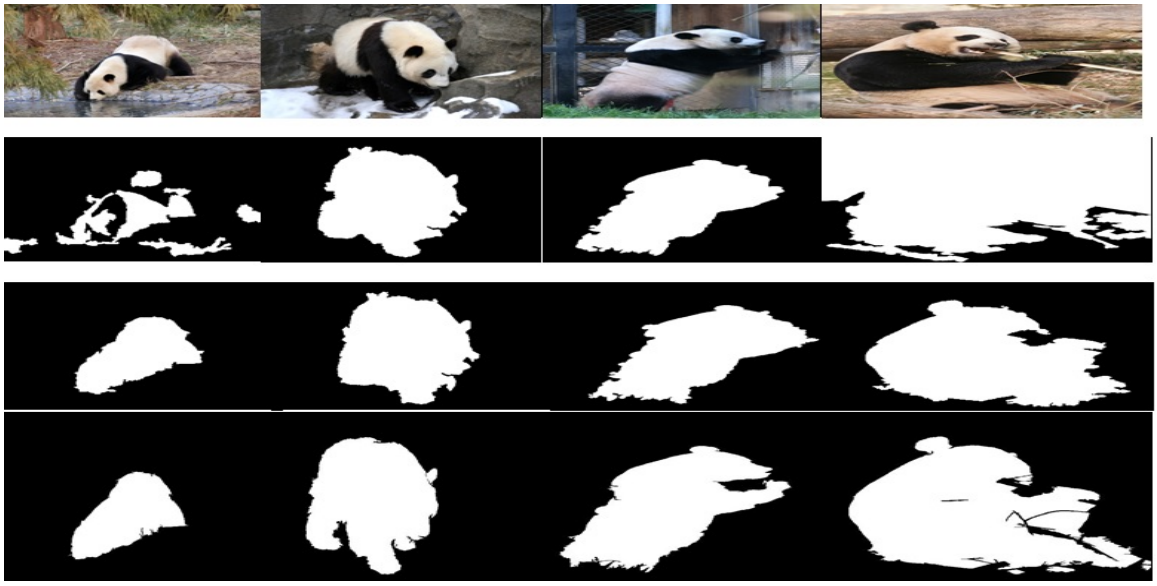

Figure 5.1: Co-segmentation results on the panda set of images from iCoseg. The first row: input images. The second row: result of the proposed algorithm without bipartite graph. The third row: co-segmentation results by full algorithm. The last row: ground truth. The algorithm is run on all 20 images and 4 are selected for illustration purpose.

ing higher probability of getting misclassified. Thus selecting those segments which have close probability of getting classified into both foreground and background region with the scribbled segments and the adjacent classified segments. Further complex foreground/background images must be provided with more scribbles compared to those with simple images as they contain lots of information. In summary, the full co-segmentation algorithm provides more foreground information to each image, thus improving the final results.

Chapter 6

Experimental Results and Discussion

6.1 Simulation

In this section, the performance of the algorithm is evaluated by making a qualitative and quantitative study on certain benchmark datasets. The proposed co-segmentation method is evaluated on two benchmark datasets: MSRC dataset[32] and iCoseg dataset[1] which have been widely used by previous work to evaluate the performance of the image co-segmentation method. The iCoseg dataset consists of 38 groups with total 643 images that each group of images has a common object.The

experiment has been evaluated on randomly selecting 35 such group of images for perform evaluation. Similarly the MSRC dataset contains 20 group of images and we randomly selected 9 such group for performance evaluation.Quantitative evaluation includes two performance metrics: Accuracy (A) and Jaccard similarity (J). 'A' indicates the ratio of correctly classified pixels into foreground and background region. 'J' indicates intersection over union of the segmentation results and ground truth mask.

6.1.1 Parameter settings

There are some suggestions for our interactive method on how to provide user scribbles.

- Some images with complex background should be provided scribbles first as it provides required information for making good estimation of the background regions and to provide sparse scribbles to images with simple background.
- The scribble should contain as many colour variations as possible, so the regions with variable colour inside foreground/background are good choice to put the scribbles on.

