TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
PULCHOWK CAMPUS

THESIS NO: 072/MSI/610

Improved Salient Object Extraction Using Structured Matrix Decomposition and Contour Based Spatial Prior

by
Ramesh Bhandari

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING
LALITPUR, NEPAL

NOVEMBER, 2017

Improved Salient Object Extraction Using Structured Matrix Decomposition and Contour Based Spatial Prior

by

Ramesh Bhandari

072/MSI/610

Thesis Supervisor

Mr. Sharad Kumar Ghimire

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Information and Communication Engineering

Department of Electronics and Computer Engineering

Institute of Engineering, Pulchowk Campus

Tribhuvan University

Lalitpur, Nepal

November, 2017

COPYRIGHT©

DECLARATION

I declare that the work hereby submitted for Master of Science in Information and Communication Engineering (MSICE) at IOE, Pulchowk Campus entitled **"Improved Salient Object Extraction Using Structured Matrix Decomposition and Contour Based Spatial Prior"** is my own work and has not been previously submitted by me at any university for any academic award.

I authorize IOE, Pulchowk Campus to lend this thesis to other institution or individuals for the purpose of scholarly research.

Ramesh Bhandari
072/MSI/610
Date: November, 2017

RECOMMENDATION

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering for acceptance, a thesis entitled **"Improved Salient Object Extraction Using Structured Matrix Decomposition and Contour Based Spatial Prior"**, submitted by **Ramesh Bhandari** in partial fulfillment of the requirement for the award of the degree of **"Master of Science in Information and Communication Engineering"**.

.....

Supervisor: Mr. Sharad Kumar Ghimire,
Department of Electronics and Computer Engineering,
Institute of Engineering, Tribhuvan University

.....

External Examiner: Mr. Subhash Dhakal,
Director,
Department of Information Technology,
Government of Nepal

.....

Committee Chairperson: Dr. Dibakar Raj Pant,
Head of Department,
Department of Electronics and Computer Engineering,
Institute of Engineering, Tribhuvan University

Date: November, 2017

DEPARTMENTAL ACCEPTANCE

The thesis entitled **"Improved Salient Object Extraction Using Structured Matrix Decomposition and Contour Based Spatial Prior"**, submitted by **Ramesh Bhandari** in partial fulfillment of the requirement for the award of the degree of **"Master of Science in Information and Communication Engineering"** has been accepted as a bonafide record of work independently carried out by him in the department.

.....

Dr. Dibakar Raj Pant

Head of the Department

Department of Electronics and Computer Engineering,

Pulchowk Campus,

Institute of Engineering,

Tribhuvan University,

Nepal.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor **Mr. Sharad Kumar Ghimire** for his encouragement, suggestions and continuous guidance throughout the course of my thesis work.

I am very grateful and pay my sincere gratitude to our program coordinator **Dr. Basanta Joshi** and Head of Department **Dr. Diwakar Raj Pant** for their guidance and every bit of help during this research work. I am highly indebted to **Prof. Dr. Subarna Shakya** and **Dr. Sanjeeb Prasad Panday** for their insights and opinions regarding the thesis work.

I would like to express my heartily gratitude towards the **Institute of Engineering, Pulchowk Campus** along with all my respected teachers, my friends, my family for giving me continuous support for their invaluable help without whom I wouldn't have come so far in my thesis.

Ramesh Bhandari
072/MSI/610

ABSTRACT

Salient object detection is a useful and important technique in computer vision, which can scale up the efficiency of several applications such as object detection, object segmentation and content based image editing. In this thesis, an improved technique of salient object extraction using structured matrix decomposition and contour based spatial prior is implemented. In order to assist efficient background-salient objects separation, structured matrix decomposition model with two structural regularizations namely, tree-structured sparsity-inducing regularization and Laplacian regularization are used. Tree-structured sparsity-inducing regularization captures the image structure and enforces the same object to have similar saliency values and Laplacian regularization enlarges the gap between background and the salient object. In addition, integrating general high level priors and contour based spatial prior obtained from biologically inspired model is implemented to improve the efficiency of saliency related tasks. The performance of the proposed method is evaluated on three demanding datasets, namely, MSRA10K, ICOSEG and PASCAL-S, including single object, multiple object and complex scene images. For MSRA10K dataset precision-recall curve of proposed method starts from 0.98 and follows top and right-hand border more than structured matrix decomposition which starts from 0.95 and similarly structural similarity is also more by 0.137. Similarly, both visual and comprehensive evaluation using receiver operating characteristics curve and mean absolute error for other datasets also shows improved result.

Keywords: Salient Object Detection, Structured Matrix Decomposition, Low Rank, Contour Based Spatial Prior, Salient Object Extraction

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AER	: Average Edge Response
AUC	: Area Under Curve
CBSP	: Contour Based Spatial Prior
FPR	: False Positive Rate
LGN	: Lateral Geniculate Nucleus
MAE	: Mean Absolute Error
MATLAB	: Matrix Laboratory
OR	: Overlap Ratio
PR	: Precision-Recall
ROC	: Receiver Operating Characteristics
SLIC	: Simple Linear Iterative Clustering
SMD	: Structured Matrix Decomposition
SSIM	: Structure Similarity
TPR	: True Positive Rate
WF	: Weighted F-measure

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Determining visual saliency has been a fundamental research problem in vision perception for a very long period of time. It alludes to the recognition of crucial visual information for further processing. Identifying and segmenting the most conspicuous object from the scene referred to as salient object detection and salient object extraction is an important branch of visual saliency. Our society has become more technologically advanced than ever but our most advanced machines and computers still struggle at describing what it sees in series of photos. This is due to fact that human visual system has innate capability to extract crucial information from a scene but for machine same task is difficult. Little by little, we are giving sight to the machines or computers using state-of-the-art methods from computer vision. Computer vision is one of the most frontier and potentially revolutionary technologies in computer science. Most of the tasks in computer vision relies on object detection, object extraction and object recognition. During detection and recognition of object, identifying object-of-interest plays an important role for the real world computer vision applications.

Salient objects are dominant objects in a given image, for example: animals, people, flowers, or any other objects. Salient objects can be extracted successfully if it can be located and detected correctly. Salient object detection is an important branch of visual saliency and mainly deals with identifying the most noticeable objects from an images. Due to its wide range of applications in computer vision, such as object detection, object extraction, object recognition and automatic image editing, it has grabbed attention over last decade. To accomplish the task of salient object detection many saliency models have been proposed. Based on whether prior knowledge is used or not, current models fall in two classes, namely, bottom-up and top-down. Bottom-up models [1] and [2] are based on low level

features such as color, texture, location etc. The main downside of bottom-up methods is detected salient regions may only contain parts of salient object. On the other side, top-down salient object detection methods [3], [4] and [5] use high level human perceptual knowledge to identify potential region of salient object. These high level human perceptual knowledge are generally context, semantics and background knowledge which guide the saliency related tasks. However, efficiency of such models suffer from diversity of object types which are encountered in real world applications. Recent development for salient object detection shows that trend is to combine bottom-up cues with top-down cues to increase the efficiency of saliency related tasks.

This research work mainly focuses on saliency related tasks, salient object detection and salient object extraction, and improved salient object extraction is implemented using structured matrix decomposition (SMD) and contour based spatial prior (CBSP).

1.2 Problem Definition

Object detection and extraction from images are fundamental tasks in real world computer vision applications. One of the most challenging task in object detection and extraction is shape and size of the object that varies from application to application. And, another challenging task is complex-scene image in which it is difficult to locate salient objects. So, powerful and computationally feasible methods are required to accomplish this task which are invariant to size and features of objects. So to overcome this problem efficient salient object extraction using structured matrix decomposition and contour based spatial prior is proposed. Proposed approach will be invariant to size, shapes and features of objects.

1.3 Objectives

The prime objective of this research work are:

- To detect salient object automatically using structured matrix decomposition and contour based spatial prior.
- To extract salient object from detected region.

1.4 Scope of the Work

This research will have application in following fields:

- **Computer Vision**

  Most of the important tasks in computer vision depends on object detection and object recognition. Salient object extraction using structured matrix decomposition and contour based spatial prior can make the process of detecting salient object from an images will be a lot easier. After detecting the salient object this work can also extract the salient object from images which can be used for object recognition.

- **Automatic Image Editing**

  This research work automatically extracts the salient object from an images which can be subjected to different background as required changing their appearances.

- **Feature Extraction**

  Different image features such as object-of-interest, blobs, edges etc. can be extracted with little modification using this work which can be used for supervised and unsupervised learning in artificial neural networks.

CHAPTER 2

LITERATURE REVIEW

In the past, to detect salient object many saliency models have been proposed. Based on whether prior knowledge is used or not, current methods fall in two classes, namely, bottom-up model and top-down model.

Bottom-up models [1] - [2] are based on low level features such as color, texture and location. The main limitation of these methods are that detected salient regions may only contain parts of the target objects, or can be easily mixed with background representing background as salient objects.

On the other hand, top-down models [3] - [5] are based on high-level human perceptual knowledge, such as context, semantics and background priors, to guide the subsequent saliency computation. However, generalization and scalability of these models suffers from high diversity of object types or when image contains complex scene.

Fixation prediction methods proposed by L. Itti, C. Koch, and E. Niebur [1] mainly focus on high contrast boundaries but ignores object surfaces and shapes. In contrast, methods proposed by Q. Yan, L. Xu, J. Shi, and J. Jia [4] may lacks its performance when there are no dominant objects in the scene.

Recent state-of-the-art method for saliency related task shows that trend is to combine bottom-up cues with top-down cues [6],[7],[8],[9],[10] and [11] to assist efficient salient object detection. Low-rank recovery models have shown potential for salient object detection, where matrix is decomposed into a low rank matrix representing image background and a sparse matrix identifying salient objects. Methods [6]- [11] are based on low rank matrix recovery theory [12]. Performance of research work [6] - [10] suffers when there are similarities between the salient objects and background. Peng et. al. [11] proposes salient object detection via structured matrix decomposition to overcome these problems which shows its decent accuracy for complex scene images.

Shen and Wu [7] put forward a unified LR model (ULR) with feature transformation to combine traditional low-level features with high-level prior knowledge. In which an image is represented as a low-rank matrix plus sparse noises in a certain feature space, where the non-salient regions (or background) can be explained by the low-rank matrix, and the salient regions are indicated by the sparse noises.

Lang et al. [8] present a low-rank representation (LRR) [13] based multi-task learning method, in which top-down priors are weighted and combined with multiple features to estimate saliency collaboratively. Zou et al. [9] introduce the segmentation priors derived from image background and boundary cues to assist the low-rank matrix recovery.

A representative series of research work [6] - [10] lack performance when there are similarities between the salient objects and background or when background is complicated. Generally, these LR-based saliency detection methods assume that an image can be represented as a combination of a highly redundant information part (e.g., visually consistent background regions) and a sparse salient part (e.g., distinctive foreground object regions). The redundant information part usually lies in a low dimensional feature subspace, which can be approximated by a low-rank feature matrix representing background. In contrast, the salient part deviating from the low-rank subspace can be viewed as noise or errors, which are represented by a sparse sensory matrix representing salient object in the given scene. Peng et. al. [11] suggests salient object detection via structured matrix decomposition to overcome these snags. However, this method fails to suppress some small background region with distinctive appearances because feature vectors of those regions are not in low dimensional subspace and may incorrectly highlighted as foreground salient objects.

In this research work, integrating contour based spatial prior obtained by using biologically inspired model with method proposed in [11] is implemented to improve the task of detection which ultimately results better for salient object extraction.

CHAPTER 3

METHODOLOGY

3.1 Block Diagram

To carry out series of image processing operation required to an input image to detect and extract salient object from the given image, following block diagram shown in Figure 3.1 has been formulated.

