

August 2013

Master's Degree Thesis

A Photograph Reconstruction Method Based on Parameter-Free Spectral Matting for Better Composition

Graduate School of Chosun University

Department of Computer Engineering

Sidra Riaz

비모수 스펙트럴 매칭에 기반한 보다 나은 구도를 위한 사진 재구성 방법

A Photograph Reconstruction Method Based on Parameter-Free Spectral Matting
for Better Composition

2013 년 8 월 23 일

Graduate School of Chosun University

Department of Computer Engineering

Sidra Riaz

A Photograph Reconstruction Method Based on Parameter-Free Spectral Matting for Better Composition

Advisor: Prof. Sang-Woong Lee

This Thesis is submitted to Graduate School of Chosun University in
partial fulfillment of the requirements for a Master's degree

April 2013




Graduate School of Chosun University

Department of Computer Engineering

Sidra Riaz

□

Thesis Examination Committee

위원장	조선대학교 교수	문인규	
위원	조선대학교 교수	양희덕	
위원	조선대학교 교수	이상웅	

2013 년 4 월

조선대학교 대학원

© 2013
Sidra Riaz
All rights reserved.

I dedicate this thesis to my beloved family, my loving parents (Mr. Riaz Hussain & Mrs. Farkhandah Parveen) for their endless love, support and encouragement, my grandmother (Mrs. Faiz Bibi) for her pure love and prayers, my brothers (Mr. Sheraz Hussain, Mr. Aleem Riaz, Mr. Abbass Riaz) for their reinforcement, my cute Niece (Aleeza Sheraz) for being a shining star of my life, and especially to my very nice and caring sister- Aqsa Riaz.

Table of Contents

Table of Contents	i
List of Figures	iv
List of Tables	v
List of Acronyms	vi
초록	vii

I. Introduction	1
A. Compositional Photography	1
1. Rule-of-Thirds	1
2. Balancing Elements	2
3. Leading Lines	2
4. Symmetry and Patterns	2
5. View Point	3
6. Background Simplicity	3
7. Depth of Field	3
8. Natural Framing	3
9. Cropping (Macros)	3
B. Research Motivation	3
C. Research Objectives	4
D. Thesis Contributions	5
1. ROI-RT Algorithm Proposed	5
2. Revised Spectral Matting Technique	5
3. Study on Segmentation Approaches	5
4. Aesthetic Score Evaluation	5
E. Thesis Organization	6

II. Overview and Related Work.....7

III. Photograph Reconsruction Method for Better Composition9

- A. Overview of Proposed Approach..... 9
 - 1. Background-Foreground Estimation..... 10
 - 1.1 Automatic Spectral Matting* 11
 - 1.2 Discarding Alpha Components* 13
 - 2. Main Object Segmentation..... 14
 - 3. State-of-the-art Techniques in Segmentation..... 15
 - 3.1 Bayesian Matting* 15
 - 3.2 Spectral Matting*..... 16
 - 3.3 Magic Wand*..... 16
 - 3.4 Intelligent Scissors* 17
 - 3.5 Graph Cut* 17
 - 3.6 Grab Cut* 18
 - 4. Segment-based Feature Extraction..... 20
 - 5. Background Best Patch Detection..... 23
 - 6. Background Texture Synthesis 24
 - 6.1 Nearest-Neighbor Function (NNF)*..... 25
 - 7. Retargeting Algorithm ROI-RT 27

IV. Retargeting Approach ROI-RT.....29

- A. Composition Guidelines..... 29
- B. Explanation of ROI-RT Algorithm 30

V. Aesthetic Score Measurement33

- 1. Composition Rule (RoT) Feature..... 33

2. Position Feature.....	34
3. Mean Intensity Feature.....	34
4. Average Saturation Feature.....	34
5. Wavelet Feature	35
6. Color Feature.....	35
7. Aspect Ratio Feature	36
8. Average Hue Feature	36
9. Aesthetic Score Function	36
VI. Experimental Results.....	37
A. Database	37
B. Discussions on Results.....	38
VI. Conclusions and Future Enhancements	44
Bibliography.....	45
ABSTRACT	48
ACKNOWLEDGMENT	49

List of Figures

Fig. 1.1: Example of a professional photograph for ‘Rule-of-thirds’	1
Fig. 1.2: Example of a professional photograph for ‘Balancing elements’	2
Fig. 1.3: Example of a professional photograph for ‘Leading lines’	2
Fig. 1.4: Example of a professional photograph for ‘Symmetry and Patterns’	2
Fig. 1.5: Example of a professional photograph which respect the ‘Rule-of-Third’	4
Fig. 2.1: Crop and retarget approach presented by Liu et al, in [5]	7
Fig. 2.2: In camera photo composition rules.	8
Fig. 3.1: Overview of proposed approach for better composition	9
Fig. 3.2: Alpha components of photographs by spectral matting	11
Fig. 3.3: A comparison of segmentation techniques	18
Fig. 3.4: Alpha components estimation and segmentation results	19
Fig. 3.5: Tilt angle feature to determine the position of the main object	20
Fig. 3.6: Laplacian matrix for photographs convolution	23
Fig. 3.7: Best patch detection results based on features	24
Fig. 3.8: The schematic diagram for texture synthesis by NNF	25
Fig. 3.9: Illustration of results for texture synthesis	27
Fig. 3.10: Example of rule-of-thirds lines for gridded mesh	28
Fig. 4.1: Composition guideline, rule-of-thirds	29
Fig. 4.2: Composition guideline, balancing elements	30
Fig. 4.3: Visualizing composition guideline, size ratio	30
Fig. 6.1: Example of photographs taken from online public websites	37
Fig. 6.2: Automatic photo composition by ROI-RT method	38
Fig. 6.3: Automatic photo composition optimization by ROI-RT method	40
Fig. 6.4: Aesthetic score results comparison	41
Fig. 6.5: Results of photographic composition optimization	42

List of Tables

Table 1: System accuracy for automatic photographic composition optimization.....	43
Table 2: Efficiency comparison of composition based on aesthetic scores.....	43

List of Acronyms

FMB	Floral Macros with Blurriness
FMD	Floral Macros with Difficult Background
NFP	Non-Floral Photographs
PPs	(Original) Positive Photographs
NPs	(Manipulated) Negative Photographs
ROI-RT	Retargeting Objects for Implementation of Rule-of-Third
RoT	Rule-of-Thirds
NNF	Nearest-Neighbor Function
GMM	Gaussian Mixer Model
FFT	Fast Fourier Transform

초록

비모수 스펙트럴 매칭에 기반한 보다 나은 구도를 위한 사진 재구성

시드라 리아즈
지도교수: 이상웅, 조교수, Ph.D.
컴퓨터공학과
조선대학교 대학원

디지털 사진 촬영 기법에서 사진 구도에 대한 규칙은 심미적 사진 촬영을 위해 필수적이다. 3분할법은 심미적으로 훌륭한 사진을 촬영하기 위해 사진사들이 이용하는 가장 기본적인면서도 중요한 규칙이다. 3분할법은 주요 피사체는 전체 사진의 3분할선이나 3분할선의 교차점에 배치되었을 때, 균형을 가진다는 내용을 담고 있다. 전통적인 방식의 카메라에 의해 촬영된 사진은 심미적으로 3분할법의 법칙을 따를 때 향상된다고 알려져 있다. 본 논문에서는 3분할법에 기반한 피사체의 재배치 방법을 제안한다. 이 방법은 자동적으로 3분할법을 따르도록 사진의 구도를 개선한다. 자동적인 3분할법을 따르기 위해서는 배경 - 전경 추출, 주 객체 분할, 특징 추출, 가장 적합한 배경 패치 탐지, 배경을 채우기 위한 합성 등의 과정이 필요하며, 마지막으로 3분할법을 이용한 사진을 재창조하기 위해서 주요 피사체를 합성된 배경에 재배치한다. 공인데이터베이스를 이용하여 실험한 결과 제안된 방법은 기존 기술에 비해 전반적으로 심리적 점수에서 94.30%의 향상을 보여주었다.