- The regions with similar colours in foreground and background region must be scribbled. User should add scribbles until these scribbles have contained most colour information.

Once the scribbles are provided, mean shift over-segmentation algorithm [6] parameter, spatial bandwidth $h_s = 8$, range bandwidth $h_r = 7$ and minimum size of the segment $M = 20$ are set experimentally based on training images. Unless mentioned otherwise, the following parameters are used in the proposed algorithm, $\epsilon_1 = 0.1$ and $\epsilon_2 = 0.1$.

6.1.2 Simulation Environment

The matlab program is simulated under the stated computer configuration: Intel Core $i5 - 6200$U CPU @ 2.3GHz with $4GB$ RAM. The complexity of the proposed algorithm can be divided into two phases. For finding the probability of each pixel to foreground/background region, the complexity is $O(\sum_{i=1}^{n} (N(R^i))^2 * L$. In case of bipartite graph construction, let the number of regions used for increasing the segmentation quality is $S(R^i)$ which is less than that in computing the probability of each segment using the Bhattacharya Distance. Hence the time complexity of the algorithm is $O(\sum_{i=1}^{n} (N(R^i))^2 * L$.

6.1.3 Co-segmentation Results

In the experiment, we collect a variety of image groups from well known image databases such as iCoseg dataset[1] or Microsoft MSRC dataset[32]. These two dataset are most popular dataset for image co-segmentation experiments where the ground truth segmentation mask are also provided.

In the Fig.6.1 we have shown the scribbled images of 002 Alaskan Brown Bear-Eukaryote museums Milwaukee Zoo 2006-Cmlburnett from iCoseg dataset as it is one of the most complex image group in the whole dataset.

In Fig.6.2 we have shown the scribbled images of Flower class from

Figure 6.1: Scribbled Images (Alaskan Brown Bear) from iCoseg dataset

MSRC dataset. Their corresponding co-segmentation result using the

Figure 6.2: Scribbled Images (Flower) from MSRC dataset

proposed algorithm is shown in Fig.6.3 and Fig.6.4 respectively.

6.2 Performance Evaluation

In this section we have evaluated performance of our proposed algorithm with respect to accuracy. The comparison is done by comparing pixel by pixel of the mask of ground truth with that obtained from the algorithm. In Fig.6.5 a comparison of the proposed algorithm is made with Co-segmentation by Composition algorithm[9] using iCoseg dataset and in Fig.6.6 that using the MSRC dataset.

6.2.1 Accuracy Evaluation

In this section, we list the precision statistics i.e Accuracy (A) and Jaccard Similarity (J) for each of iCoseg dataset group of images in Table6.1 and MSRC dataset group of images in Table6.2 selected for evaluation. We have also compared the corresponding statistics with state-of-the-art co-segmentation by composition algorithm[9]. In Table6.1 and 6.2, the first column indicates the name of the group image followed by number of images in each group. The third column indicates the accuracy of our proposed algorithm followed by which we have provided Jaccard Similarity value. To evaluate the performance of out algorithm, we have compared the result with Co-segmentation by Composition algorithm[9] in the last column. From the accuracy value, it can be seen that the proposed algorithm outperforms state-of-the-art algorithm[9]. The average accuracy of the proposed algorithm on iCoseg dataset is 96.33% and that of MSRC dataset is 91.19% which is less compared to iCoseg as the images in iCoseg are less complex than that of MSRC dataset.

6.2.2 Time Comparison

In this section, Table 6.3 provides the total run time on each group of image of MSRC dataset and Table 6.4 provides the total run time for each group of image of iCoseg dataset. In Table6.3 and Table6.4 run time of the proposed algorithm has been depicted on MSRC and iCoseg dataset respectively. All the run time values are measured in seconds on Intel Core $i5-6200$U CPU @ 2.3GHz and $4GB$ RAM. It can be observed that time taken for co-segmentation in MSRC dataset is more than that of iCoseg dataset, as the images in MSRC are highly complex compared to that of iCoseg. Thus it produces more number of segments resulting in more time consumption.