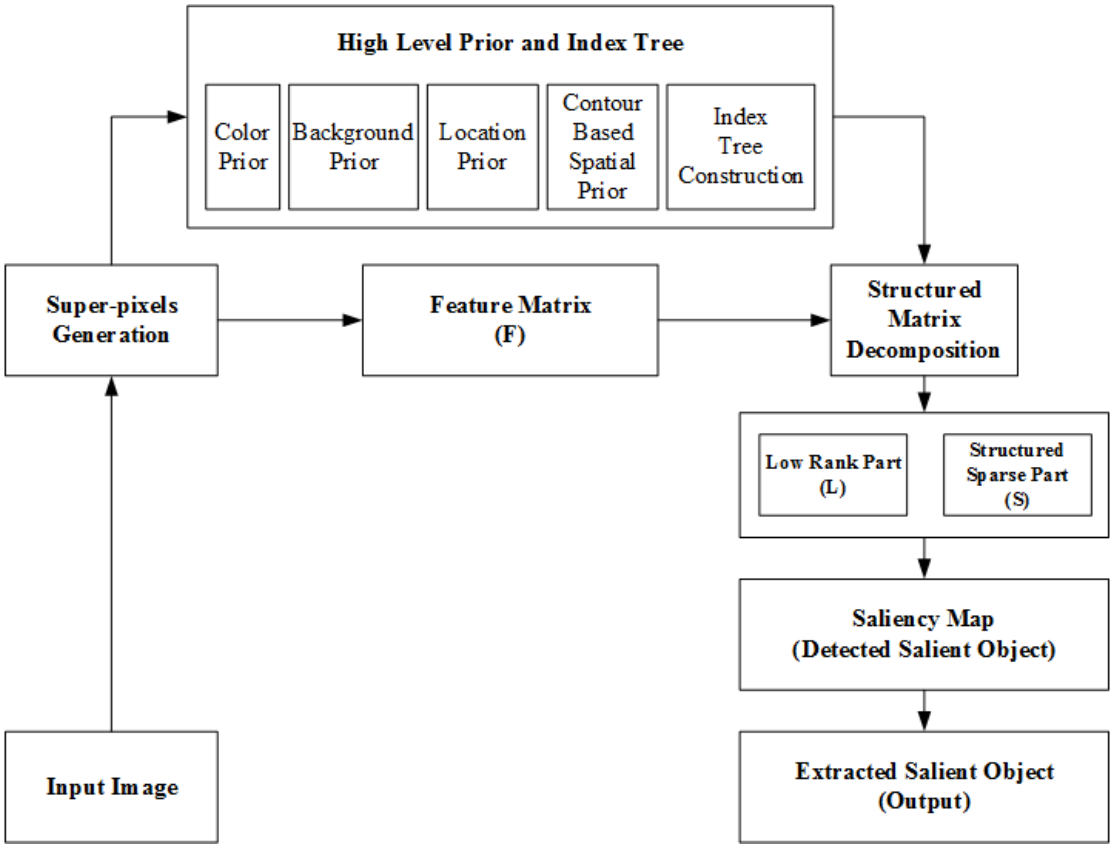

Figure 3.1: Block Diagram

3.2 Data Collection

Images required for this research work would be captured through a digital camera and these images are converted to appropriate format like .bmp, .tiff, .png, .jpg according to needs. To validate final results, standard datasets images are used

from following datasets:

(a) MSRA10K Salient Object Database [14].

(b) CMU-Cornell ICOSEG Dataset [15].

(c) PASCAL-S Dataset [16].

3.3 Procedure of Work

In this research work, salient object is defined as a dominant object (e.g., people, animals, cars, flowers, or any other structures that is dominant on image) in the given image. Having an input image, input image is first partitioned to perceptually homogeneous elements based on low rank matrix recovery model for salient object detection [7]. After partitioning, feature matrix (F) is computed which consists low level features and then simple linear iterative clustering (SLIC) algorithm [17] is performed to over-segment the image to generate super-pixels. On the other hand, high level features like - color prior, location prior, background prior, contour based spatial prior are computed and index tree is also constructed to encode structure information. Graph based segmentation algorithm [18] is applied to merge spatially neighboring patches. After obtaining feature matrix (F) and index tree, structured matrix decomposition is applied to decompose feature matrix (F) into low-rank part (L) and structured sparse part (S). After decomposition, saliency map is calculated by transferring structured sparse part (S) from feature domain to spatial domain. Finally, with the help of original image salient object is extracted by making decision whether a pixel corresponding to original image falls in detected salient object or not.

3.3.1 Structured Matrix Decomposition

Given an input image I, it is first partitioned into N patches (super pixels) $P = P_1, P_2, P_3, ...P_N$. For an input image shown in Figure 3.2 generated super-pixels using simple linear iterative clustering [17] is shown in Figure 3.3.

For each super pixel P_i a D dimensional feature (here $D = 53$) vector is extracted

Figure 3.2: Input Image

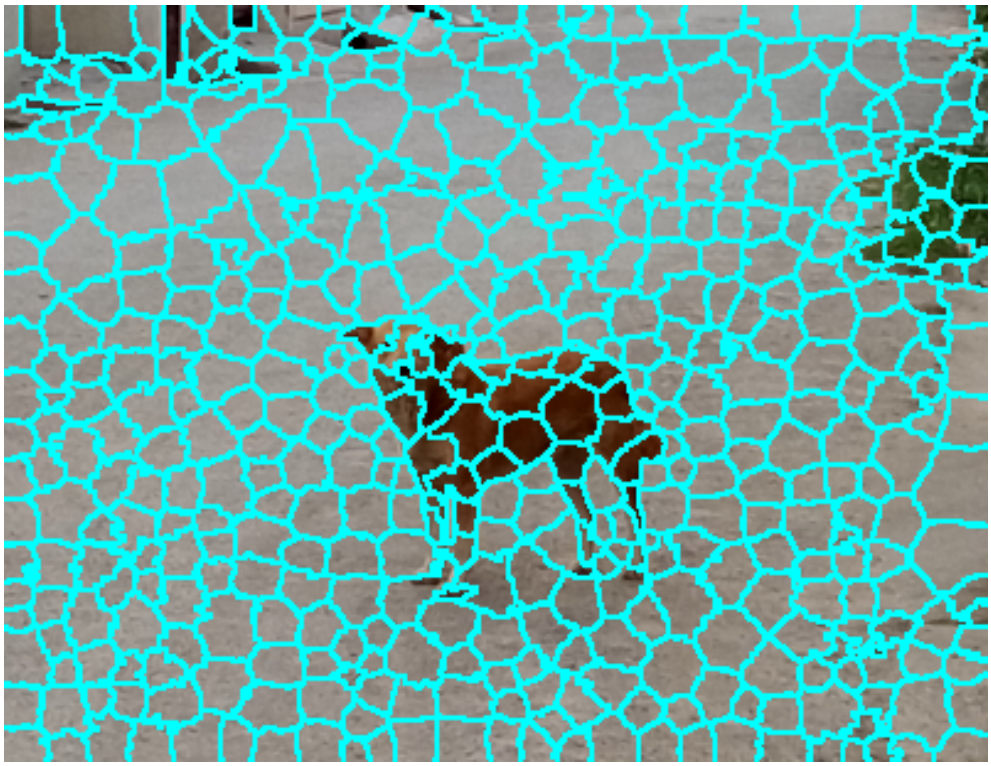

Figure 3.3: Generated Super-pixel for Input Image

and denoted as f_i. Feature vector forms a matrix representation of I, denoted as $F = f_1, f_2, f_3.., f_N$. Feature matrix (F) consists of low level feature including RGB color, steerable pyramids [19] and Gabor filter [20] to construct 53-dimensional feature representation. On top of super-pixels, an index tree is constructed to encode structure information via hierarchical segmentation. Then graph based segmentation algorithm [18] is applied to merge spatially neighboring patches. The problem of salient object detection is to design an effective model to decompose the feature matrix F into a redundant information part L which indicate non-salient background part and a structured distinctive part S which indicates salient background. In this research, when both the feature matrix (F) and index tree are ready, structured matrix decomposition model is applied as proposed in [11] as follows:

$$\min_{L,S} \Psi(L) + \alpha\Omega(S) + \beta\Theta(L,S) \;\; s.t. \;\; F = L + S \tag{3.1}$$

Where $\Psi(.)$ is a low-rank constraint to allow identification of the intrinsic feature subspace of the redundant background patches, $\Omega(.)$ is a structured sparsity regularization to capture the spatial and feature relations of patches in S, $\Theta(.,.)$ is an interactive regularization term to enlarge the distance between the L and S, and α, β are positive trade-off parameters. Similarly, three interactive regularizations are performed as proposed in [11] they are: Low-rank regularization for image background, Structured-sparsity regularization for salient objects and Laplacian regularization to enlarge the gap between L and S.

The structured matrix based salient object detection is further extended to integrate high level priors and contour based spatial prior. Three types of high level priors are used, namely, location prior, color prior and background priors. Specifically, the location prior is generated by Gaussian distribution based on the distance of the pixels from an image center and it is denoted by *lp*. The color prior used here is same as [7], which measures human eye sensitivity to red and yellow color. Computed color prior is denoted by *cp*. The background prior calculates the probabilities of image regions connected to image boundaries [21] and is denoted by *bp*. These three high level priors are combined by taking weighted sum to get

high level prior denoted by hp as follows:

$$hp = w_1 * lp + w_2 * bp + w_3 * cp \tag{3.2}$$

Where, w_1, w_2 and w_3 are weight given to each prior. Values of w_1, w_2 and w_3 lies between 0 and 1, and satisfies the following relations:

$$w_1 + w_2 + w_3 = 1 \tag{3.3}$$

Contour based spatial prior obtained from biologically inspired model is also computed later and integrated with obtained high level prior (hp) in Equation (3.2) to get final high level prior map (fp).

3.3.2 Contour Based Spatial Prior

To estimate contour based spatial prior, first, the edge response and corresponding Orientations is estimated by using, biologically inspired method, efficient color

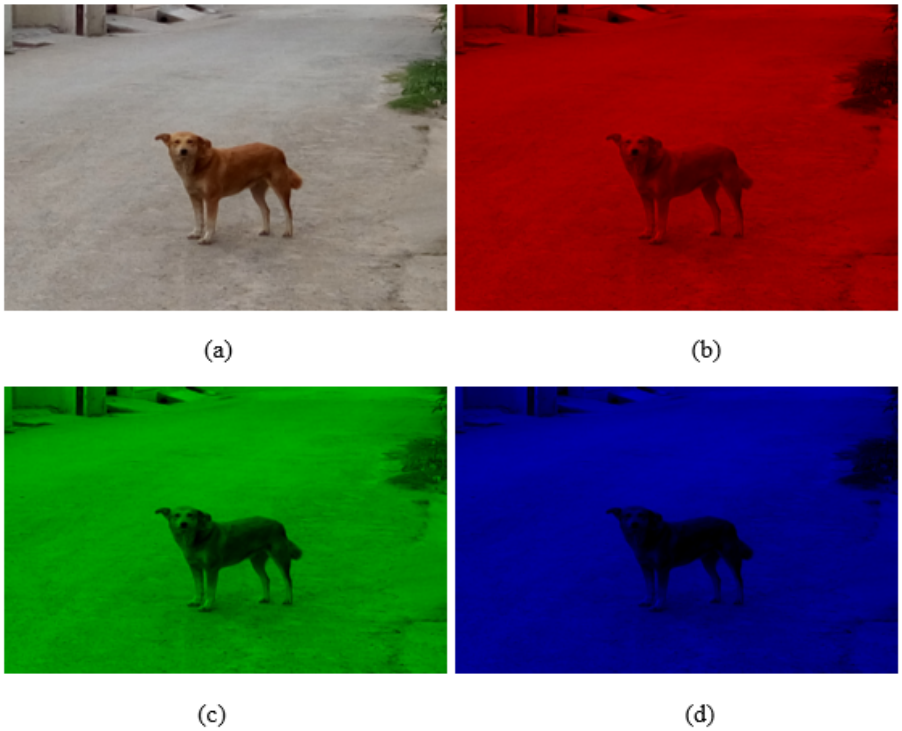

Figure 3.4: Original image and Red, Green and Blue component of an image. (a) Original image. (b) Red component. (c) Green component. (d) Blue component.

boundary detection with color-opponent mechanisms proposed in [22] and contour based spatial prior is computed using method proposed in [23].

Efficient color boundary detection with color-opponent mechanisms imitate the working of human visual system to identify the edges from an images. This mechanism involves processing of an image information in three layers, namely, Cone layer, Ganglion/Lateral Geniculate Nucleus (LGN) layer, and Cortex layer, of our visual system.