I. Introduction

A. Compositional Photography

Digital photography has become an exciting area of modern research in recent decades. People are more interested in taking photographs which are aesthetically pleasing. In taking the best photograph from a digital camera, people seek visual aesthetic quality as a measure in their photographs. ‘Aesthetic’ quality is ‘concerned with beauty and art’. The definition of aesthetics gives a good explanation for composition of photographs i.e. ‘made in an artistic way and beautiful to look at’ according to Advanced Learners Dictionary [1].

The composition plays an important role in making aesthetically pleasing photographs. Composition is defined as the ‘way in which the whole is made up’, or the ‘action of putting things together’ [2]. The poor composition of a photograph makes the main object out of focus or less prominent in the photograph. For improving the photograph’s visual aesthetics, the various established composition guidelines exist, such as 1) rule-of-thirds, 2) visual balance, 3) leading lines, 4) symmetry and patterns, 5) view point, 6) background simplicity, 7) depth-of-field, 8) framing, and 10) cropping (macro photography). The composition rules are briefly highlighted below:

1. Rule-of-Thirds

The rule-of-thirds says that you should position the most important elements in your scene along thirds lines, or at the points where the lines intersect. The photographic example for the Rule-of-third’s composition guideline is given in Figure 1.1.

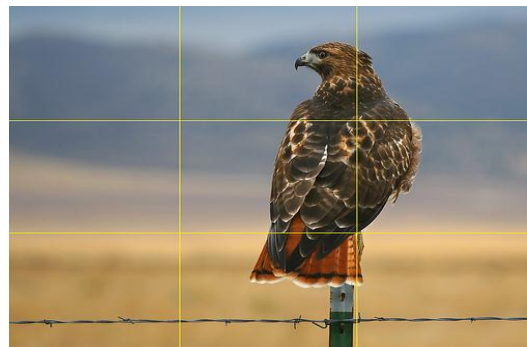


Fig. 1.1: Example of professional photograph which respect composition guidelines, ‘Rule-of-Thirds’.

2. Balancing Elements

To make a visual balance in the photograph, another object of lesser importance is included with the main object in the photograph to fill the space. The example for the 'Balancing Element' is given in Figure 1.2.



Fig. 1.2: Example of professional photograph which respects the composition guideline, i.e. 'Balancing Elements' in photography.

3. Leading Lines

The eye's attention is naturally drawn to the leading lines in the photograph. A photograph can be more pleasant to the eyes if these lines are placed carefully. Example is shown in Fig. 1.3.



Fig. 1.3: A professional photograph which respect the composition rule, i.e. 'Leading Lines'.

4. Symmetry and Patterns

In our surroundings, there are various natural and man-made patterns exist that are very eye catching. Including these patterns in the photographs improves the composition.



Fig. 1.4: A professional photograph which respect the composition rule, i.e. 'Symmetry and Patterns'.

5. View Point

The view point guideline approves that the main subject should be photographed by different angles, i.e. from above, from sides, from very close up, and from far away. The view point also improves the composition of the photograph.

6. Background Simplicity

The busy background draws out the focus from the main subject in the photograph. So, a simple background sometimes improves the composition and aesthetics of the photograph.

7. Depth-of-Field

The depth-of-field composition gives the sensation of presence in the very scene. The photographs are captured in a way to include the foreground, the middle ground, and the background. Capturing the photograph in this way improves the composition.

8. Natural Framing

To isolate the main object from outer world, natural frames, i.e. trees, and mountains are placed on the horizontal extremes of the photograph. The eye focus draws to the main object providing more visual appreciation.

9. Cropping (Macros)

Sometimes a smaller object in the huge photograph makes it very difficult to draw the attention of the viewer's. So the cropping of the photograph sometimes improves the composition.

B. Research Motivation

It is factual that composition plays an important role for capturing aesthetically pleasing shots. The Rule-of-Thirds (RoT) is the most important composition rule/guideline in photography. If digital photographs are captured in a way to follow RoT, visual appreciation is likely to be more towards these photographs. Rule-of-thirds divides a whole image into 3 x 3 grids by means of 2 vertical and 2 horizontal lines for appropriate placement of the main object as shown in Figure. 1.1 and Figure 1.4. The vertical and horizontal thirds-lines play an

important role for drawing viewer's attention towards the important object in the photograph. The placement of main objects along the thirds lines or one of the focus points (intersections) is called Rule-of-Thirds (RoT). So, our research motivation is to automate the composition of the photographs by considering RoT. The input shots to our system are automatically transformed into very highly aesthetically pleasing photographs by improving their composition.

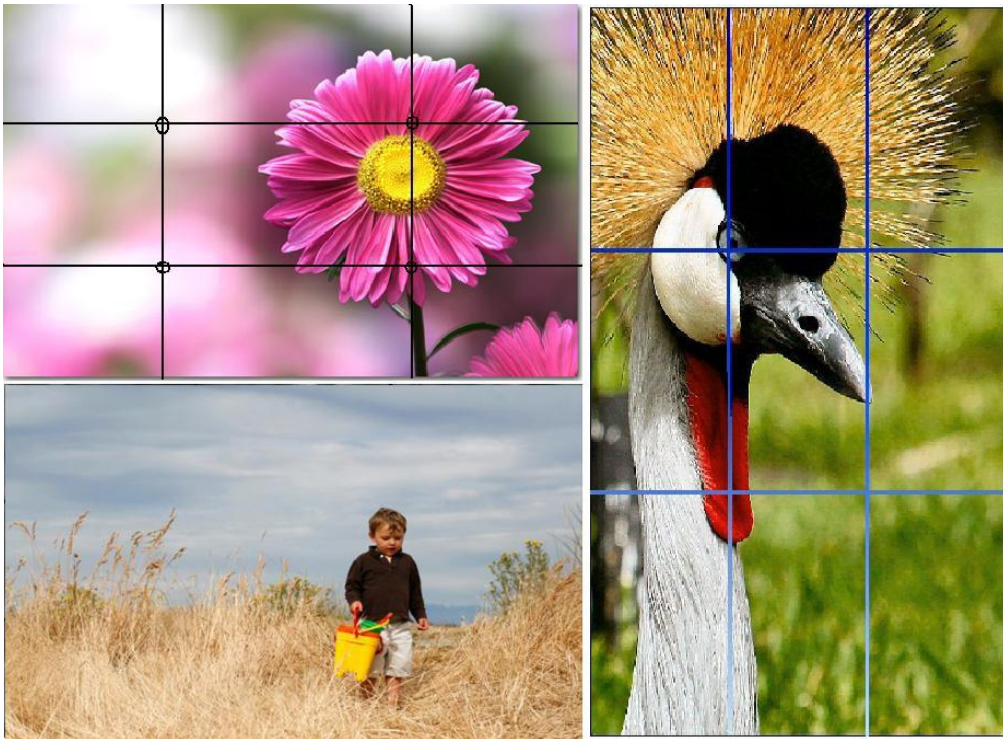


Fig. 1.5: Example of professional photographs which respect the rule-of-thirds. Main objects of interest are placed along the thirds-lines or intersection of thirds-lines in the photographs.

C. Research Objectives

In this paper, we present our computational approaches to automatically implement the rule-of-thirds (RoT) in digital photography without cropping or losing background information. The aesthetic score of resultant photographs is computationally estimated which is comparative to the human user ratings. The proposed system is a useful approach for camera automation, an onsite guide to non-photographers, and has various interesting applications in

computer vision. This approach also provides us an insight study of aesthetic score improvement in digital photography.

D. Thesis Contributions

1. ROI-RT Algorithm Proposed

Photo composition rules play an important role in creating highly aesthetic photographs. In this paper, we explain our computational approach for automating an important photo composition rule, rule-of-thirds (RoT) in digital photography. We have proposed an algorithm namely ‘Retargeting Objects for Implementation of Rule-of-Thirds’, (ROI-RT).

2. Modification in Spectral Matting Technique for High Accuracy

We have performed essential modifications in original spectral matting technique [3] for achieving results with more accuracy. Before grouping alpha mattes and selecting the best alpha, we propose some steps for selecting best alpha components and grouping them. The results are compared to original spectral matting technique which shows high accuracy rate by modified method.