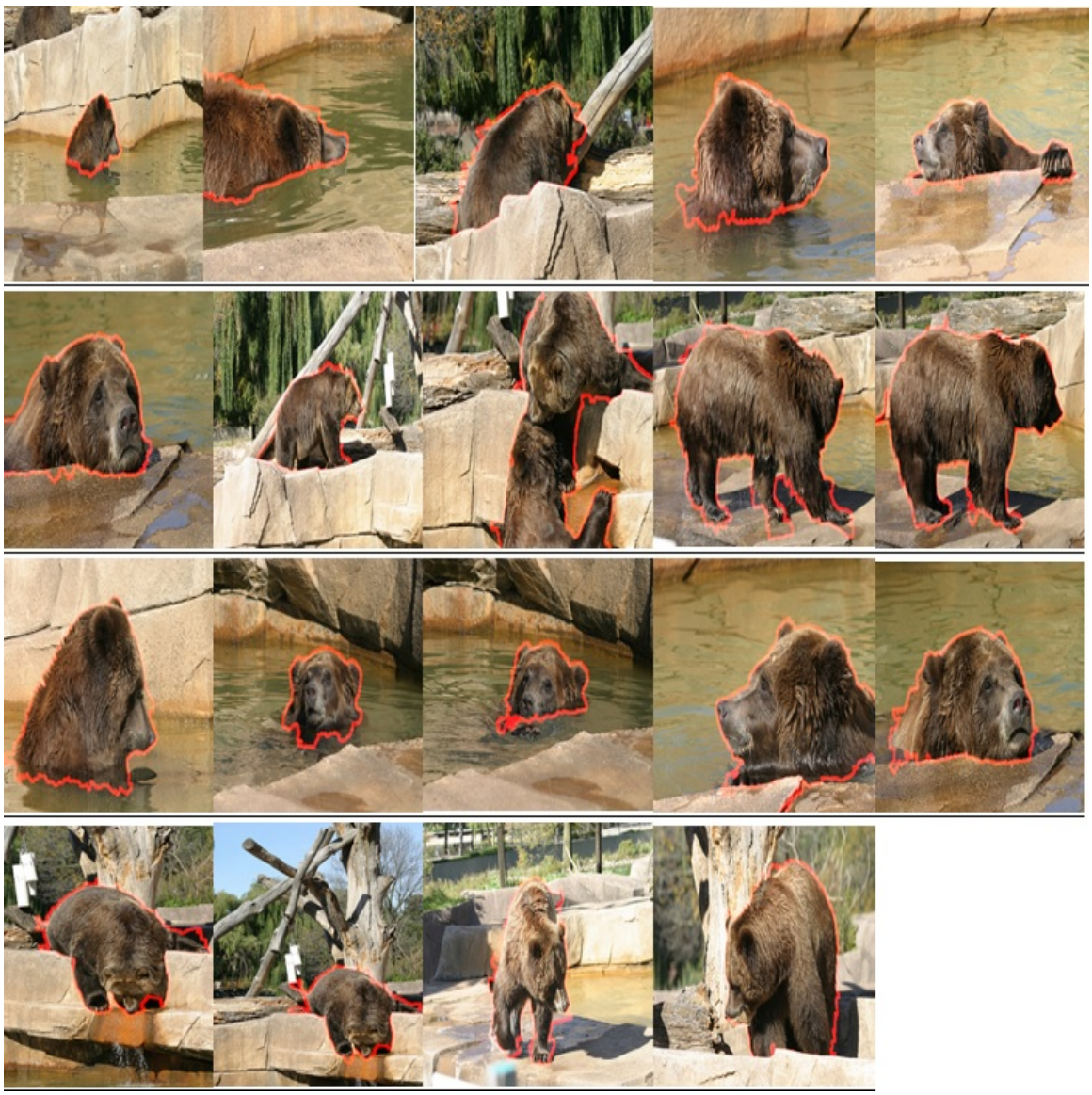

Figure 6.3: Co-segmentation Results (Alaskan Brown Bear): By the proposed algorithm