In the Cone layer, three color components namely red, green and blue from an input color image is extracted and denoted by r,g and b respectively. These three channels are shown in Figure 3.4. To implement color opponent mechanisms, yellow component shown in Figure 3.5 of a color image is computed by taking average of red and green component as follows:

$$y = \frac{r + g}{2} \tag{3.4}$$

Figure 3.5: Computed Yellow component of original image.

In order to obtain local color information, extracted four components from an

image are subjected to Gaussian filters which simulates the receptive fields of the cones in the retina as suggested by [24] and [25]. Here, all the four components are smoothed with Gaussian filters having same standard deviation (σ). The output of the four components after simulated by Gaussian filters are output of Cone layer and are denoted by R_g, G_g, B_g, and Y_g. Now the outputs of Cone layer are passed to Ganglion/Lateral Geniculate Nucleus (LGN) layer.

In the Ganglion/Lateral Geniculate Nucleus (LGN) layer, single opponent mechanism is implemented. Ganglion and Lateral Geniculate Nucleus layers are implemented in single layer because these cells have similar receptive field properties. Response of this layer to Cone layer outputs is mathematically described as follows:

For R-G channel:

$$S(x, y) = w_1.R_g(x, y; \sigma) + w_2.G_g(x, y; \sigma) \tag{3.5}$$

For B-Y channel:

$$S(x, y) = w_1.B_g(x, y; \sigma) + w_2.Y_g(x, y; \sigma) \tag{3.6}$$

Where,

$$w_1 w_2 \leq 0 \ \ and \ \ |w_1|, |w_2| \in [0, 1] \tag{3.7}$$

Here, w_1 and w_2 are connection weights from cone layer cells to Ganglion/Lateral Geniculate Nucleus (LGN) layer. Weight w_1 and w_2 have always opposite sign. In single opponent mechanisms, for R-G channel, with $w_1 < 0$ and $w_2 > 0$ obtained response is of R off /G on cells, and with $w_1 > 0$ and $w_2 < 0$ obtained response is of R on /G off cells. And similarly, for B-Y channel, with $w_1 < 0$ and $w_2 > 0$ obtained response is of B off /Y on cells, and with $w_1 > 0$ and $w_2 < 0$ obtained response is of B on /Y off cells. In general, boundaries in four channels $(S_{rg}, S_{gr}, S_{by}, S_{yb})$ are calculated as:

$$S_{rg} = R_g + w.G_g \tag{3.8}$$

$$S_{gr} = w.R_g + G_g \tag{3.9}$$

$$S_{by} = B_g + w.Y_g \tag{3.10}$$

$$S_{yb} = w.B_g + Y_g \tag{3.11}$$

Where $w \in [-1, 0]$ is cone input weight and other weight is set to 1.

Outputs (S_{rg}, S_{gr}, S_{by} and S_{yb}) of Ganglion/Lateral Geniculate Nucleus (LGN) layer is passed to Cortex layer and then double opponent mechanisms is implemented in this layer. In the cortex layer of V_1, the receptive fields of most color- and color-luminance-sensitive neurons are both chromatically and spatially opponent [22]. In particular, the oriented double-opponent cells are considered to play an important role in color boundary detection [26]. In this layer boundary is detected by using set of filters having orientation $\theta \in [0, 2\pi]$. Boundary responses at each orientation is calculated using,

$$\begin{aligned} D(x, y; \theta_i) = & \sum_{m,n \in N_{r+g-}} S_{r+g-}(x+m, y+n) * RF(m, n; \theta_i) \\ & + \\ & \sum_{m,n \in N_{r-g+}} S_{r-g+}(x+m, y+n) * RF(m, n; \theta_i) \end{aligned} \tag{3.12}$$

Here, $S_{r-g+} = -S_{r+g-}$, N_{r-g+} and N_{r+g-} are the R off/G on and R on/G off neurons in the V1 region. Where θ_i is given as:

$$\theta_i = \frac{2(i-1)\pi}{N_\theta} \tag{3.13}$$

And $RF(x, y; \theta_i)$ is determined by using following equations,

$$RF(x, y; \theta_i) = \left| \frac{\partial f(\tilde{x}, \tilde{y})}{\partial \tilde{x}} \right| \tag{3.14}$$

$$f(x,y) = \frac{1}{\sqrt{(2\pi(k\sigma^2))}} exp\left(\frac{-(\tilde{x}^2 + \gamma^2\tilde{y}^2}{2(k\sigma^2)}\right) \tag{3.15}$$

$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} = \begin{pmatrix} xcos(\theta) + ysin(\theta) \\ -xsin(\theta) + ycos(\theta)\end{pmatrix} \tag{3.16}$$

γ in Equation (3.15) is the spatial aspect ratio of Gaussian that controls the ellipticity of receptive field. Based on physiological findings [27] and [28], value of γ is generally taken as 0.5. Similarly, product of k and σ determines size of V_1 neurons in cortex layer.

After calculating response at different orientation, maximum response from each orientation is calculated as follows:

$$D(x,y) = max\{D(x,y;\theta_i)|i = 1,2,....N_\theta\} \tag{3.17}$$

Now output of each channel after taking maximum at N_θ orientation are represented by D_{rg}, D_{gr}, D_{by} and D_{yb}. These outputs are normalized linearly. Final response of cortex layer is calculated by taking maximum response at each channel, which is done as follows:

$$r(x,y) = max\{D_{c_i}(x,y;\theta_i)|c_i \in D_{rg}, D_{gr}, D_{by}, D_{yb}\} \tag{3.18}$$

$r(x,y)$ is the output of color opponent mechanisms for color boundary extraction. $r(x,y)$ gives the detected color boundary which is shown in Figure 3.6.

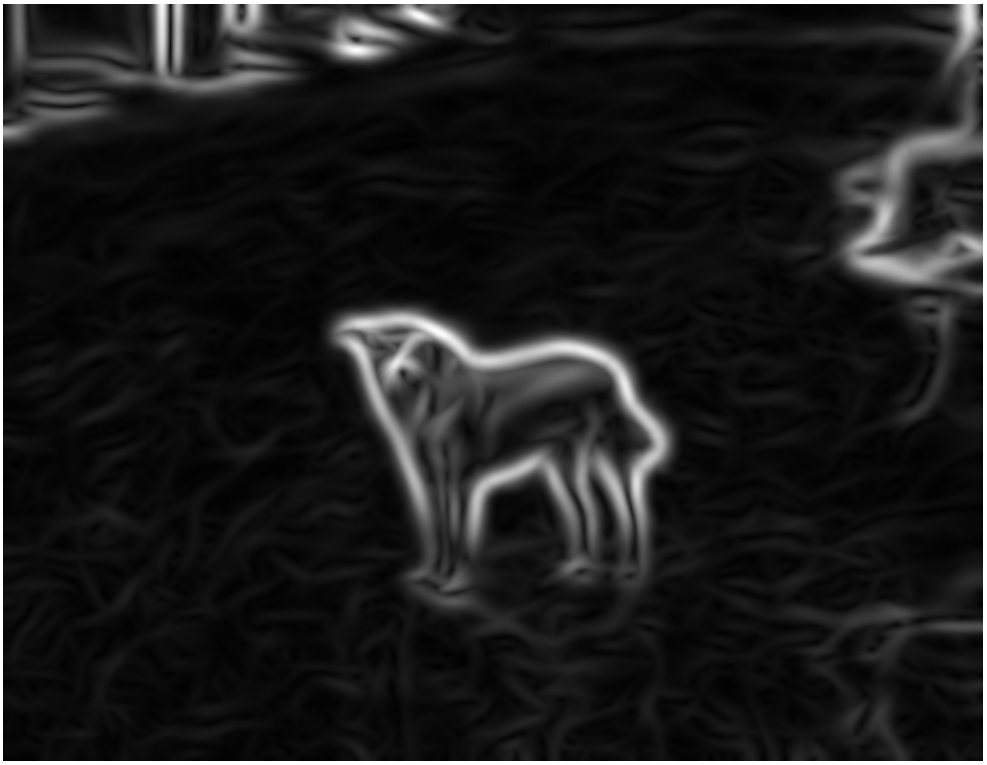

Figure 3.6: Detected color boundaries or edges using color opponent mechanisms.

Figure 3.7: Obtained edges using color opponent mechanisms and non-maximum suppression.

Figure 3.8: Dominant edge obtained after thresholding.

Detected color boundaries or edges are further processed using non-maximum suppression for thinning edge. Edges in the original images after applying non-maximum suppression technique is shown in Figure 3.7.

Obtained edges after color opponent mechanisms and non-maximum suppression are further thresholded to find dominant edges only (shown in Figure 3.8). In this research work, threshold value 0.33 is used (assuming all the pixel's values are between 0 and 1). After obtaining dominant edge response and corresponding orientations, for each edge pixel, Average Edge Response (AER) is calculated in the left and right half disk around it with disk radius $d_r = min(W, H)/3$. Where W and H are width and height of given image. Then simple voting is carried out to compute rough spatial weights of saliency S_e. Voting here is: all of the pixels within half disk having higher average edge response between two half disks are voted 1, and the pixels in the other half are voted 0.

Computed saliency weight using average edge response is shown in Figure 3.9 (b). To obtain final contour based spatial prior, center-bias weighting [29], [30] modeled by Gaussian masks is also considered with standard deviation $\sigma_c = d_r$ and this saliency is represented by S_c.

Computed center-bias weight by considering $\sigma_c = d_r$ is shown in Figure 3.9 (c). Using saliency measure S_e and S_c, contour based spatial prior denoted by cbsp is obtained by:

$$cbsp = S_e + S_c \tag{3.19}$$

and shown in Figure 3.9(a).

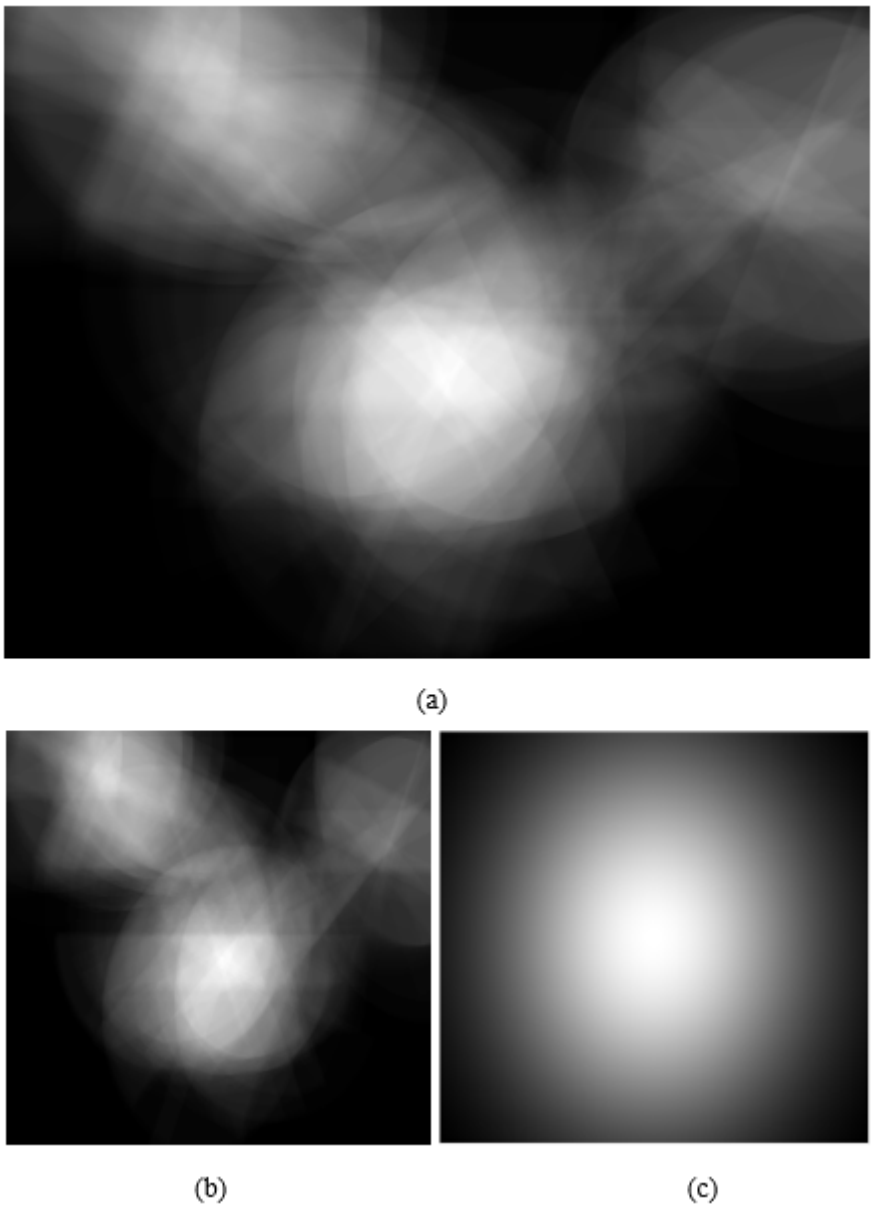

Figure 3.9: Computed contour based spatial prior, saliency and center bias weight (a) Contour based spatial prior. (b) Saliency weight obtained by calculating average edge response. (c) Center bias weight with standard deviation $\sigma_c = min(W, H)/3$.