3. Study on Segmentation Approaches

For automatic photo composition rules, we segment the main object from original photographs and then retarget on texturized background by using ROI-RT algorithm. A thorough study is made on many segmentation approaches to evaluate system accuracy and performance.

4. Aesthetic Score Evaluation

Aesthetic score of resultant photographs is also computed for validating the results in terms of visually aesthetically pleasing photographs. The comparison is made with existing methods for automation of composition rules.

E. Thesis Organization

The rest of the thesis is organized as follows. In Chapter 2, a comprehensive overview of the system and essential description of the related work is provided. Chapter 3 discusses our proposed approach and provides a full synopsis of our system, and Chapter 4 is related to ROI-RT scheme for composition improvement. In Chapter 5, we have proposed an image evaluation scheme such as aesthetic score based on generic features and aesthetic score function (ASF). Chapter 6 includes the experimental results for validating the proposed methods. In Chapter 7, conclusion of the thesis is presented by setting an insight on our future plans.

II. Overview and Related Work

Photographic composition is a well-researched topic for quality assessment in computational photography. This topic has received a large attention of the researchers. Recent works include rule of thirds detection from photographs based on features which are designed by saliency and generic objectness map for detecting visual elements within the photographs [4]. Further work on optimizing the photo composition in [5] by crop and retarget approach for better aesthetic quality is presented by Liu et al. In this crop and retarget method, photo composition is improved by selecting the region of the photograph which has high aesthetic score, but the limitation of the system is that a large area in the image is removed, i.e. cropped as shown in Figure 2.1.

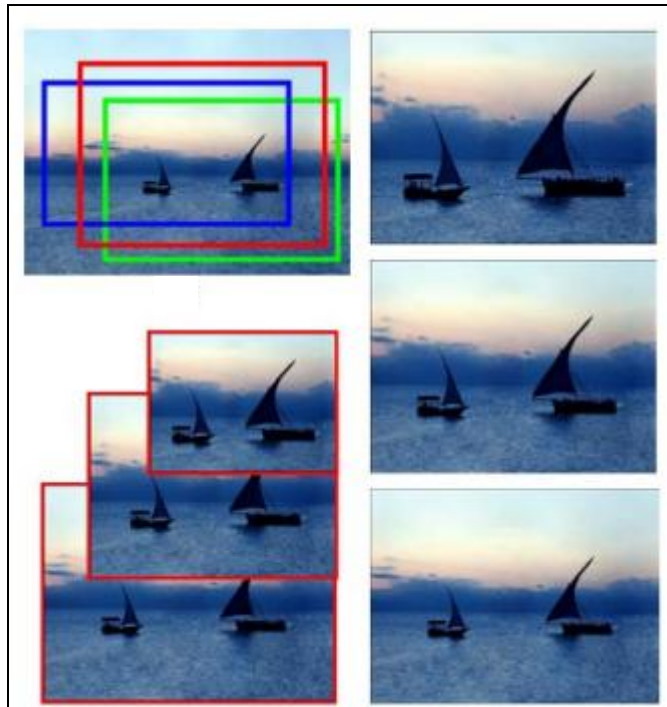


Fig. 2.1: Crop and retarget approach presented based on aesthetic score by Liu et al, in [5].

Similar methods are image retargeting based on global energy map [6] and automatic image retargeting based on these steps, i.e. segmentation, important object recognition, removing and

filling the gaps, resizing backgrounds and re-inserting important regions on the image [7]. This retargeting technique is useful to display large images on smaller displays i.e. mobile phones. Image retargeting with scene consistency is presented in [8] based on depth information is given. To make a professional photograph while taking the shot of the panoramic scene, to find the good composition and to find good views in panorama by saliency and GIST descriptors is presented in [9]. The similar work for photographic composition rules is presented in [10] which is a useful approach for in camera automation. The user takes the shot of the scene and main object is segmented and repositioned along the thirds-lines by signaling i.e. to move camera in some particular direction or to shift or crop the picture so that the center of mass should fall onto nearest corners. The example picture is shown in Fig. 2.2. The limitation of the scheme is that the background is cropped in order to make the center of the mass of the main object on the thirds lines.

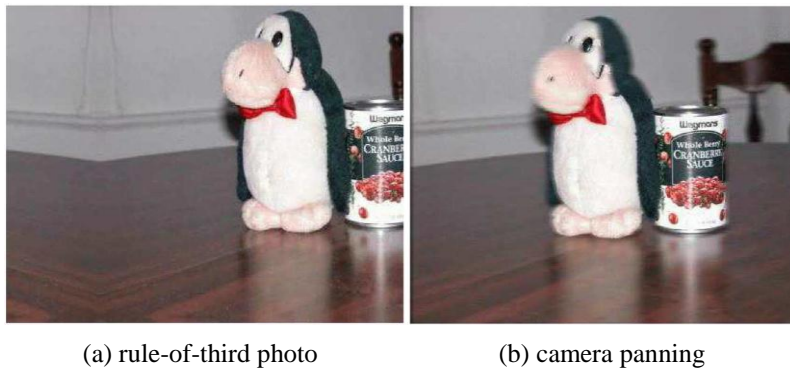


Fig. 2.2: In camera photo composition rules by detecting main subject and placing on thirds-lines by moving camera, or by shifting or cropping the picture.

These existing methods of image retargeting and seam carving are designed and presented for specific requirements and applications i.e. energy-map optimization, on mobile display, resizing for different aspect ratios etc. including systematic limitations for retargeting objects. In existing work, we also do not find the aesthetic measure of the resultant photographs which is the fundamental requirement and important aspect in digital photography.

III. Photograph Reconstruction Method for Better Composition

A. Overview of Proposed Approach

In this section, we present a detailed overview of the proposed system for automation of composition rules in digital photographs by our approach ‘Retargeting Object for Implementation of Rule-of-Thirds (ROI-RT)’. The main steps of our methods are shown in block diagram in Figure. 3.1. The first step is to determine the background and the foreground of the photographs.

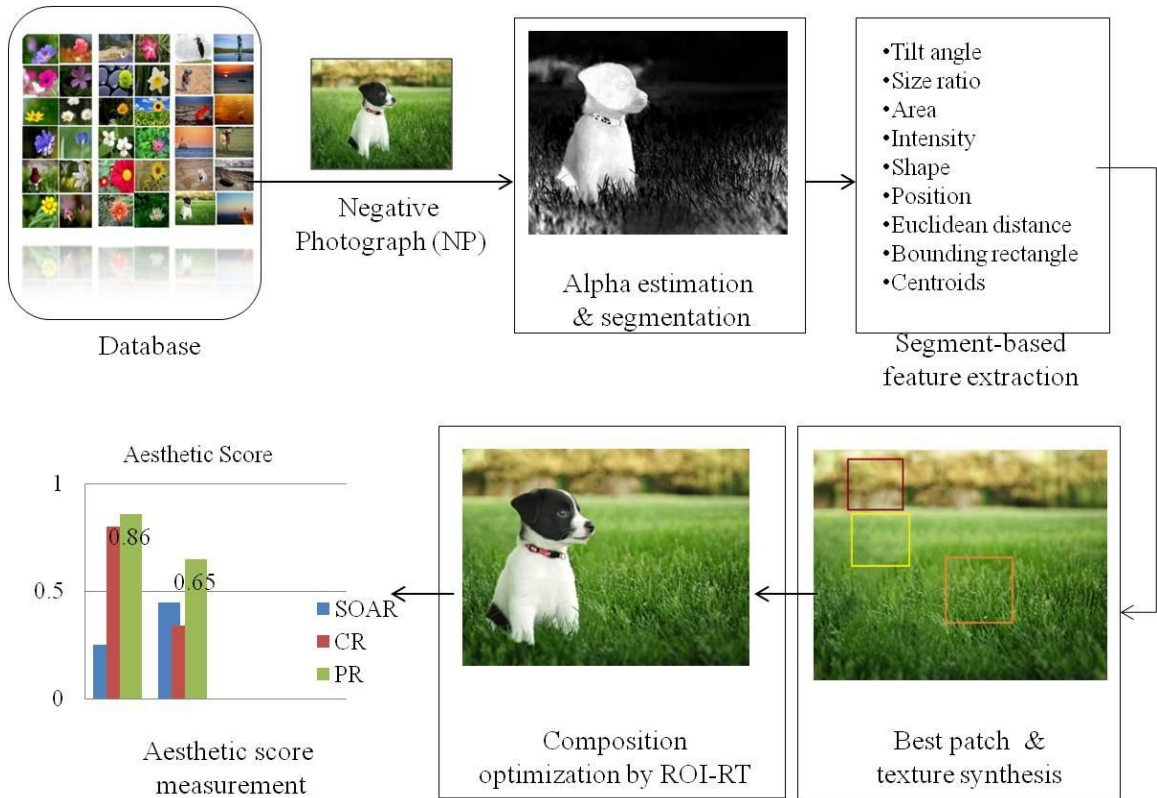


Fig. 3.1: Overview of the proposed approach for better composition.