Figure 6.4: Co-segmentation Results (Flower): By the proposed algorithm

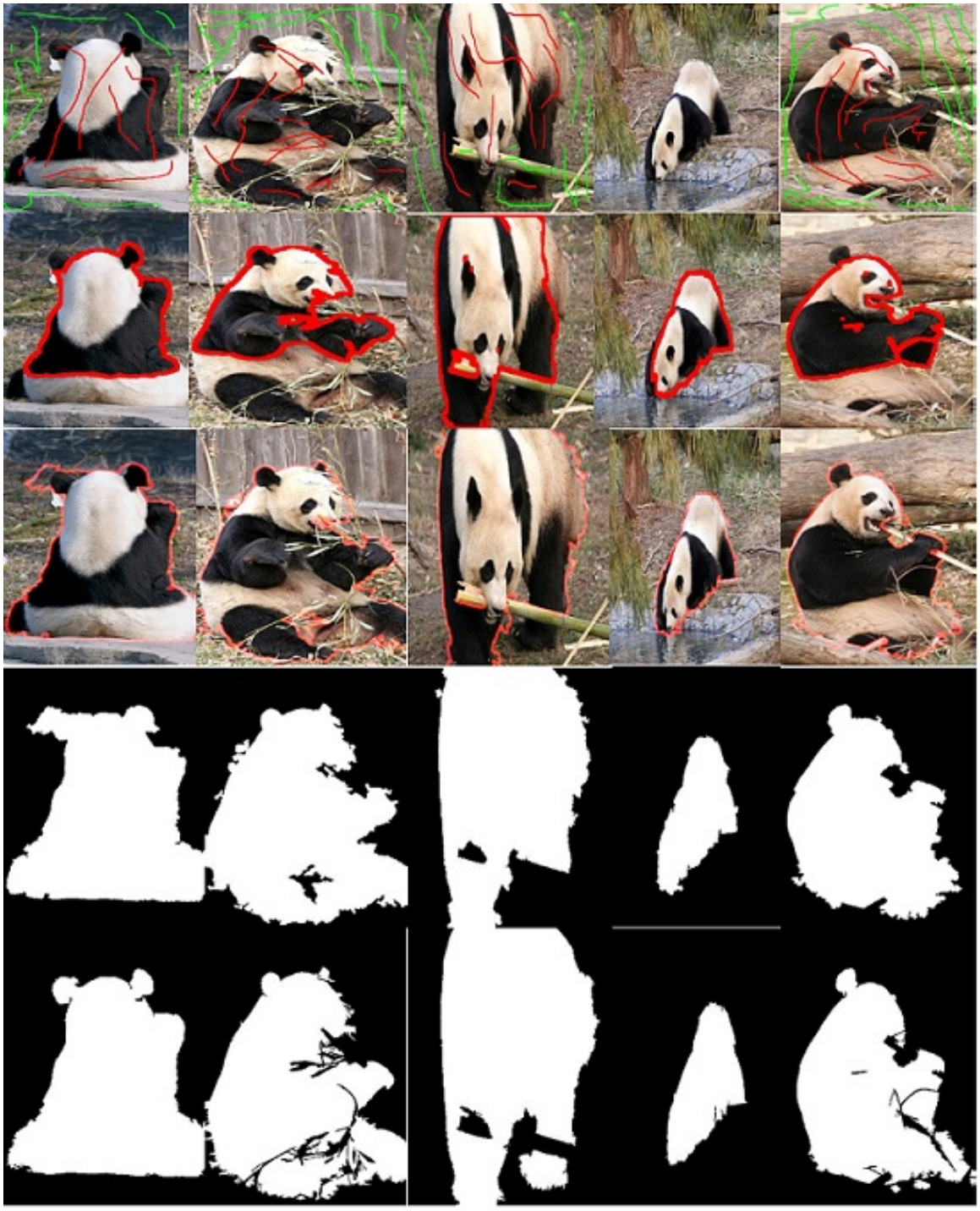

Figure 6.5: Comparison Results: The First row contains the scribbles from ICOSEG [1] dataset. The Second row is the co-segmentation results obtained by Co-segmentation by Composition [9]. Third row contains co-segmentation results by the proposed algorithm. Fourth row contains the mask obtained from the proposed algorithm and fifth row contains the ground truth.