3.3.3 Integrating Contour Based Spatial Prior

Since contour based spatial prior is determined pixel wise while high level priors are computed for super-pixels. To integrate contour-based spatial prior ($cbsp$) with high level prior (hp) obtained earlier in Equation (3.2), generated N patches (super-pixels) are mapped over contour based spatial prior and average value of all the pixels value within each patch is calculated to obtain final contour based spatial prior. Final contour based spatial prior is denoted by $fcbsp$. Mapped N patches (super-pixels) over contour based spatial prior for input image shown in Figure 3.2 is shown in Figure 3.10.

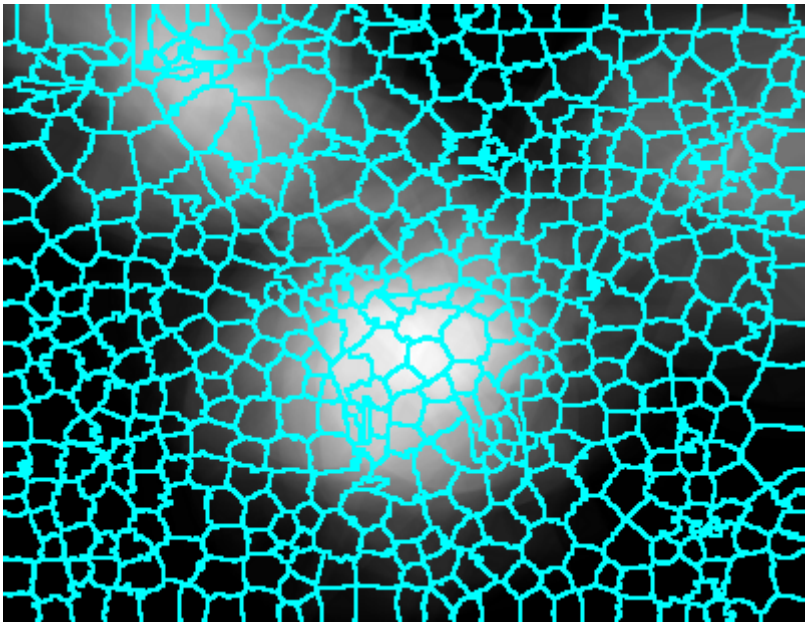

Figure 3.10: Mapped N patches over contour based spatial prior.

After finding final contour based spatial prior ($fcbsp$), this prior is combined with previously obtained high level prior (hp) by taking weighted sum to get final high level prior denoted by fp as follows:

$$fp = m_1 * hp + m_2 * fcbsp \tag{3.20}$$

Where, m_1 and m_2 are weight given to each prior. Values of m_1 and m_2 lies between 0 and 1, and satisfies the following relations:

$$m_1 + m_2 = 1 \tag{3.21}$$

Final high level prior, $fp \in [0, 1]$ for each patch P_i indicates the likelihood that P_i belongs to a salient object based on high level information. This prior is encoded into the structured matrix decomposition by multiplying each component in feature matrix (F) to guide matrix decomposition.

3.3.4 Salient Object Detection

After decomposition of feature matrix (F) into low-rank part (L) and structured sparse part (S), saliency map is calculated by transferring structured sparse part (S) from feature domain to spatial domain [11]. For an input image shown in Figure 3.2 detected salient object is shown in Figure 3.11.

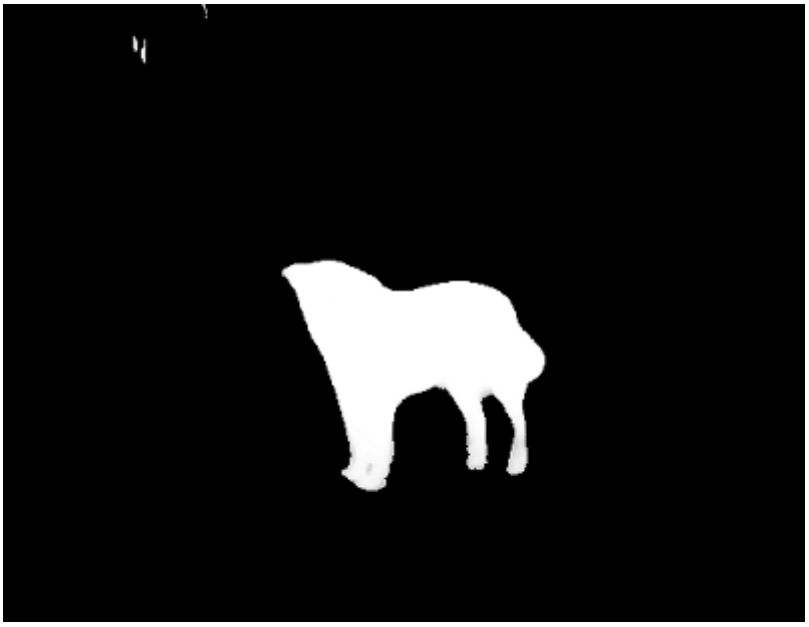

Figure 3.11: Detected Salient Object (Saliency map).

3.3.5 Salient Object Extraction

Finally, salient object is extracted by making decision whether a pixel corresponding to original image falls in detected salient object (saliency map) or not. This can be done by taking each pixel from original image and checking whether that pixel belongs to detected salient region or not. If that pixel belongs to detected salient region, then value of that pixel is taken otherwise some standard color is set which will be background color (black color is used in this research) in extracted image.

After performing this calculation obtained final extracted object for input image shown in Figure 3.2 is shown in Figure 3.12. In this research work, obtained saliency map is thresholded to 0.33 (assuming all the pixels values are between 0 and 1) to remove weakly detected region in extracted image.

Figure 3.12: Extracted Salient Object.

3.4 Tools

Various programming environment like Matrix Laboratory (MATLAB), C++, Python etc. can be used to implement the different computational tasks in this research work. MATLAB and C++ is used to carry out all the image processing operations required for the implementation of this research work. All the images used in this report are obtained by the implementation of this project in MATLAB.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Experimental Setup

To evaluate the performance of proposed method to detect and extract salient object from an image, series of experiments are conducted using different images and standard datasets involving various scenes. In this experiment evaluation is conducted on different images having single object images, multiple object images and complex scene images from different datasets, namely, MSRA10K [14], ICOSEG [15] and PASCAL-S [16]. Experimental analysis is also performed on MSRA10K [14], ICOSEG[15] and PASCAL-S [16] datasets to evaluate metrics like receiver operating characteristic (ROC) curve, precision recall (PR) curve and Mean Absolute Error (MAE) for comprehensive evaluation. The detailed characteristics of different datasets are tabulated in Table 4.1.

Table 4.1: Characteristics of Datasets.

Name of Datasets	**Number of Images**	**Characteristics**
MSRA10K [14]	10,000	Single and Multiple Object Images, Simple Background, High Contrast Images
ICOSEG [15]	643	Single and Multiple Object Images, Various Numbers Objects with Different Sizes
PASCAL-S [16]	850	Complex Scene Natural Images, Various Object Categories

Different parameters for the implementation of this research work are set as follows. While computing contour based spatial prior, in color opponent mechanisms, value of sigma (σ) is set to 1.5, cone input weights are set to -0.6 and 1, and number of orientation for color opponent mechanisms are set to 8. Similarly, value of γ is chosen to be 0.5 based on physiological findings [27] and [28].

Similarly, while integrating high level priors and contour based spatial prior values

of weight w_1, w_2 and w_3 while finding high level prior are set to $\frac{1}{3}$ and value of weights m_1 and m_2 while integrating high level prior with final contour based spatial prior is set to $\frac{1}{2}$. Which indicates equal priority is given to high level prior and final contour based spatial prior.

4.2 Evaluation Metrics

Widely used metrics including precision-recall (PR) curve, receiver operating characteristics (ROC) curve and mean absolute error (MAE) are used to evaluate the overall performance of the proposed method comprehensively.

In precision-recall curve, precision measures the fraction of the salient pixels that are correctly assigned while recall measures the fraction of ground truth salient pixels which are correctly labeled as salient object pixels. For each obtained salient map, binary maps are obtained by binarizing the saliency map with fixed thresholds ranging from 0 to 255. On a dataset, then the averaging precisions and recalls for different thresholds are computed to plot precision-recall curve.

In addition to precision-recall curve, receiver operating characteristics curve is also plotted. The receiver operating characteristics curve is generated using true positive rates (TPR) and false positive rates (FPR). Receiver operating characteristics curve is plot of true positive rates versus false positive rates. The false positive rate computes the fraction of true background pixels which are erroneously labeled as salient object pixels while true positive rate computes the fraction of true salient object pixels labeled as salient object pixels. For each obtained salient map, binary maps are obtained by binarizing the saliency map with fixed thresholds ranging from 0 to 255 to compute mean false positive rates and mean true positive rates, then receiver operating characteristics curve is plotted.

Besides PR curves and ROC curves, mean absolute error (MAE) is also computed for quantitative analysis, which considers the saliency assignment for the non-salient pixels. Which means it provides an indication of all the pixels which are erroneously marked as salient object pixels. Mean absolute error is obtained by calculating the mean difference between the saliency map and the ground truth in

pixel level.

To evaluate structural similarity between ground truth and detected salient object, structural similarity index (SSIM) is used. Structural similarity index measures the similarity between two images. Structural similarity index is computed for both methods by comparing saliency map obtained by structured matrix decomposition with ground truth and saliency map obtained by structured matrix decomposition and contour based spatial prior with ground truth.

4.3 Experimental Results

4.3.1 Results on Single Object Images

Visual analysis is subjective way to evaluate performance of saliency related tasks. Figure 4.1 shows the performance of proposed method to detect and extract salient

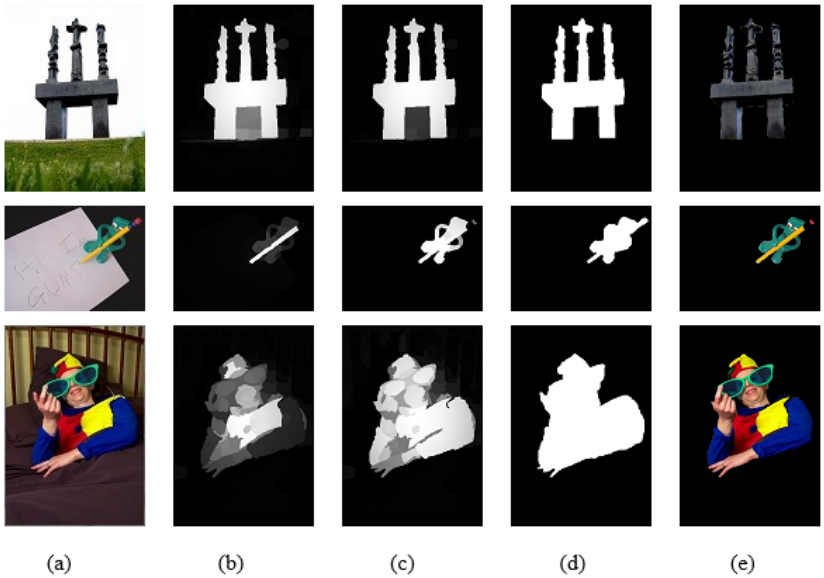

Figure 4.1: Results on single object images from MSRA10K dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.

object from MSRA10K [14] dataset images containing single object on it along with output of structured matrix decomposition method for object detection. Similarly, Figure 4.2 shows the performance of proposed method to detect and extract salient

object from ICOSEG [15] dataset images containing single object on it along with output of structured matrix decomposition method for object detection. For single object images, proposed method accurately detects and extracts entire salient object and shows visual improvement over structured matrix decomposition method. Images used in this experiment are taken from MSRA10K [14] and ICOSEG [15] datasets.