Second step is the segmentation of main object from photographs. Third step involves the segment-based features extraction from the main object. These features are used in retargeting the main object by using ROI-RT approaches. In fourth step, best region from the background is detected based on features extracted from the background portion of the photograph. Next step involves seamless texture synthesis for generating background of composite or resultant photograph. Final step is to properly retarget the main object on textured background to achieve a photograph which follows the rule-of-thirds (RoT). The output photograph respects the rule-of-thirds, RoT to provide a better visual appreciation to the viewers. Each step is thoroughly explained in the following sections.

1. Background-Foreground Estimation

At first step, background-foreground estimation is done by automatic spectral matting techniques [3]. Foreground information of input photographs is given by opacity estimation parameter, α . The reason of using matting techniques at first step is that 1) we have a better estimate of background and foreground pixels before performing hard segmentation, 2) Opacity estimation, α is also used for combining the segmented foreground with a new textured background by using ROI-RT algorithm. The fine pixels information on edges of the main object might lose in segmentation, so opacity estimation of foreground gives us a better tool for fine segmentation from background also for repositioning the foreground on new background. We have performed essential modifications in original spectral matting technique for achieving high accuracy. In Figure 3.2, results of alpha mattes are given with spectral matting technique and by our modified method. Here we explain the background foreground estimation method using automatic spectral matting techniques.

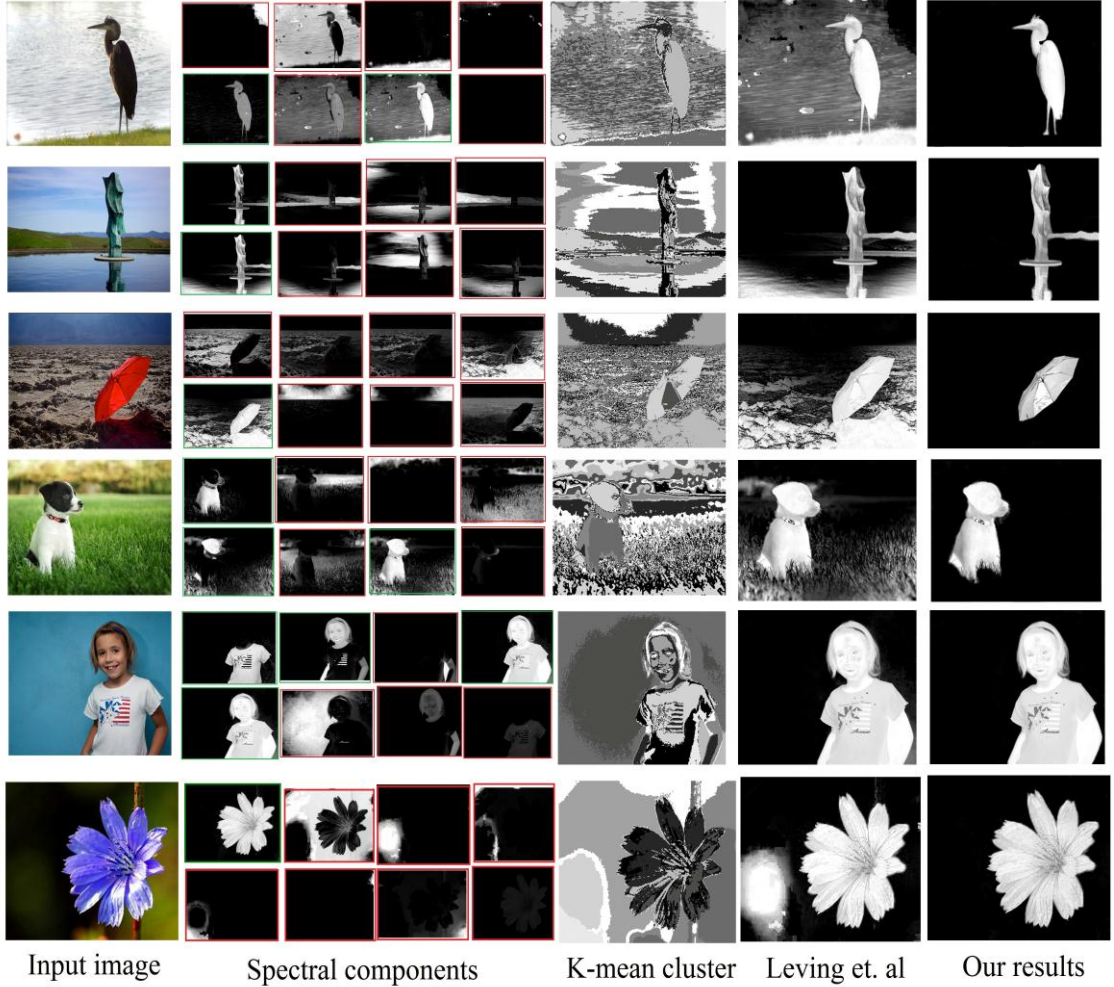


Fig. 3.2: Alpha components extracted by revised spectral matting technique also compared to k-means clustering and Levin et. al [3].

1.1 Automatic Spectral Matting

Image matting typically assume that each pixel $\{x_i \in X, y_i \in Y\}$ in an input image $I(x, y)$ is a linear combination of a foreground color F_i , a background color B_i , and alpha matte, α_i . Alpha matte, α_i is the pixels foreground opacity parameter. In Eq. (1), we can see the composition equation of an image.

$$I(x, y)_i = \alpha_i F_i + (1 - \alpha) B_i \quad (1)$$

The generalized composition equation in spectral matting is achieved by assuming that each pixel of the image is a convex combination of K image layers $L^1 : L^K$ as given by Eq. (2).

$$\sum_{k=1}^K \alpha_i^k F_i^k = 1; \alpha_i^k \in [0,1] \quad (2)$$

Alpha matte, α_j of pixel 'j' is written as given in Eq. (3) below:

$$\alpha_j = \alpha_i^R I_j^R + \alpha_i^G I_j^G + \alpha_i^B I_j^B + b_i \quad (3)$$

where, I_j^R, I_j^G, I_j^B are the *RGB* values of pixel 'j'. Estimation of alpha matte, α is done by minimizing the cost function as given by Eq. (4).

$$J(\alpha, a, b) = \sum_{m \in I \in w_m} (\alpha_i - \alpha_m^R I_i^R - \alpha_m^G I_i^G - \alpha_m^B I_i^B - b_m)^2 + \varepsilon \|a_m\|^2 \quad (4)$$

where, $\varepsilon \|a_m\|^2$ is a regulation term. Matting components, α_i^k in Eq. (1) must satisfy the condition given in Eq. (2). The K vectors of α_i^k specify the fractional contribution to the final color by all layers at each pixel position. The matting components are non-negative and sum to 1 at every pixel as given in Eq. (2). Suppose that the input photograph contains K components $C^1 : C^K$ such that they are unique $C_i \cap C_j = \emptyset$ for $i \neq j$. The Laplacian matrix, $L = \sum_n A_n$

is an $N \times N$ matrix given by Eq. (5).

$$L = \sum_{k((i,j) \in w_k)} \left(\delta_{ij} - \frac{1}{|w_k|} \left(1 + \frac{1}{\frac{\varepsilon}{|w_k|} + \sigma_k^2} (I_i - \mu_k)(I_j - \mu_k) \right) \right) \quad (5)$$

where, δ_{ij} is the Kronecker delta, μ_k and σ_k^2 are the mean and variance of the intensities in the window w_k around k , and $|w_k|$ is the number of pixels in this window. M eigenvectors of the Laplacian matrix in Eq. (5) are computed as $E = [e^1, e^2, \dots, e^M]$ to form an $N \times M$ matrix. Then spectral segmentation with k -means clustering is performed to obtain the distinct components, α_k of an image based on eigenvector matrix, E . The distinct components of the image are given by Eq. (6).

$$\alpha^k = EE^T m^{c^k} \quad (6)$$

Before grouping different components of alpha matte and selecting the best alpha estimate in spectral matting technique, we discard the components with less opacity value. For a reminder, a value “0” represents fully transparent color and a value “1” represents fully opaque color. Since, we know in natural images, the majority pixels are opaque, so alpha, α values are in the range of [0, 1].