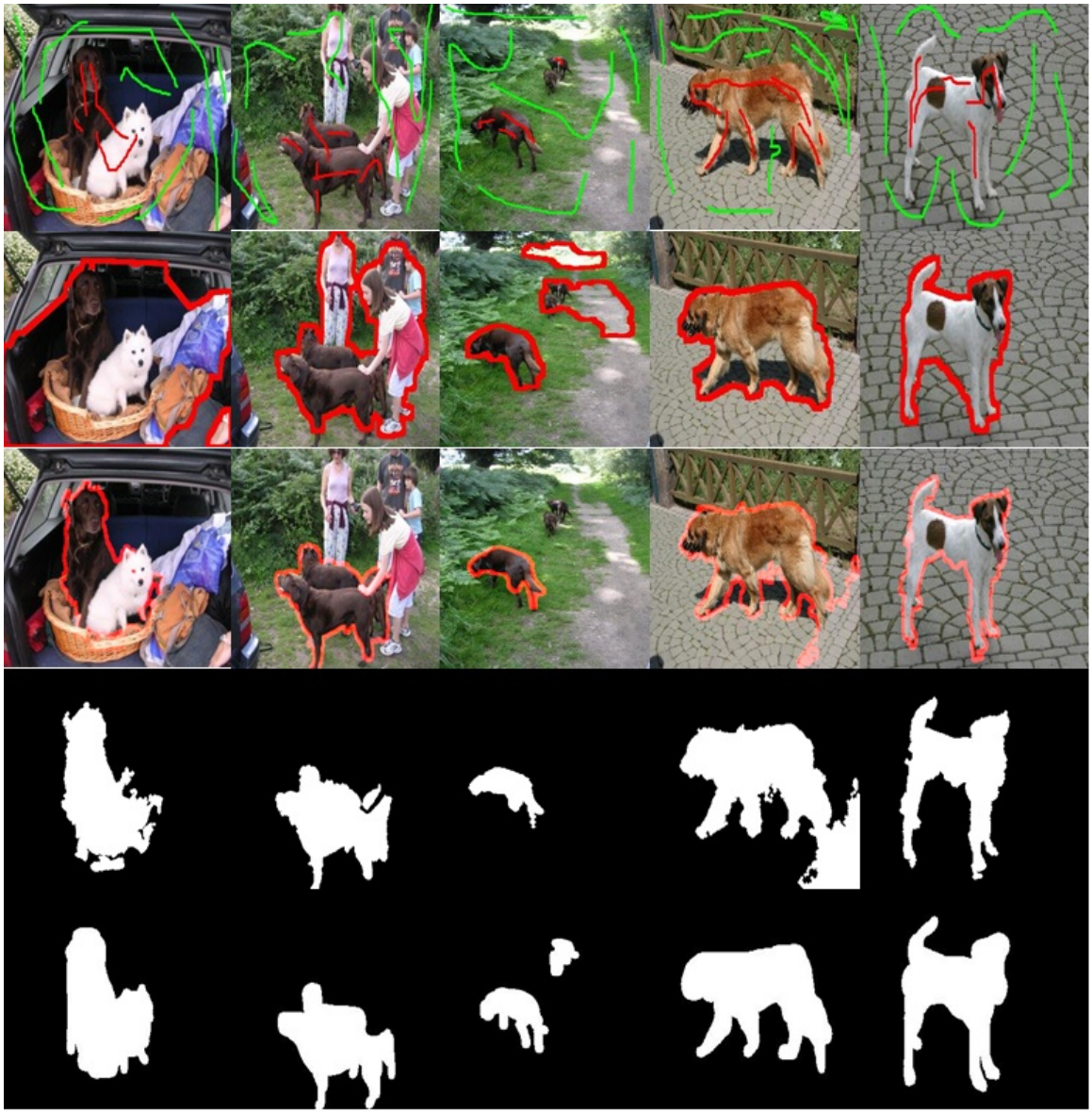

Figure 6.6: Comparison Results: The First row contains the scribbles of Dog class images from MSRC dataset[32]. The Second row is the co-segmentation results obtained by Co-segmentation by Composition [9]. Third row contains co-segmentation results by the proposed algorithm. Fourth row contains the mask obtained from the proposed algorithm and fifth row contains the ground truth.