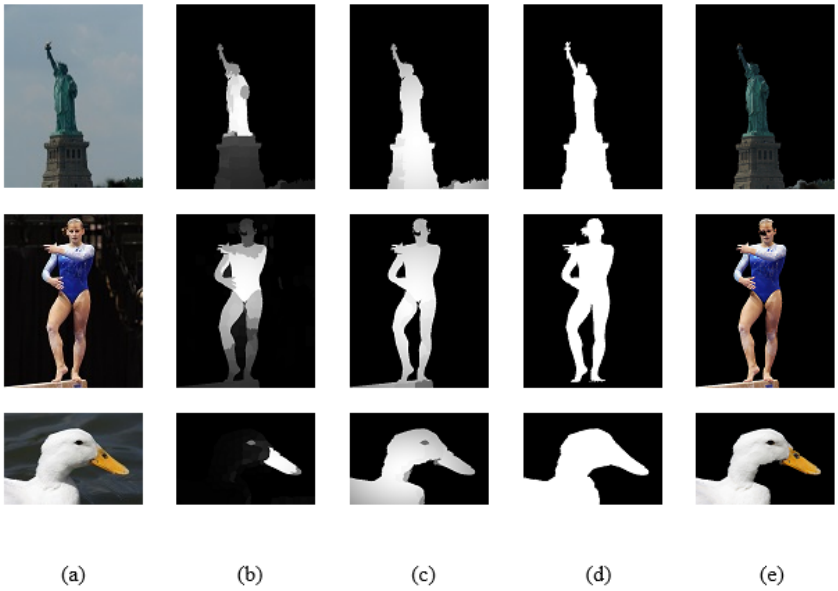

Figure 4.2: Results on single object images from ICOSEG dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.

4.3.2 Results on Multiple Object Images

Like visual analysis on single object images, experiment is further extended for multiple object images. Figure 4.3 shows the performance of proposed method to detect and extract multiple salient object from MSRA10K [14] dataset images containing multiple object on it along with output of structured matrix decomposition method for object detection. Similarly, Figure 4.4 shows the performance of proposed method to detect and extract multiple salient object from ICOSEG [15] dataset images containing multiple object on it along with output of structured matrix decomposition method for object detection. For multiple object images,

proposed method gives satisfactory visual result. Images used in this experiment are taken from MSRA10K [14] and ICOSEG [15] datasets.

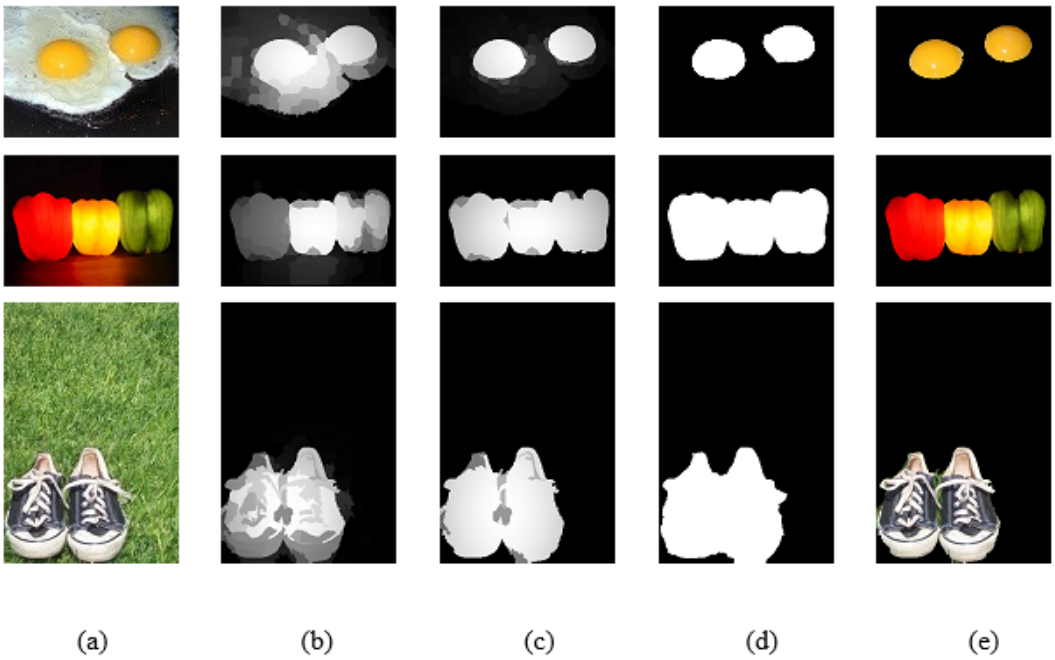

Figure 4.3: Results on multiple object images from MSRA10K dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.

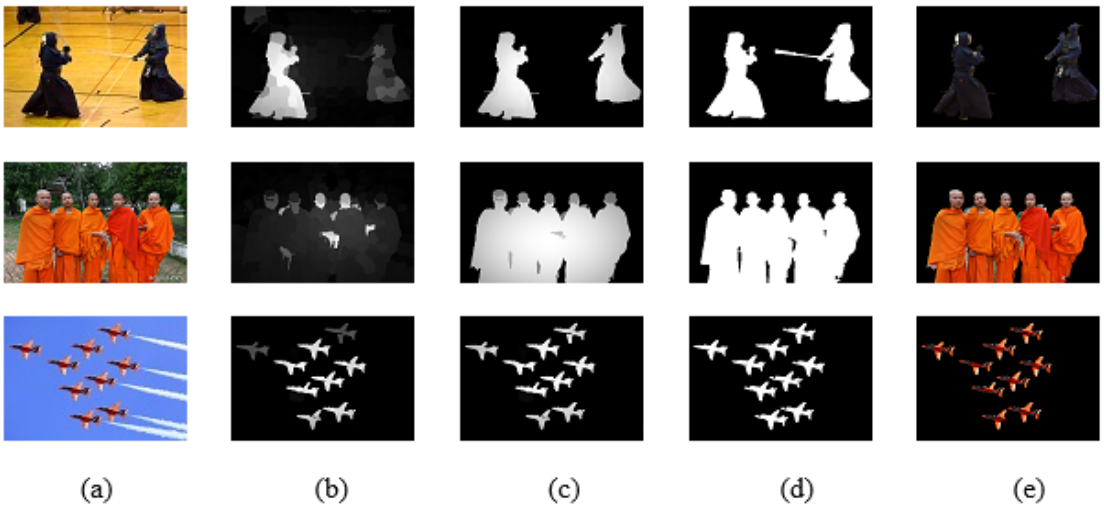

Figure 4.4: Results on multiple object images from ICOSEG dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.

4.3.3 Results on Complex Scene Images

Visual analysis is further extended for complex scene images. For this experiment PASCAL-S [16] dataset is used which contains complex scene images. Figure 4.5 shows the performance of proposed method to detect and extract object from PASCAL-S [16] dataset images containing complex scene on it along with output of structured matrix decomposition method for object detection.

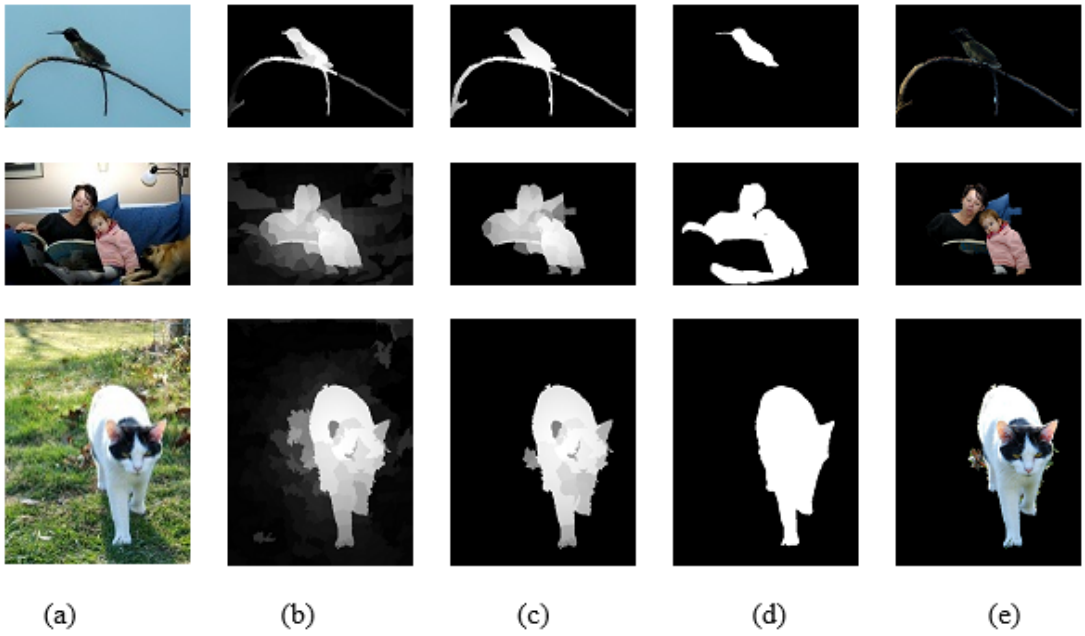

Figure 4.5: Results on complex scene images from PASCAL-S dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.

4.3.4 Evaluation Using ROC Curve

For comprehensive evaluation, receiver operating characteristics (ROC) curve for three different datasets, namely, MSRA10K [14], ICOSEG [15] and PASCAL-S [16] is used. For each dataset, ROC curve is plotted for both result obtained by structured matrix decomposition [11] and proposed method structured matrix decomposition with contour based spatial prior. Receiver operating characteristics curve for MSRA10K [14] dataset is shown in Figure 4.6, for ICOSEG [15] dataset is shown in Figure 4.7 and for PASCAL-S [16] dataset is shown in Figure 4.8.

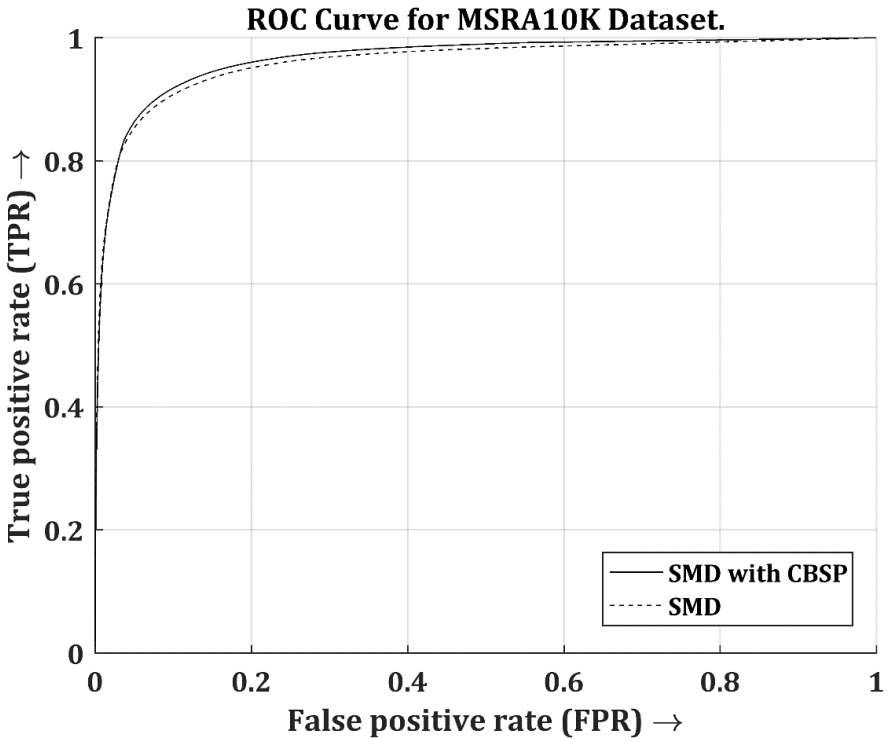

Figure 4.6: ROC curve for MSRA10K dataset.