1.2 Discarding Alpha Components

1) For each of K components of alpha matte, (α_i^k) compute a $K \times 1$ matrix

$$V[X]_{|k=1,2,\dots,K} = \sum_{i=1}^N \alpha_i^k.$$

2) Compare $(x+1)st$ element in the matrix $V[X]$ with xth element sequentially as if $(x+1) < m$ and replace $(x+1)st$ element appropriately. For $x+1 = X$, stop; otherwise, replace m by $m+1$. Every element in $V[X]$ corresponds to an alpha matte, α_i . The first element in $V[X]$ represents α_i with a higher value in the range 0 : 1.

3) For every element in $V[X]$, check: (i) number of connected and closed opacity components,

$$C_{comp}^k \text{ for } k = 1, 2, \dots, m, \text{ and (ii) size of each connected component } S_k = size(C_{comp}^k).$$

4) For $C_{comp}^k = 0$ discard it.

5) For $C_{comp}^k = 1$, if $S_k = size(C_{comp}^k) > \tau$ where τ is a threshold value, select the alpha component *else* $S_k = size(C_{comp}^k) < \tau$ discard it.

6) For $C_{comp}^k = 0$ and $S_k = size(C_{comp}^k) < \tau$ discard the alpha matte component.

Remaining alpha components are, $\alpha_i^r = \alpha_i^k - \alpha_i^d$ where α_i^d are discarded components. By combining remaining components, α_i^r through Eq. (7) we find the final opacity estimation parameter which can distinguish the foreground and background.

$$\alpha_T = \sum_i w^i \alpha_i^r \quad (7)$$

where, w is a weight matrix with values $[0,1]$ for alpha components representing background and foreground respectively.

2. Main Object Extraction

In second step, segmentation techniques are applied with the help of alpha matte for object of interest extraction from negative photographs (NPs). First we review state-of-the-art methods for segmenting the ROI with interactive and non-interactive segmentation and matting techniques. Here are mentioned these methods for comparison with our technique, i.e. 1) Bayesian Matting [11], 2) Spectral matting [12], 3) Magic Wand [13], 4) Intelligent Scissors [14], 5) Graph Cut [15], and 7) Grab Cut [16] etc. In our methods of segmenting the ROI, no user effort is required for initial tri-map provision. Also, there is no need for marking the margins of ROI to accurately estimate the boundaries. We take advantage of automatic spectral alpha matte extraction in Levin et al., [3] and improve its accuracy by discarding non-fudicial alpha components. Then we mask the object with seamless boundaries using alpha matte.

3. State-of-the-art Techniques in Segmentation

3.1 Bayesian Matting

In Bayesian framework, the assumption is made to have information of Background (B), Foreground (F) and unknown region (C). Given the observation C, the problem is to find the most likely estimates for F, B, and α . It is expressed as a maximization over a sum of log likelihoods:

$$\begin{aligned}
 \arg \max_{F,B,\alpha} P(F, B, \alpha | C) \\
 &= \arg \max_{F,B,\alpha} P(F)P(B)P(\alpha)/P(C) \\
 &= \arg \max_{F,B,\alpha} L(C | F, B, \alpha) + \dots \\
 &\quad L(F) + L(B) + L(\alpha)
 \end{aligned} \tag{8}$$

where $L(\cdot)$ is the log likelihood $L(\cdot) = \log P(\cdot)$. The first term $L(C | F, B, \alpha)$ in Eq. (8) is defined as:

$$L(C | F, B, \alpha) = \frac{-\|C - \alpha F - (1 - \alpha)B\|^2}{\sigma_C^2} \tag{9}$$

The log likelihood corresponds to a Gaussian probability distribution centered at $\bar{C} = \alpha F + (1 - \alpha)B$ with standard deviation σ_C . To estimate the foreground term, $L(F)$, spatial coherence of the image is used as given below:

$$L(F) = -(F - \bar{F})^T \sum_F^{-1} (F - \bar{F})/2 \tag{10}$$

The log likelihood of $F(B)$ depends on the matting type, i.e. for natural image matting, the term used is analogous to that of $L(F)$ by setting w_i to $(1 - \alpha_i)^2$. The log likelihood of $L(\alpha)$

is considered as constant. To solve Eq. (8), two assumptions are made: (i) α is constant, (ii) F and B are constant. With the assumption of alpha, α is constant, partial derivatives with respect to F and B of Eq. (8) carried out and set them equal to zero to results in Eq. (11).

$$\begin{bmatrix} \sum_F^{-1} + I\alpha^2/\sigma_C^2 & I\alpha(1-\alpha)/\sigma_C^2 \\ I\alpha(1-\alpha)/\sigma_C^2 & \sum_B^{-1} + I(1-\alpha)^2/\sigma_C^2 \end{bmatrix} \begin{bmatrix} F \\ B \end{bmatrix} = \begin{bmatrix} \sum_F^{-1} \bar{F} + C\alpha/\sigma_C^2 \\ \sum_B^{-1} \bar{B} + C(1-\alpha)/\sigma_C^2 \end{bmatrix} \quad (11)$$

So, for a constant α , F and B are solved through linear equations given above. For solving alpha, α , F and B are assumed constant which yield a quadratic equation in α as:

$$\alpha = \frac{(C - B).(F - B)}{\|F - B\|^2} \quad (12)$$

3.2 Spectral Matting

In spectral matting technique, initial tri-map or scribble information is not required [3]. Matting components are extracted automatically and then summed up as given in Eq. (7). In details, we have discussed automatic spectral matting technique in section 1.1.

3.3 Magic Wand

Magic wand is widely used for segmentation and included in many image editing tools. The segmentation results are good for object with high contrast compared to the background of uniform color. The operation of Magic Wand is explained below:

Magic wand takes only one input point $Input = i$. Based on color similarity with the input seed it selects the areas on image given a threshold value, τ .

For: N pixels exist that directly linked with i :

if $W(x, y) > \tau$

1) compute a matrix $L[N] = P(x, y)_n$ for $N > 1$

2) classify pixel $n \in N$ belong to segment s

3) Go to step (2) to classify all the pixels in N.

3.4 *Intelligent Scissors*

In intelligent scissors technique, two initial seeds are required. Based on the input seeds, a minimum weighted path between the two is achieved. The minimum weighted path might not be optimal and also the algorithm is generalized to two dimensional images. The bi-directional link between two input seeds is given by:

$$L(p, q) = \begin{bmatrix} (q - p) & \text{if } Dp \times (q - p) \geq 0 \\ (p - q) & \text{if } Dp \times (q - p) < 0 \end{bmatrix} \quad (13)$$

where p and q are two seed points. The bidirectional links $L(p, q)$ are horizontal, vertical or diagonal with the position of p and q .