Table 6.1: Performance Evaluation of the co-segmentation result obtained after the implementation of the proposed algorithm with respect to the ground truth based on accuracy (A) and Jaccard Similarity (J)(iCoseg Dataset)

Name of group-image	#images	$A(OurAlgorithm)$	J	A [9]
002 Alaskan Brown Bear	19	97.502892	0.873295	72.561460
006 Red Sox Players	25	96.535165	0.721738	81.933203
009 Stonehenge	5	96.513261	0.731919	72.703047
012 Stonehenge	18	96.261414	0.865248	54.087859
015 Ferrari	11	95.864048	0.836310	67.507976
017 Agra Taj Mahal	5	97.315093	0.841502	80.452388
018 Agra Taj Mahal	5	97.102474	0.837948	79.366940
020 Pyramids	10	97.403243	0.846149	69.739093
021 Elephants-safari	15	96.369770	0.820358	73.647929
022 Goose-Riverside	30	97.166887	0.889424	70.403854
023 Pandas-Tai-Land	25	92.957098	0.874784	55.740881
025 Airshows-helicopter	12	97.908192	0.861047	90.350304
025 Airshows-planes	39	96.589485	0.570366	91.423372
026 Airshows-Huntsville	22	99.346050	0.668831	88.533857
028 Cheetah-National Zoo-1	15	96.082177	0.731300	61.842540
028 Cheetah-National Zoo-2	18	95.974143	0.732347	60.793210
029 Pandas-National Zoo	20	96.421212	0.771936	43.487349
032 Kite-Brighton kite Festival	18	96.820681	0.685849	88.534857
033 Kite-kitekid	10	96.511733	0.720002	61.273948
034 Kite-Margate Kite Festival	7	96.970021	0.722711	64.935258
035 Kite-Colt Park	11	98.403306	0.698228	87.890952
036 Gymnastics-1	6	98.238750	0.704026	88.726963
036 Gymnastics-2	4	82.997000	0.698005	81.780667
036 Gymnastics-3	6	97.309679	0.703186	82.840889
038 Skating-ISU	12	98.218711	0.704730	88.551300
039 Women Soccer Players	27	95.707246	0.739985	74.375408
040 Monks-LAO PDR	13	94.925941	0.744649	65.479467
041 Hot Balloons-Skybird	23	98.805484	0.799611	84.675067
042 Statue of Liberty	40	98.216293	0.928956	73.727867
043 Christ the Redeemer	13	96.972800	0.924596	62.273722
048 Windmill	17	98.317459	0.812920	88.171369
049 Kendo-Kendo	30	95.915531	0.876146	73.800025
050 Kendo-EKC 2008	11	92.680533	0.848536	74.375234
brown bear	5	95.369518	0.826762	64.155689

Table 6.2: Performance Evaluation of the co-segmentation result obtained after the implementation of the proposed algorithm with respect to the ground truth based on accuracy (A) and Jaccard Similarity (J)(MSRC Dataset)

Name of group-image	#images	A(Our Algorithm)	J	A [9]
$Bird$	34	95.258065	0.746541	79.288703
Cat	24	86.291968	0.648297	67.583714
$Cows$	30	92.909918	0.781766	71.746577
Dog	30	95.415200	0.801606	67.683955
$Face$	30	85.187451	0.612260	70.083138
$Flower$	32	88.799500	0.764719	56.607330
$Horse$	31	93.559808	0.817670	55.498203
$Plane$	30	92.156005	0.746793	67.583284
$Sheep$	30	94.247457	0.820877	68.882825

Table 6.3: Total Run time of each image group in MSRC Dataset

Name of group-image	#images	Time in seconds
$Bird$	34	348.664514
Cat	24	373.634910
$Cows$	30	215.129028
Dog	30	312.903621
$Face$	30	282.047000
$Flower$	32	364.819080
$Horse$	31	427.936972
$Plane$	30	328.946860
$Sheep$	30	205.964417