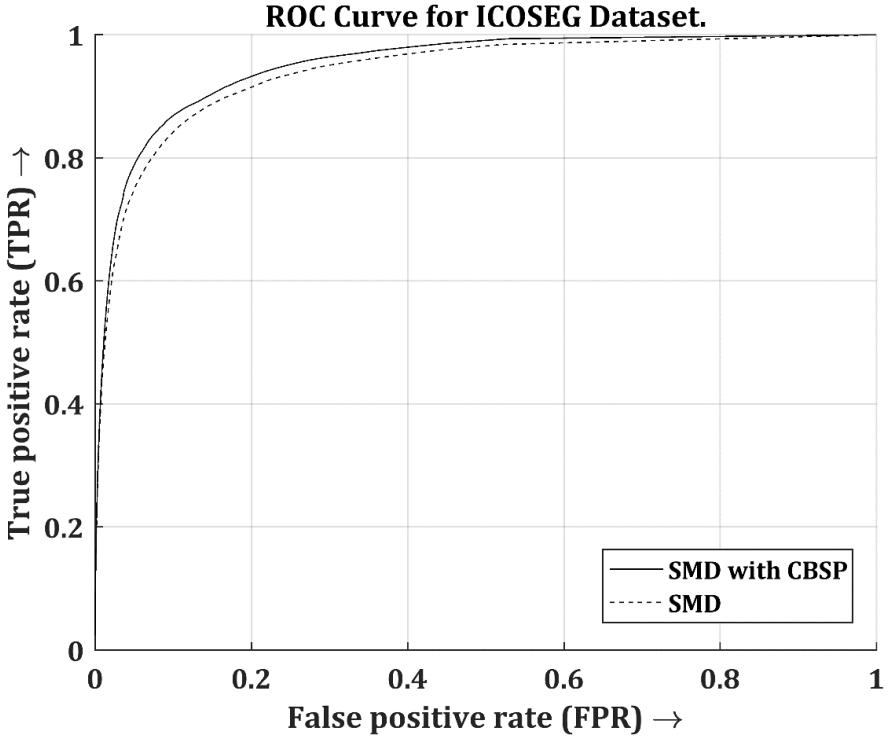

Figure 4.7: ROC curve for ICOSEG dataset.

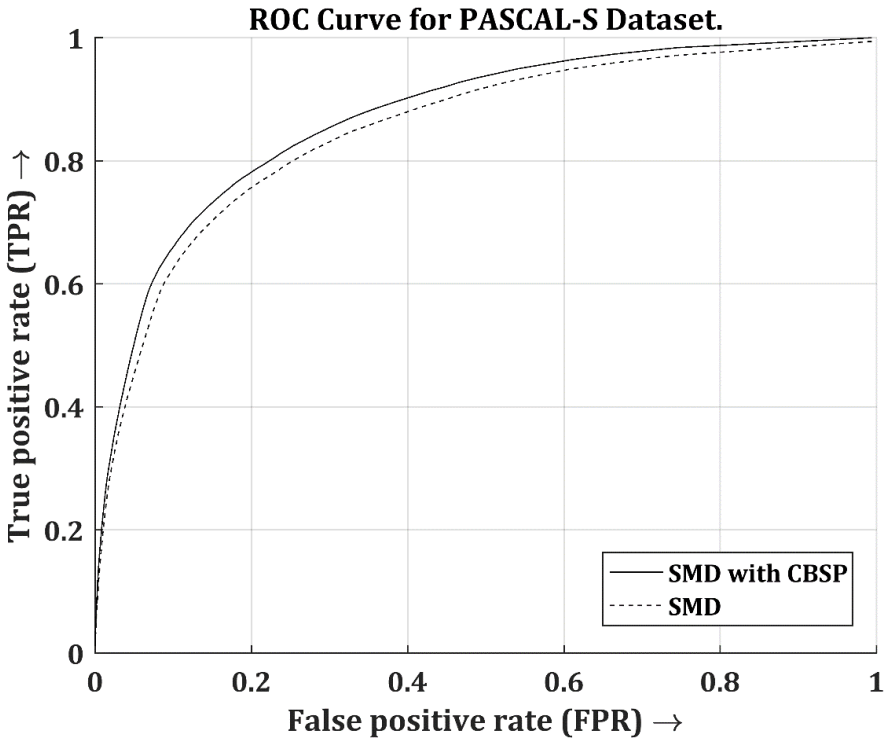

Figure 4.8: ROC curve for PASCAL-S dataset.

4.3.5 Evaluation Using PR Curve

In addition to ROC curves, precision-recall (PR) curve for three different datasets, namely, MSRA10K [14], ICOSEG [15] and PASCAL-S [16] is also plotted. For each dataset, the PR curve for both result obtained by structured matrix decomposition [11] and proposed method structured matrix decomposition with contour based spatial prior is obtained. Precision-recall (PR) curve for MSRA10K [14] dataset is shown in Figure 4.9, for ICOSEG [15] dataset is shown in Figure 4.10 and for PASCAL-S [16] dataset is shown in Figure 4.11.

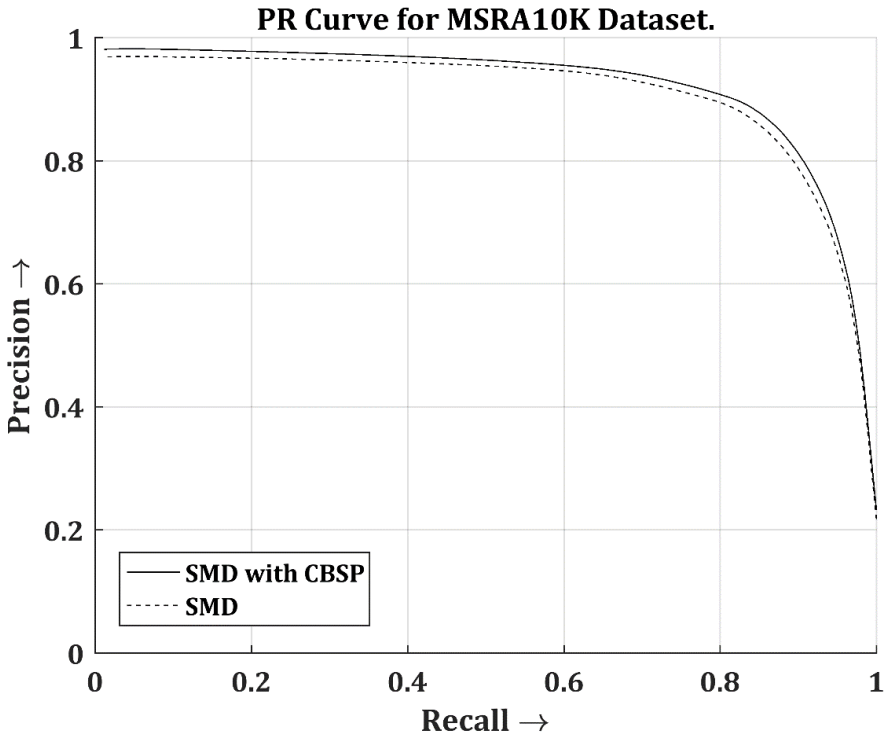

Figure 4.9: PR curve for MSRA10K dataset.

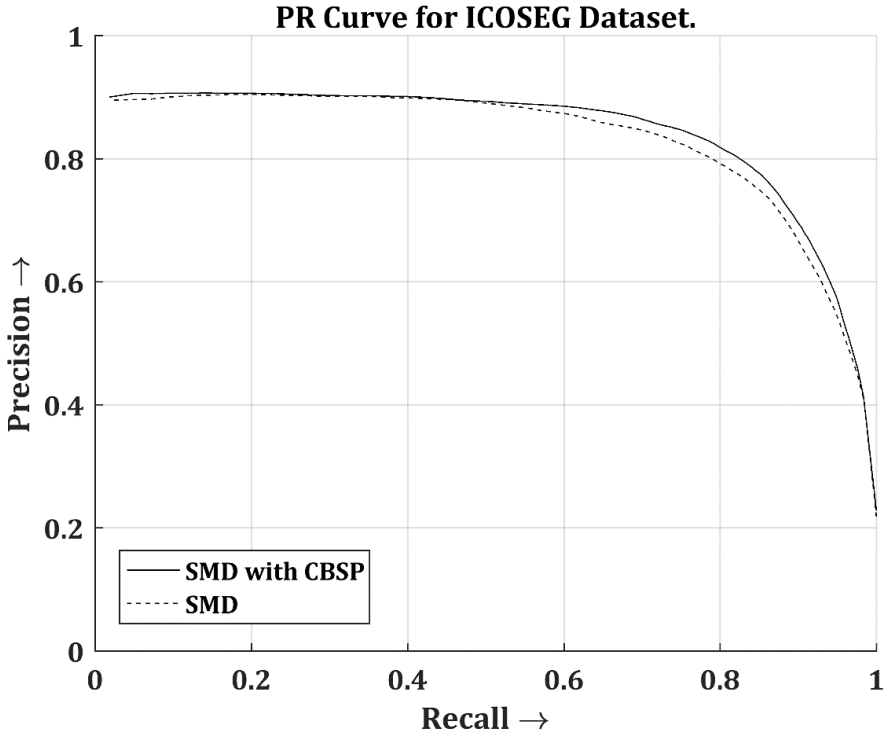

Figure 4.10: PR curve for ICOSEG dataset.

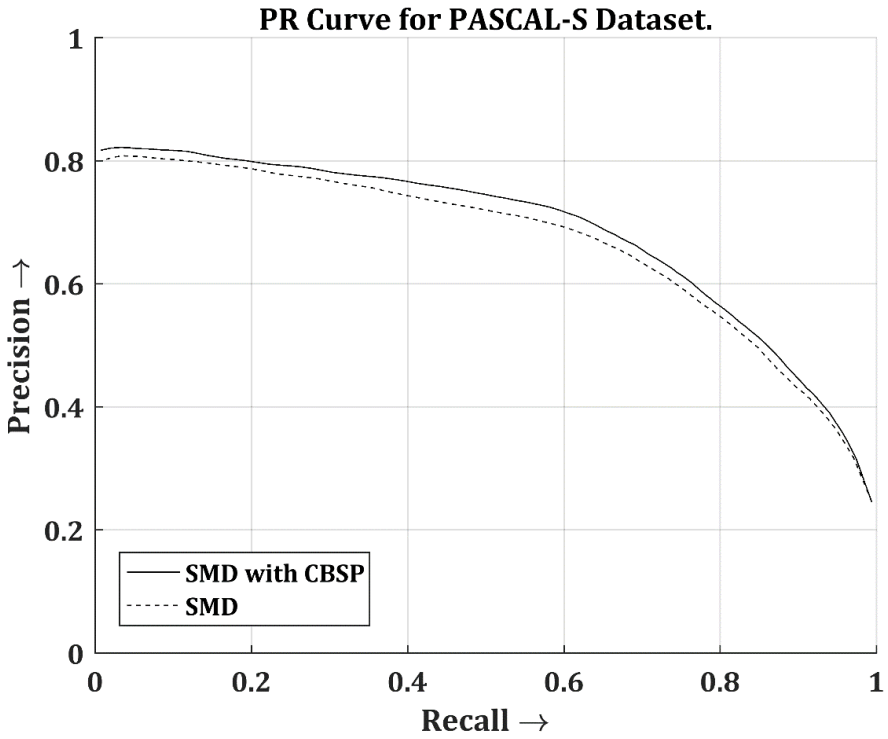

Figure 4.11: PR curve for PASCAL-S dataset.

4.3.6 Evaluation Using MAE

As complementary to ROC curves and PR curves, Mean Absolute Error (MAE) is evaluated for quantitative analysis, which considers the saliency assignment for the non-salient pixels. Which means it provides an indication of all the pixels which are erroneously marked as salient object pixels. Mean absolute error is obtained by calculating the mean difference between the saliency map and the ground truth in pixel level. Computed mean absolute error for three different datasets, namely, MSRA10K [14], ICOSEG [15] and PASCAL-S [16] are tabulated in Table 2 and plot is shown in Figure 4.12.