3.5 *Graph Cut*

Segmentation is defined as min-cut and max-flow graph cuts problem. The maximum flow is to find the flow of the maximum value on a flow network G. A cut of a flow network G is a partition of V into S and $T = V - S$ such as $s \in S$ and $t \in T$. For a given flow f , the net flow is defined as $f(S, T) = \sum_{x \in S} \sum_{y \in T} f(x, y)$. Also the capacity of the cut is defined as

$c(S, T) = \sum_{x \in S} \sum_{y \in T} c(x, y)$. A min-cut of a flow is a cut whose capacity or cost is the least of all

the s-t cuts of the network [15]. The cost of the cut is defined as $c = \sum_{i, j \in E_i, j \in C_{i, j}} W(i, j)$, where

$E_{i, j}$ represent edges that when removed divide the graph into two discrete segments. Weights, $W(i, j)$ represents the flow. The graph cut can lead us to an accurate segmentation results on the cost of slow implementation and user input for scribbles.

3.6 Grab Cut

Grab Cut [16] is very famous interactive segmentation technique. It is an improvement over Graph Cut technique with GMM (Gaussian Mixture Model) and iterative optimization. The user interaction is also reduced compared to existing methods of magic wand, intelligent scissors etc. Its algorithm is explained below:

- 1) For each pixel $p(x, y)$ assign a value $\alpha \in [0, 1]$ i.e. $\alpha = 0$ for background and $\alpha = 1$ for unknown region.
- 2) Initialize two sets of GMM i.e. one for foreground and one for background and assign GMM labels to each pixels.
- 3) Using graphcut minimize a Gibbs energy: $\alpha = \operatorname{argmin} E(\alpha, \theta, k, z)$ where, α is background or foreground label, θ is GMM parameter, k is GMM labeling parameter, and z is a pixel value.
- 4) Update GMM parameters and repeat from step 2.

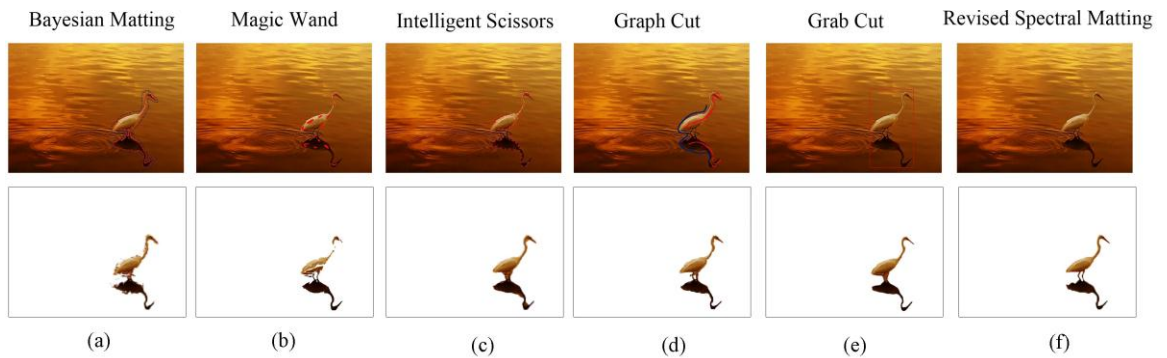


Fig. 3.3: A comparison between segmentation techniques, (a) bayesian matting, (b) magic wand, (c) intelligent scissors, (d) graph cut, (e) grab cut, (f) revised spectral matting (our approach).

In Figure 3.3, the comparison of various segmentation techniques with revised spectral matting technique is provided. The example image is very challenging and chosen due to the reason of same foreground-background color distributions and intensities in the photograph. Revised spectral matting technique performs better for foreground segmentation in the photograph when some alpha components are discarded before summation.

Figure 3.4 shows some more results for foreground (column c) and background extraction (column d) using alpha matte estimation (column b).

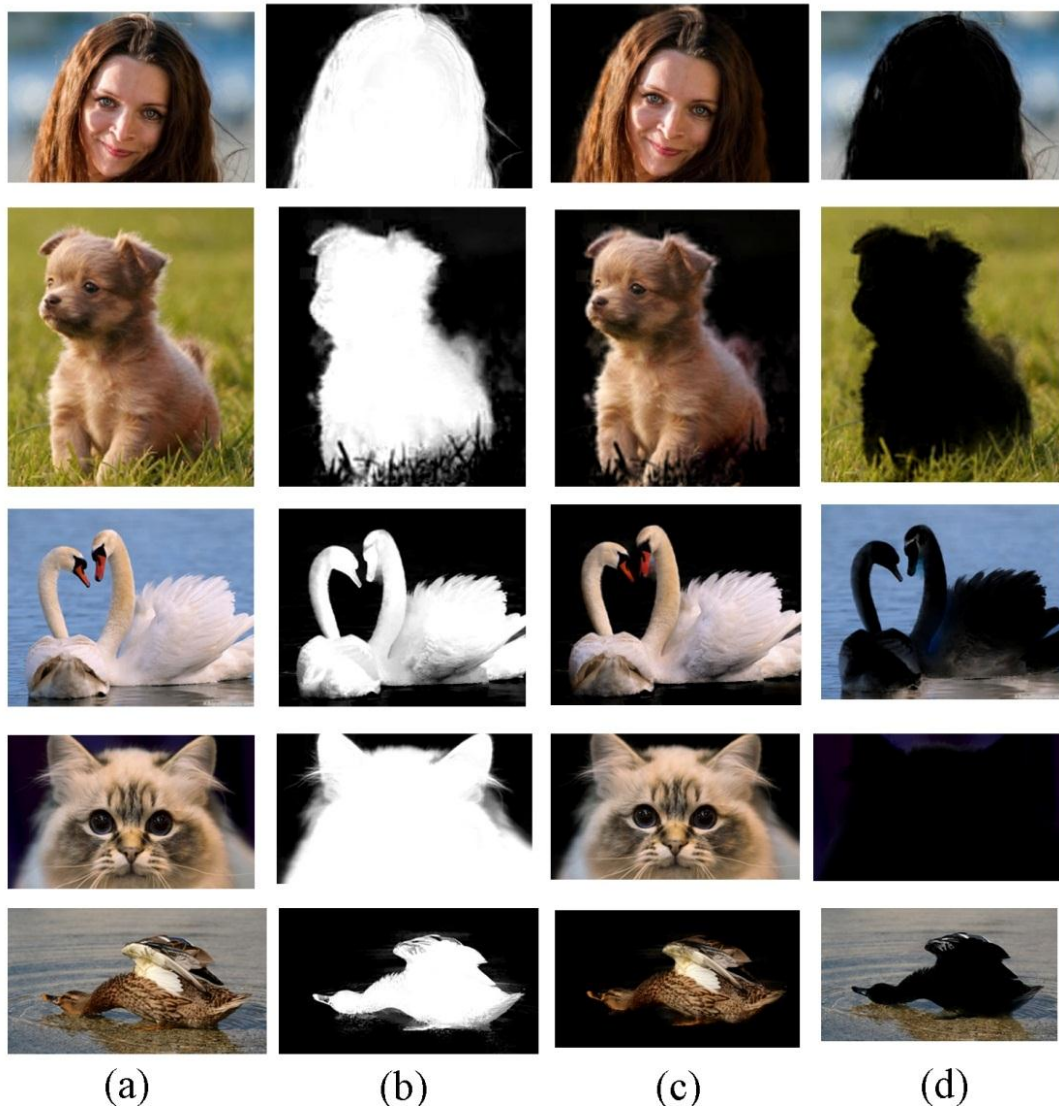


Fig. 3.4: Alpha components estimation and segmentation results, (a) input image, (b) alpha estimation, (c) foreground extraction, (d) background extraction.

4. Segment-based Feature Extraction

Some important features are extracted, such as tilt of an object (θ) with respect to vertical line, size ratio (S_r), area (A), intensity (I_n), shape (M_s), vertical and horizontal position [v_R, h_R], Euclidian distances from thirds-lines (E_d), bounding rectangle (R_β), and centroids (α) on a gridded mesh G_i in the photograph etc.

Tilt angle (θ) is extracted with respect to vertical lines. Major lines along the object are drawn, and a vertical line passing from the center α of the ROI is sketched. The angle between these two lines is called as tilt angle feature. For a tilt angle $\theta > \pm 45^\circ$, the object is shifted towards right or left. The figurative sketch is given in Fig. 3.5.