Table 6.4: Total Run time of each image group in iCoseg Dataset

Name of group-image	#images	Time in seconds
002 Alaskan Brown Bear	19	273.378694
006 Red Sox Players	25	123.091477
009 Stonehenge	5	23.474778
012 Stonehenge	18	92.575113
015 Ferrari	11	85.999046
017 Agra Taj Mahal	5	27.914493
018 Agra Taj Mahal	5	32.061741
020 Pyramids	10	31.593410
021 Elephants-safari	15	67.858999
022 Goose-Riverside	30	125.441273
023 Pandas-Tai-Land	25	147.521080
025 Airshows-helicopter	12	29.975221
025 Airshows-planes	39	95.618304
026 Airshows-Huntsville	22	58.404713
028 Cheetah-National Zoo-1	15	158.205491
028 Cheetah-National Zoo-2	18	288.341946
029 Pandas-National Zoo	20	122.379261
032 Kite-Brighton kite Festival	18	75.099656
033 Kite-kitekid	10	157.997829
034 Kite-Margate Kite Festival	7	24.750165
035 Kite-Colt Park	11	35.523028
036 Gymnastics-1	6	34.626192
036 Gymnastics-2	4	17.952251
036 Gymnastics-3	6	30.970778
038 Skating-ISU	12	27.694076
039 Women Soccer Players	27	156.078715
040 Monks-LAO PDR	13	57.381250
041 Hot Balloons-Skybird	23	64.288494
042 Statue of Liberty	40	217.014995
043 Christ the Redeemer	13	68.684512
048 Windmill	17	67.754779
049 Kendo-Kendo	30	152.994766
050 Kendo-EKC 2008	11	34.333841
brown bear	5	29.915491

Chapter 7

Conclusion and Future Work

In this work, a new framework for solving the interactive co-segmentation problem has been presented. The proposed algorithm consists of one global unary function responsible for providing prior probability to each segment belonging to foreground/background region by matching each segment histogram to all the scribbled segments histograms using the Bhattacharya distance. The local function is responsible for maintaining smoothness of the outcome by providing similar probability to those segments which are similar in colour appearance. Bipartite graph further improves the segmentation quality by considering the segments which are prone to be misclassified due to similarity in both foreground/background colour regions. The experimental results show

that the proposed algorithm provide good results compared to other co-segmentation algorithm. In future the algorithm can be extended to provide co-segmentation on videos or multi class images.

Bibliography

[1] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3169–3176, June 2010.

[2] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen. Interactively co-segmentating topically related images with intelligent scribble guidance. *International Journal of Computer Vision*, 93(3):273–292, Jul 2011.

[3] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient n-d image segmentation. *Int. J. Comput. Vision*, 70(2):109–131, November 2006.

[4] B. Cheng, G. Liu, J. Wang, Zhongyang Huang, and S. Yan. Multi-task low-rank affinity pursuit for image segmentation. In *2011*

International Conference on Computer Vision, pages 2439–2446, Nov 2011.

[5] M. D. Collins, J. Xu, L. Grady, and V. Singh. Random walks based multi-image segmentation: Quasiconvexity results and gpu-based solutions. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1656–1663, June 2012.

[6] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, May 2002.

[7] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38 – 59, 1995.

[8] X. Dong, J. Shen, L. Shao, and M. H. Yang. Interactive cosegmentation using global and local energy optimization. *IEEE Transactions on Image Processing*, 24(11):3966–3977, Nov 2015.

[9] A. Faktor and M. Irani. Co-segmentation by composition. In *2013 IEEE International Conference on Computer Vision*, pages 1297–1304, Dec 2013.

[10] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, Sep 2004.

[11] H. Fu, D. Xu, S. Lin, and J. Liu. Object-based rgbd image co-segmentation with mutex constraint. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4428–4436, June 2015.

[12] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3129–3136, June 2010.

[13] D. S. Hochbaum and V. Singh. An efficient algorithm for co-segmentation. In *2009 IEEE 12th International Conference on Computer Vision*, pages 269–276, Sept 2009.