Table 4.2: Comparison of Mean Absolute Error

	Mean Absolute Error	
Datasets	**SMD**	**SMD with CBSP**
MSRA10K [14]	0.103982	0.098589
ICOSEG [15]	0.138161	0.118017
PASCAL-S [16]	0.208472	0.189488

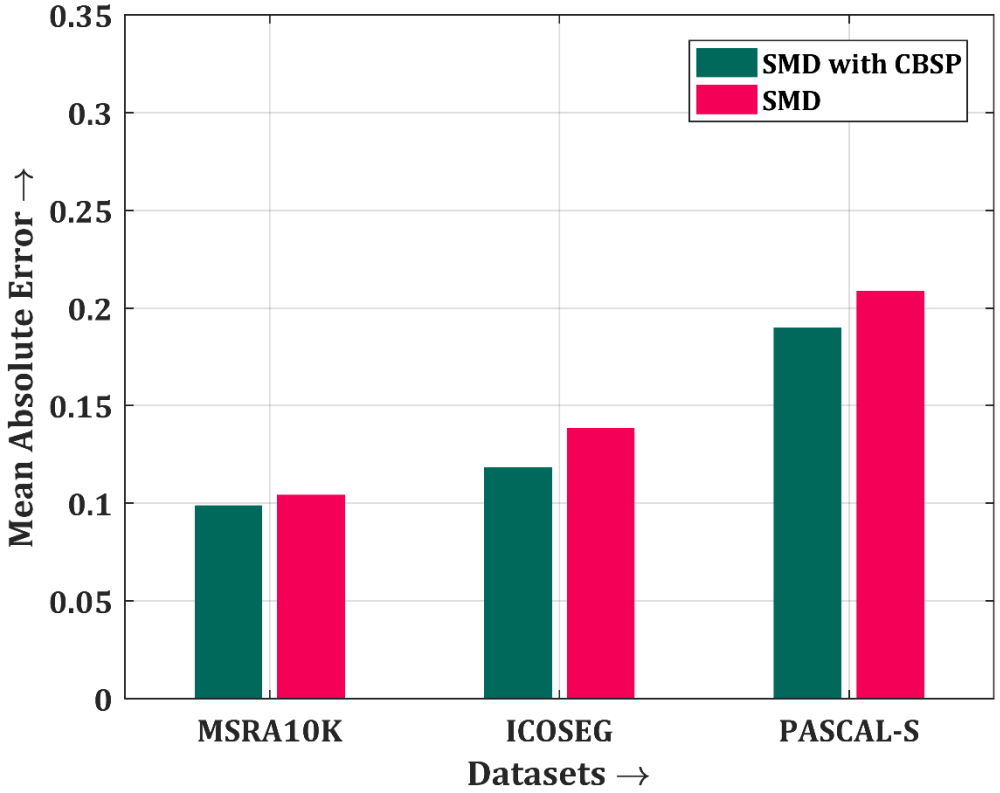

Figure 4.12: Comparison of Mean Absolute Error for Different Datasets.

4.3.7 Structural Similarity Measure

Structural similarity (SSIM) indexes measured for three different datasets using saliency map of structured matrix decomposition with contour based spatial prior and structured matrix decomposition by comparing with ground truth are tabulated in Table 4.3 and shown in Figure 4.13.

Table 4.3: Comparison of Structural Similarity (SSIM) Index

	Structure Similarity (SSIM) Index	
Datasets	**SMD**	**SMD with CBSP**
MSRA10K [14]	0.590623	0.727231
ICOSEG [15]	0.596654	0.720875
PASCAL-S [16]	0.394864	0.568001

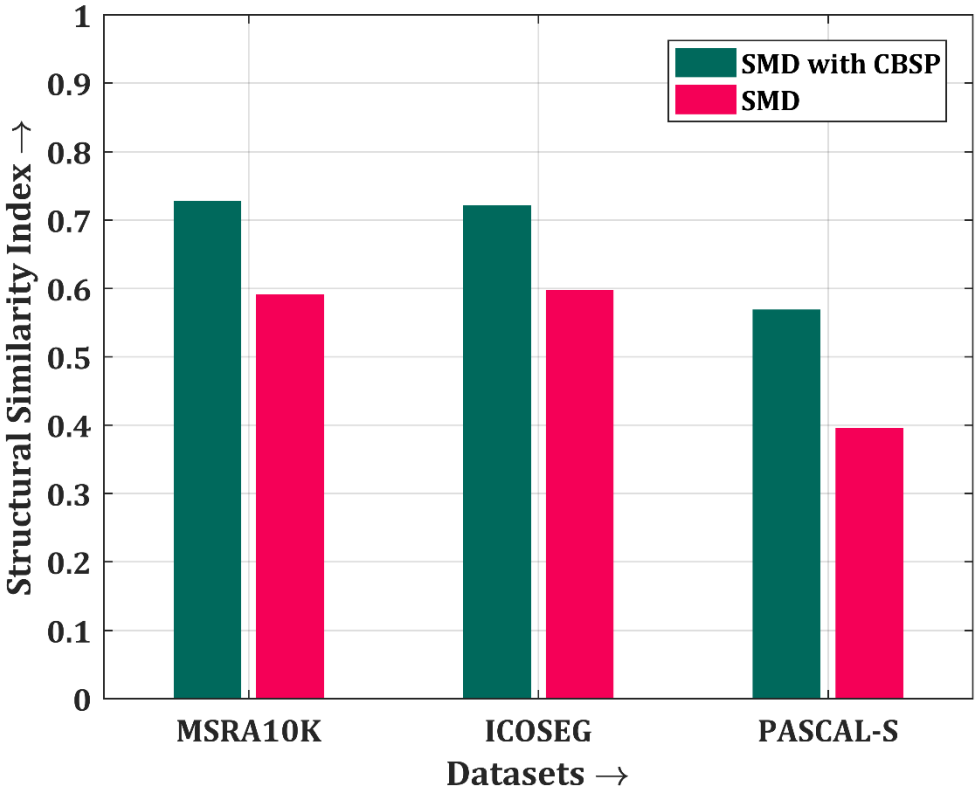

Figure 4.13: Comparison of Structural Similarity Index for Different Datasets.

4.4 Discussion

Series of experiments using different images from standard datasets involving various scenes gives promising results. Visual assessment on single object images (Figure 4.1 and Figure 4.2), multiple object images (Figure 4.3 and Figure 4.4) and complex scene images (Figure 4.5) shows that this method can accurately detects and extracts entire salient object from a given a scene and shows improvement over structured matrix decomposition method.

Comprehensive evaluation using receiver operating characteristics (ROC) curve obtained from three different datasets (Figure 4.6, Figure 4.7 and Figure 4.8) shows that proposed method has decent accuracy over saliency detection using structured matrix decomposition, since the receiver operating characteristics curve follows the left-hand border and the top border of the receiver operating characteristics space. Generally, when the curve follows closer the left-hand border and the top border of the receiver operating characteristics space specify the more accurate the result

and similarly when the curve comes closer to the 45-degree diagonal of the receiver operating characteristics space specify less accuracy.

Similarly, in terms of PR curve shown in Figure 9, Figure 10 and Figure 11 obtained from three different datasets using proposed method follows the right-hand border and the top border more than using structured matrix decomposition in precision-recall space.

As complementary to ROC and PR curves, evaluated Mean Absolute Error (MAE) for quantitative analysis shows that MAE is less by 0.005, 0.020 and 0.019 respectively for datasets MSRA10K [14], ICOSEG [15] and PASCAL-S [16] using structured matrix decomposition with contour based spatial prior than structured matrix decomposition. Since mean absolute error is evaluated in pixel level, less value obtained using structured matrix decomposition model indicates per pixel improvement in salient object detection.

Structural similarity (SSIM) indexes measured for the analysis of similarity between detected salient object and ground truth are found to be higher value by 0.137, 0.124 and 0.173 respectively for datasets MSRA10K [14], ICOSEG [15] and PASCAL-S [16] in case of proposed method than structured matrix decomposition model which indicates that structural similarity is also maintained while detecting salient object. Value of structural similarity index lies between -1 and 1. Closer to -1 indicates dissimilarity and closer to 1 indicates similarity between detected image and reference image (ground truth in this research work).

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Salient object detection and extraction is the method of localizing and segmenting most conspicuous object from an images which has a wide range of applications in image processing and computer vision. This research work aims to improve the performance of well-known method for salient object detection by structured matrix decomposition by integrating contour based spatial prior obtained from biologically inspired framework. Further this study also focuses on method to integrate contour based spatial prior to the structured matrix decomposition model.

In this thesis, performance of the proposed method is demonstrated by various experiments on three widely used datasets. The obtained result of proposed method is compared with structured matrix decomposition method on the basis of evaluation metrics. For this, visual assessment is carried for different images having single object, multiple object and complex scene on it. For comprehensive analysis, widely used evaluation metrics, receiver operating characteristics (ROC) curve, precision-recall (PR) curve and mean absolute error (MAE) are evaluated and compared with structured matrix decomposition method. Similarly, to measure structural similarity between saliency map and ground truth, structural similarity index is calculated and compared with structured matrix decomposition method. Both visual assessment as well as comprehensive evaluation indicates proposed method can detect and extract salient object accurately and shows improvement over structured matrix decomposition model.

5.2 Limitation

In this research work, low-rank regularization is used to recover image background, therefore it may be difficult to suppress some small background regions in the

image with distinctive appearances because the feature vector of those regions usually are not in low dimensional subspace. Due to this reason, sometimes, these small background regions with distinctive appearances also appear along with detected salient region and then in extracted salient object part.

Similarly, performance of the proposed method is not good in PASCALS-S dataset. The underlying reason is that PASCAL-S dataset is designed for salient object detection and eye fixation prediction and the proposed method is dealing with salient object detection. PASCAL-S dataset is mainly used to avoid the dataset bias during validation of salient object detection.

5.3 Recommendation

In this thesis, the comparative study is done only with structured matrix decomposition method. Result of this method can be further compared with other salient object detection algorithm. Similarly, other evaluation metrics like weighted F measure (WF score), area under curve (AUC) and overlapping ratio (OR) can be evaluated and compared.

Further enhancement of this work can be done by exploration and integration of more robust and general high level prior. Finding explicit way to separate the low-rank part and structured sparse part in terms of regional difference may merit further study.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur. "a model of saliency-based visual attention for rapid scene analysis". *IEEE TPAMI*, vol. 20, no. 11, pp. 12541259,1998.

[2] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu. "global contrast based salient region detection". *in CVPR*, 2011.

[3] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li. "automatic salient object segmentation based on context and shape prior". *in BMVC*, 2011, pp. 1-12.

[4] Q. Yan, L. Xu, J. Shi, and J. Jia. "hierarchical saliency detection". *in CVPR*, 2013, pp. 1155-1162.

[5] M. Cheng, J. Warrell, W. Lin, S. Zheng, V. Vineet, and N. Crook. "efficient salient region detection with soft image abstraction". *in ICCV*, 2013, pp. 1529-1536.

[6] J. Yan, M. Zhu, H. Liu, and Y. Liu. "visual saliency detection via sparsity pursuit ". *IEEE SPL*, vol. 17, no. 8, pp. 739-742, 2010.

[7] X. Shen and Y. Wu. "a unified approach to salient object detection via low rank matrix recovery". *in CVPR*, 2012, pp. 2296-2303.

[8] C. Lang, G. Liu, J. Yu, and S. Yan. "saliency detection by multitask sparsity pursuit ". *IEEE TIP*, vol. 21, no. 3, pp. 1327-1338, 2012.

[9] "W. Zou, K. Kpalma, Z. Liu, and J. Ronsin". Segmentation driven low-rank matrix recovery for saliency detection. *in BMVC*, 2013.

[10] Chang Tang, Jin Wu, Changqing, Pichao Wang, and Wanqing Li. "salient object detection via weighted low rank matrix recovery". *in LSP*, 2017.

[11] Houwen Peng, Bing Li, Haibin Ling, Weiming Hu, Weihua Xiong, and Stephen J. Maybank. "salient object detection via structured matrix decomposition". *in TPAMI*, 2016.

[12] E. Candes, X. Li, Y. Ma, and J. Wright. "robust principal component analysis?". *J. ACM*, vol. 58, no. 3, pp. 139, 2011.