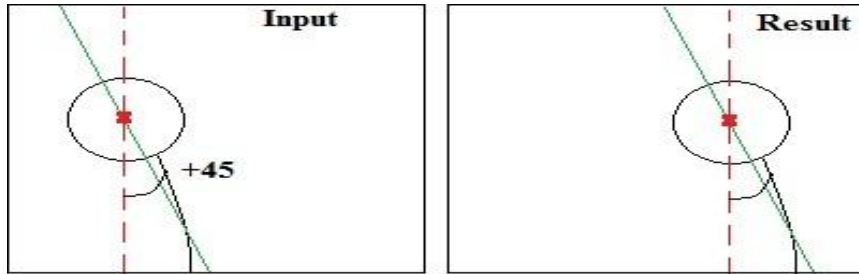


Fig. 3.5: Tilt angle feature to determine the position of the main object.

Size ratio (S_r) of an object can be calculated as $S_r = \left\{ \frac{X}{w}, \frac{Y}{h} \right\}$ where X and Y are the width and height of the image $I(x, y)$, w and h are the width and height of the object in the image. To resize the object in the image for better composition we check:

$$\left[\begin{array}{l} w = \frac{X}{3} \times 2 \quad \text{if } \frac{X}{3} < 63\% \quad \text{else } w = w \\ h = \frac{Y}{3} \times 2 \quad \text{if } \frac{Y}{3} < 63\% \quad \text{else } h = h \end{array} \right] \quad (14)$$

Area (A) of an object is the total count of the pixels $p_i(x, y)$ contained by the main object in the image, such as given in Eq. (15).

$$A = \sum_{i=1}^n p_i(x, y) \quad (15)$$

Intensity (I_n) of an object play an important role in aesthetically pleasing photograph. A dull photograph looks very boring and visually unappealing. The main object is focused in photographs and does have high intensity values than the background. We estimate the intensity (I_n) of the main object as given in Eq. (16). A comparison is then made with the background pixels intensity for placement of the main object according to rule-of-thirds (RoT).

$$I_n = \frac{1}{X \times Y} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_g(x, y) \omega_N^{(x-1)(y-1)} \quad (16)$$

where $I_g(x, y)$ is a gray-level image and $\omega_N = e^{\frac{(-2\pi)}{N}}$ is the Nth root of unity. If the background $M \times N$ pixels above or below and right or left of the main object have intensity values $I_{bg} = \varepsilon + I_n$, we determine the movement of the main object towards down, up, left and right respectively.

Shape (M_s) of an object is another parameter which helps us in understanding the appropriate position of the main object in the photograph. We check the width (w) and height (h) of the object such as if ($w < h$) and ($w \ll X$) where ($h + 0.25(X/3) < Y$) then the shape is assumed to be elongated and the placement is done along the vertical lines of rule-of-thirds.

Position [v_R, h_R] of the main object along vertical and horizontal lines is also obtained for better understanding the current position from vertical and horizontal rule-of-thirds lines. We

can consider this parameter as vector which gives us the location as well as its direction from thirds lines.

Euclidean distance (E_d) is another important measure which helps us retargeting the main object pixels. Euclidean distance is obtained depending on the current position of object $[v_R, h_R]$ as given below:

$$E_d = \sqrt{(v_R - x_1)^2 + (h_R - y_1)^2} \quad (17)$$

Similarly E_d is also calculated for $(x_2, y_2), (x_3, y_3)$ and (x_4, y_4) i.e. intersection points of thirds lines.

Bounding rectangle (R_β) is very important parameter. There is a limit or safe boundary of width, $(X/5.5, Y/5.5)$ for retargeting the object along vertical or horizontal direction of the photograph. The bounding rectangle R_β refrains the object to cross the safe boundaries.

In extraction of centroids, we take an average optimal centroid, α_p of two centroids, α_1 and α_2 as given in Eq. (18) and (19).

$$\alpha_1 = \left(\frac{R_\beta(x)}{2}, \frac{R_\beta(y)}{2} \right) \quad (18)$$

$$\alpha_2 = \left[t * s \left(C * \frac{A}{L} \right) * (v * \eta) \right] + (\alpha_c * u) \quad (19)$$

In Eq. (18), α_1 is the center of bounding rectangles, R_β , and in Eq. (19), α_2 is the composite center of the main object in the photograph [17].

5. Background Best Patch Detection

Background best patch is determined by extracting the features from background after subtracting the main object from the photograph in section 3.1.2. The background features include size of background, (S_B), background sharpness or inversely the blurriness, B_{blur} , intensity of the background, (I_{nB}) etc. The background blurriness is calculated in Fourier domain by taking image's Fast Fourier Transform, $FFT(I_B(x, y))$ as given in Eq. (20). The frequencies present in the result of Fast Fourier Transform give us with the information of the blurriness present in the background. The large amount of high frequencies tell us the image is sharp and vice versa the image is blur for low amount of frequencies in the high frequency bands. A threshold frequency, (f_τ) is set to determine low and high frequencies. We convolved background image with Laplacian kernel 1 to find frequencies as shown in Figure 3.6. Mean intensity of the background, (I_{nB}) is calculated as given by Eq. (21). The results of best patch detection from foreground of the photograph can be seen in Figure 3.7.

$$FFT(I_\beta(x, y)) = \sum_{j=1}^N I_\beta(x, y) \omega N^{(j-1)(k-1)} \quad (20)$$

$$I_{nB} = \frac{1}{XY} \sum_1^X \sum_1^Y [I_B(x, y)] \quad (21)$$

0	1	0	-1	-1	-1
-1	4	-1	-1	8	-1
0	-1	0	-1	-1	-1

Fig. 3.6: Laplacian matrix used for convolving the photographs to get a thresholding frequency.

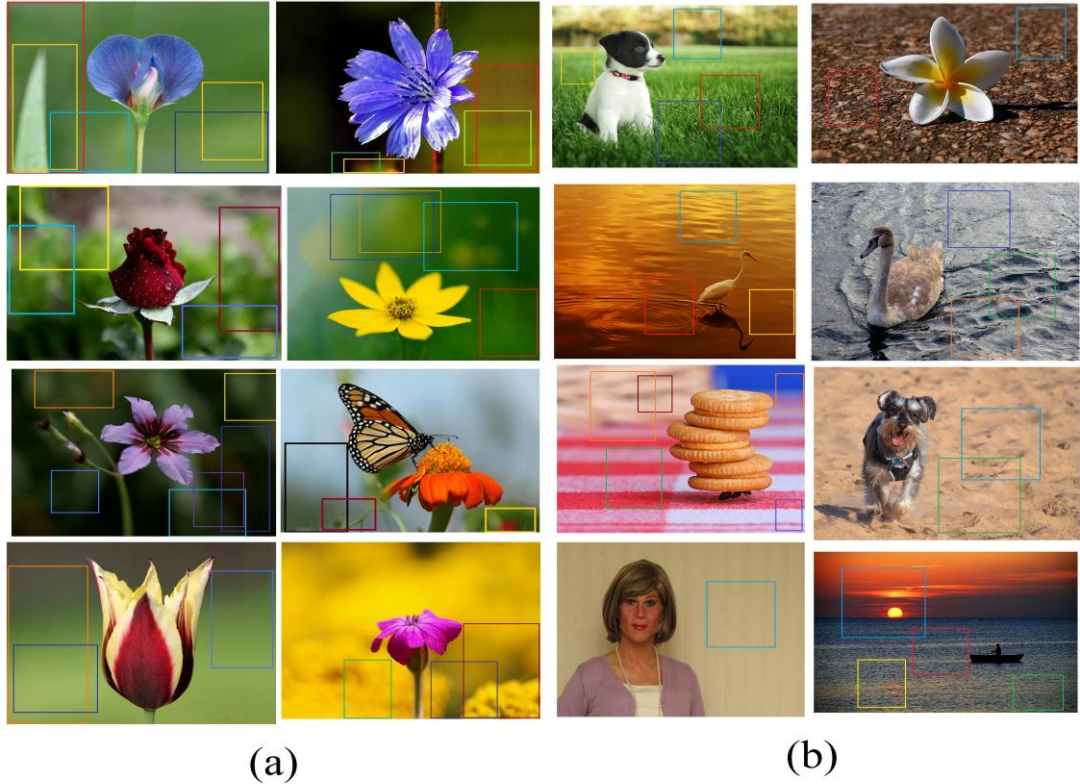


Fig. 3.7: Best patches detected from the photographs based on features.