[14] Tae Hoon, Kim Kyoung, Mu Lee, and Sang Uk Lee. Nonparametric higher-order learning for interactive segmentation.

[15] A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In *2010 IEEE Computer Society Conference*

on Computer Vision and Pattern Recognition, pages 1943–1950, June 2010.

[16] A. Joulin, F. Bach, and J. Ponce. Multi-class cosegmentation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 542–549, June 2012.

[17] Pushmeet Kohli, L'Ubor Ladický, and Philip H. Torr. Robust higher order potentials for enforcing label consistency. *Int. J. Comput. Vision*, 82(3):302–324, May 2009.

[18] Z. Lou and T. Gevers. Extracting primary objects by video co-segmentation. *IEEE Transactions on Multimedia*, 16(8):2110–2117, Dec 2014.

[19] Yadong Mu and Bingfeng Zhou. Co-segmentation of image pairs with quadratic global constraint in mrfs. In *Proceedings of the 8th Asian Conference on Computer Vision - Volume Part II*, ACCV'07, pages 837–846, Berlin, Heidelberg, 2007. Springer-Verlag.

[20] L. Mukherjee, V. Singh, and C. R. Dyer. Half-integrality based algorithms for cosegmentation of images. In *2009 IEEE Conference*

on Computer Vision and Pattern Recognition, pages 2028–2035, June 2009.

[21] L. Mukherjee, V. Singh, and Jiming Peng. Scale invariant cosegmentation for image groups. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 1881–1888, Washington, DC, USA, 2011. IEEE Computer Society.

[22] Lopamudra Mukherjee, Vikas Singh, Jia Xu, and Maxwell D. Collins. Analyzing the subspace structure of related images: Concurrent segmentation of image sets. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part IV*, ECCV'12, pages 128–142, Berlin, Heidelberg, 2012. Springer-Verlag.

[23] 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*, CVPR '06, pages 993–1000, Washington, DC, USA, 2006. IEEE Computer Society.

[24] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu. Unsupervised joint object discovery and segmentation in internet images. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1939–1946, June 2013.

[25] J. C. Rubio, J. Serrat, A. Lpez, and N. Paragios. Unsupervised co-segmentation through region matching. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 749–756, June 2012.

[26] J. Shen, Y. Du, W. Wang, and X. Li. Lazy random walks for superpixel segmentation. *IEEE Transactions on Image Processing*, 23(4):1451–1462, April 2014.

[27] S. Vicente, C. Rother, and V. Kolmogorov. Object cosegmentation. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 2217–2224, Washington, DC, USA, 2011. IEEE Computer Society.

[28] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother. Cosegmentation revisited: Models and optimization. In *Proceedings of the 11th European Conference on Computer Vision: Part II*,

ECCV'10, pages 465–479, Berlin, Heidelberg, 2010. Springer-Verlag.

[29] C. Wang, Y. Guo, J. Zhu, L. Wang, and W. Wang. Video object co-segmentation via subspace clustering and quadratic pseudo-boolean optimization in an mrf framework. *IEEE Transactions on Multimedia*, 16(4):903–916, June 2014.

[30] J. Z. Wang, Jia Li, and G. Wiederhold. Simplicity: semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, Sep 2001.

[31] W. Wang, J. Shen, X. Li, and F. Porikli. Robust video object cosegmentation. *IEEE Transactions on Image Processing*, 24(10):3137–3148, Oct 2015.

[32] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*, ICCV '05, pages 1800–1807, Washington, DC, USA, 2005. IEEE Computer Society.

[33] A. L. Yuille, D. S. Cohen, and P. W. Hallinan. Feature extraction from faces using deformable templates. In *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer Society Conference on*, pages 104–109, Jun 1989.

[34] Hongyuan Zhu, Jiangbo Lu, Jianfei Cai, Jianming Zheng, and N. M. Thalmann. Multiple foreground recognition and cosegmentation: An object-oriented crf model with robust higher-order potentials. In *IEEE Winter Conference on Applications of Computer Vision*, pages 485–492, March 2014.