[13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. "robust recovery of subspace structures by low-rank representation". *IEEE PAMI*, vol. 35, no. 1, pp. 171184, 2013.

[14] M.M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, and S.M. Hu. "global contrast based salient region detection". *IEEE TPAMI*, 2014.

[15] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. "interactively co-segmenting topically related images with intelligent scribble guidance". *IJCV*, vol. 93, no. 3,pp. 273-292 2011.

[16] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille. "the secret of salient object segmentation". *in CVPR*, 2014, pp. 280-287.

[17] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk. "slic superpixels compared to state-of-the-art superpixel methods". *IEEE TPAMI*, 2012.

[18] P. Felzenszwalb and D. Huttenlocher. "efficient graph-based image segmentation". *IJCV*, 2004.

[19] E. P. Simoncelli and W. T. Freeman. "the steerable pyramid: A flexible architecture for multi-scale derivative computation". *in ICIP*, 1995.

[20] H. G. Feichtinger and T. Strohmer. "gabor analysis and algorithms: theory and applications". *Springer*, 1998.

[21] W. Zhu, Y. Wei S. Liang, and J. Sun. "saliency optimization from robust background detection". *in CVPR*, 2014.

[22] K. Yang, S. Gao, C. Li, and Y. Li. "efficient color boundary detection with color-opponent mechanisms". *IEEE CVPR*, 2013.

[23] Kai-Fu Yang, Hui Li, Chao-Yi Li, and Yong-Jie Li. "a unified framework for salient structure detection by contour-guided visual search". *IEEE TIP*, 2016.

[24] G. Papari, P. Campisi, N. Petkov, and A. Neri. "a biologically motivated multiresolution approach to contour detection". *EURASIP Journal on Applied Signal Processing*, 2007.

[25] N. Petkov and M. A. Westenberg. "suppression of contour perception by band-limited noise and its relation to non-classical receptive field inhibition". *Biological Cybernetics*, 2003.

[26] E. N. Johnson, M. J. Hawken, and R. Shapley. "the spatial transformation of color in the primary visual cortex of the macaque monkey". *in Nature Neuroscience*, 2001.

[27] B. R. Conway. "advances in color science: from retina to behavior". *The Journal of Neuroscience*, 2010.

[28] R. Shapley and M. Hawken. "color in the cortex–single-and double-opponent cells". *in Vision Research*, 2011.

[29] T. Judd, K. Ehinger, F. Durand, and A. Torralba. "learning to predict where humans look". *IEEE ICCV*, 2009.

[30] B. W. Tatler. "the central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions". *J. Vis.*, 2007.

APPENDIX A

More Results on MSRA10K Dataset

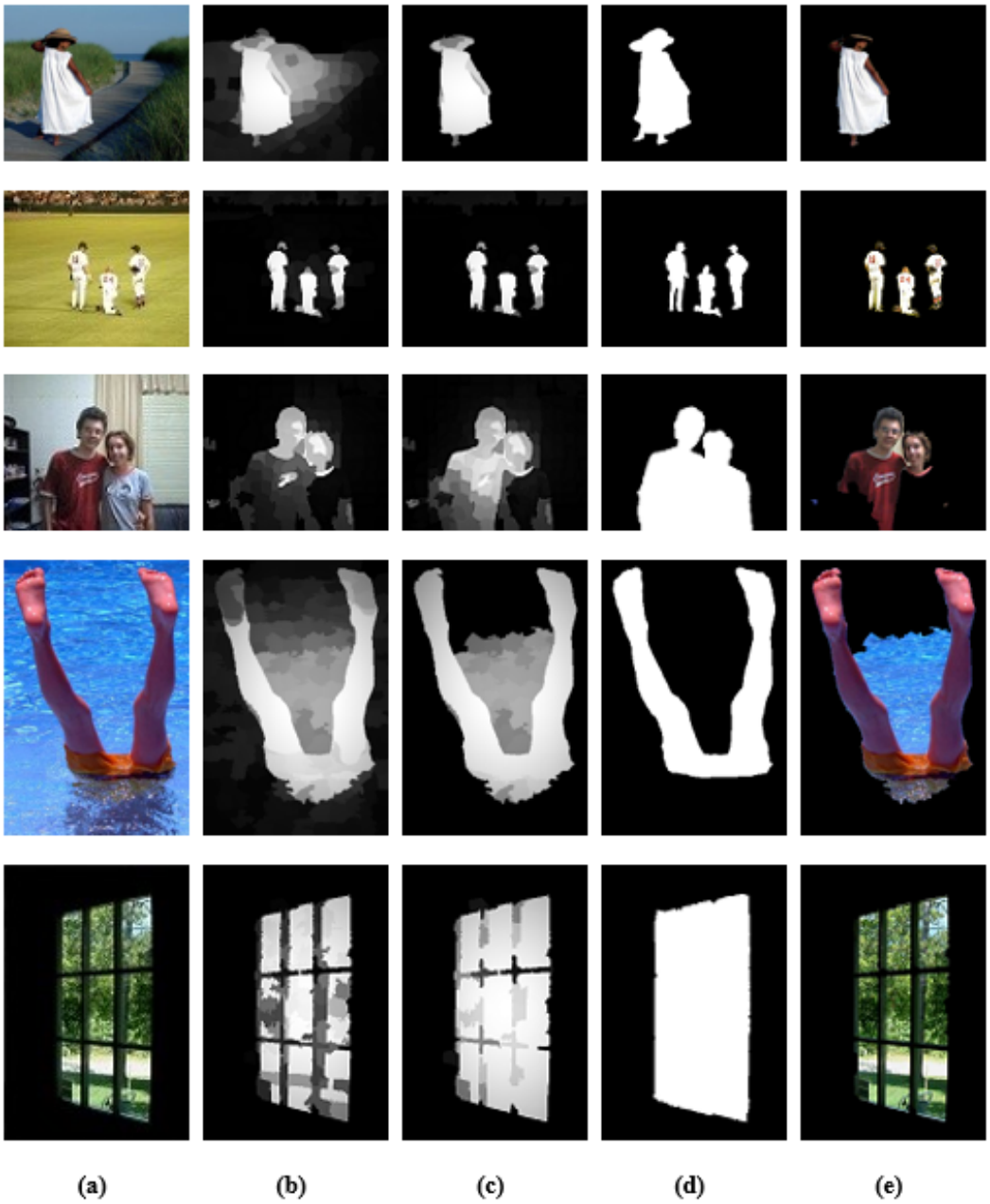

Figure A.1: More Results on images from MSRA10K dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.

APPENDIX B

More Results on ICOSEG Dataset

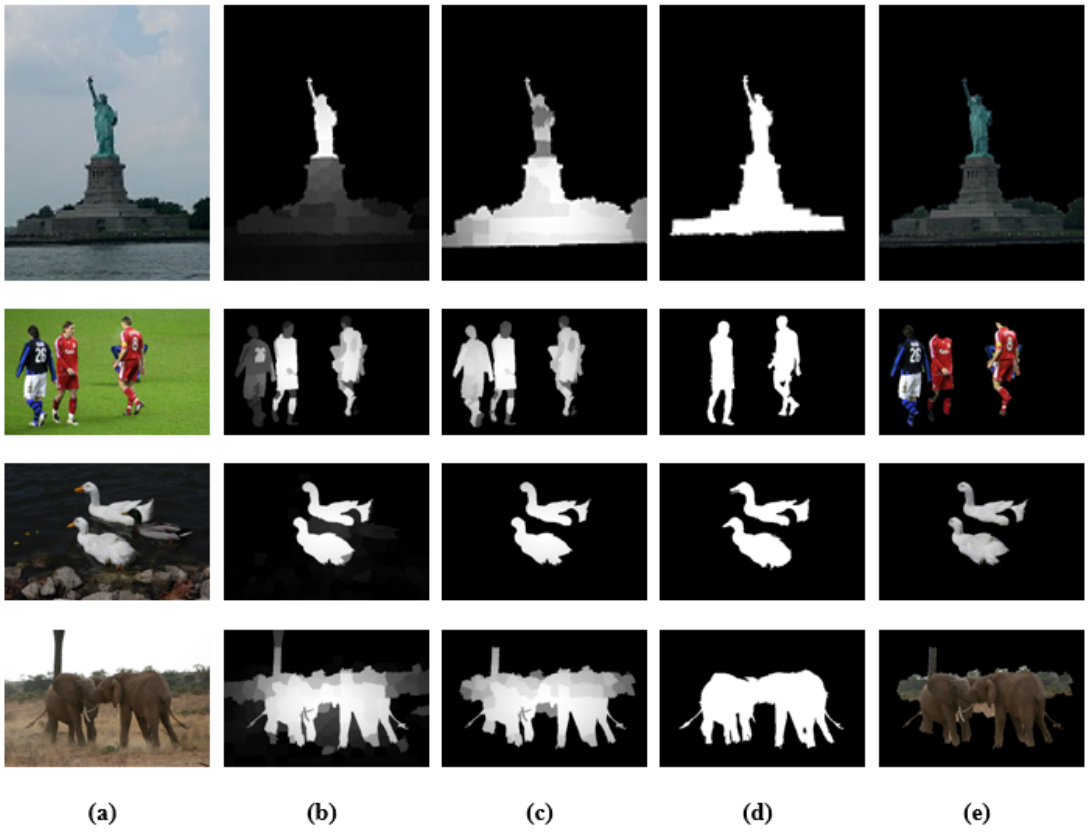

Figure B.1: More Results on images from ICOSEG dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.

APPENDIX C

More Results on PASCAL-S Dataset

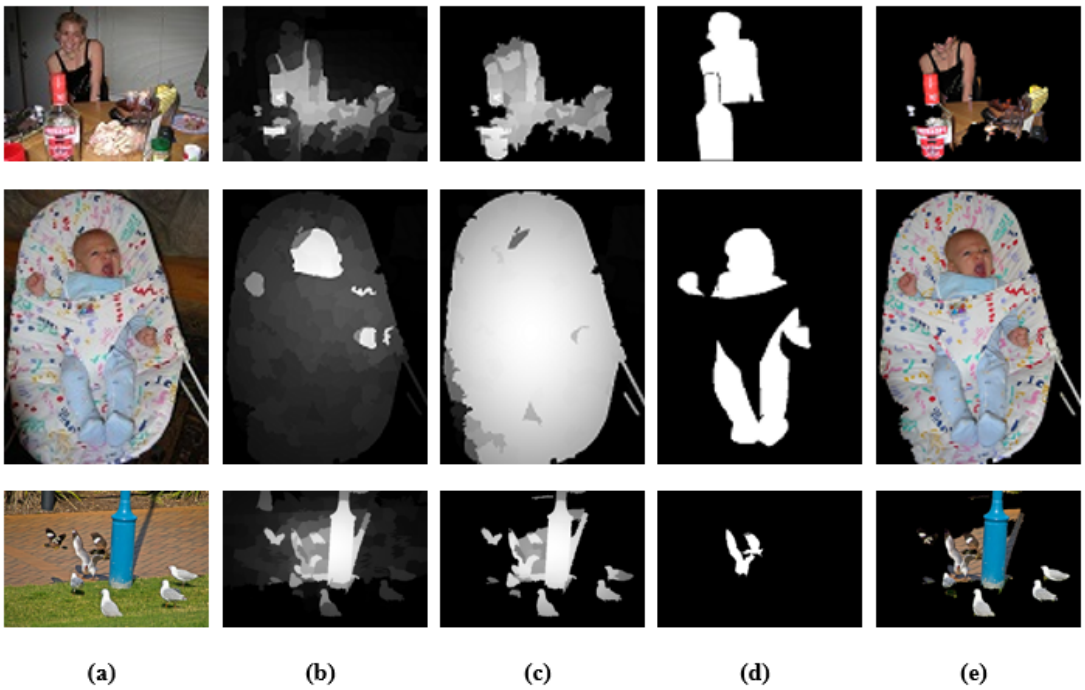

Figure C.1: More Results on images from PASCAL-S dataset. (a) Original images (b) Detected images by structured matrix decomposition. (c) Detected images by structured matrix decomposition and contour based spatial prior. (d) Ground Truth (e) Extracted images using structured matrix decomposition and contour based spatial prior.