6. Background Texture Synthesis

Texturing is performed for in-painting the occluded regions in the background. For texture synthesis, background best patch detection based on features is done in previous sections. The best patch is the source which becomes input to the function $f : A \rightarrow R^2$ called *nearest_neighbor_field* (NNF) [18]. Where, f represents a function of offsets. Given patch coordinates $C_a(x, y)$ in target image A (occluded background) and corresponding nearest neighbor patch coordinates $C_b(x, y)$ in source image B (best patch), then $f(a)$ is $C_a(x, y) - C_b(x, y)$. Here we explain NNF in more details.

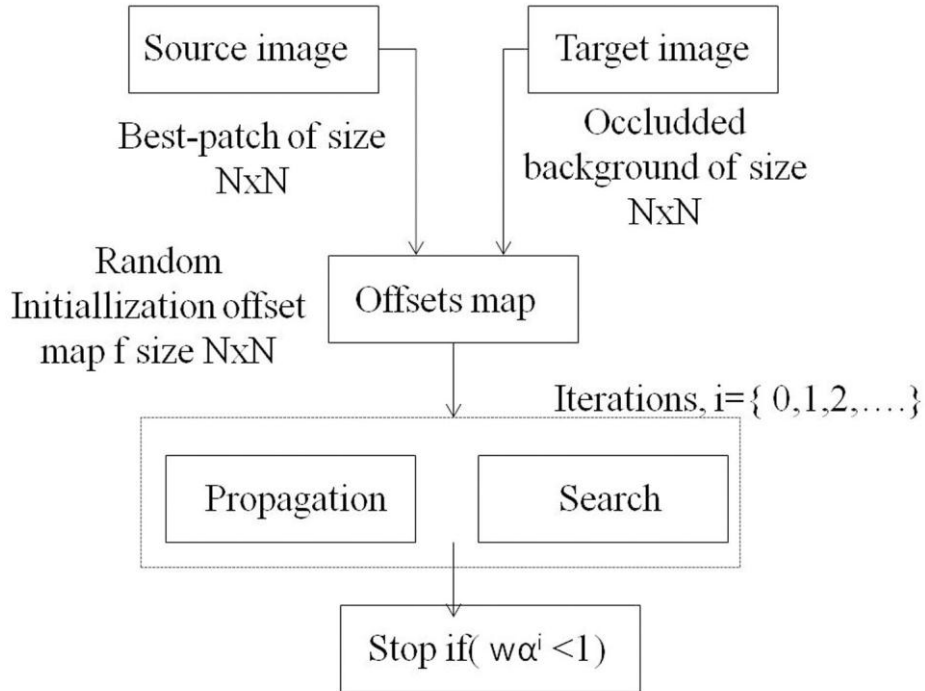


Fig. 3.8: The schematic diagram for texture synthesis by NNF.

6.1 Nearest-Neighboring Function (NNF)

There are three main steps in NNF i.e. (i) initialization, (ii) propagation, and (iii) search. Each of these steps in NNF is explained below:

(i) Initialization

Initialization is done with random offsets by using independent uniform samples from image B. The offset values are stored in an array whose dimensions are those of A . After initialization, an iterative process is performed 4 to 5 times, for improving the NNF. In every iteration, offsets are scanned from left to right, top to bottom order. Each scanned offset undergoes two more steps such as: i) *Propagation*, and ii) *Search*.

(ii) **Propagation**

Propagation step is carried out by assuming the patch offsets are likely to be the same for $f(x-1, y)$ and $f(x, y-1)$ known offsets. The patch distance between the patch at $C_a(x, y)$ in image A and patch at $C_b(x, y)$ in image B is denoted by $D(C_a, C_b)$. The new value of offset for $f(x, y)$ is updated as given in Eq. (22). In even iterations, the propagation is performed in reverse scan order by examining the offsets from right to left, bottom to top using $f(x+1, y)$ and $f(x, y+1)$.

$$f(x, y)_{updated} = \operatorname{argmin}[D(f(x, y)), D(f(x-1, y)), D(f(x, y-1))] \quad (22)$$

(iii) **Search**

For improving NNF, a sequence of candidate offsets is examined at exponentially decreasing distance from v_o as given below:

$$u_i = v_o + w\alpha^i R_i \quad (23)$$

where $v_o = f(x, y)$, w is a maximum search radius, α is a fixed ratio between search window sizes, and R is a uniform random in the range $[-1,1] \times [-1,1]$. The patches are examined for $i = 0, 1, 2, \dots$ until the current search radius $w\alpha^i$ is below 1 pixel. We use $w = \max[\operatorname{size}(I(x, y))]$ and $\alpha = \frac{1}{2}$ in our system. The schematic diagram for texture synthesis by NNF is demonstrated in Figure 3.8. An illustration of results for texture synthesis on occluded background regions is given in Figure 3.9.

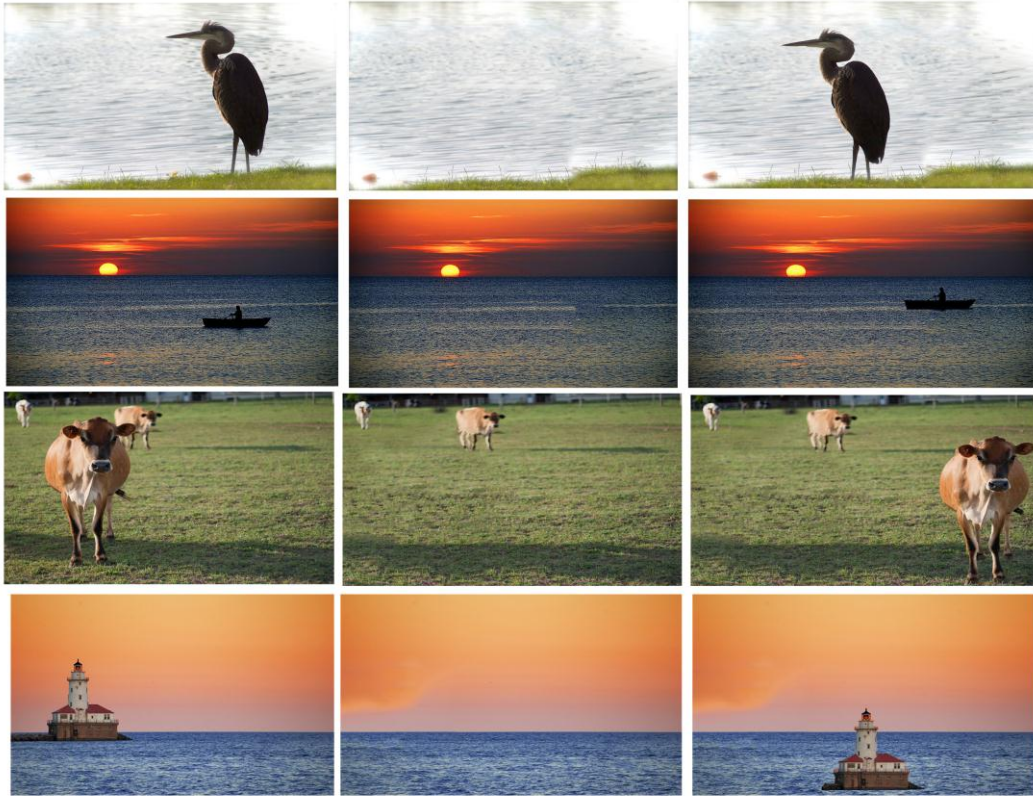


Fig. 3.9: Illustration of results for texture synthesis, (a) original non-floral photographs (NFPs), (b) texturized background, and (c) negative photographs (NPs).

7. Retargeting Algorithm ROI-RT

The main object is retargeted by ROI-RT methods and pasted on the textured background. Before explaining the ROI-RT algorithm in Chapter 4, we explain some necessary details performed. Input image is divided into 9 equal grids, G_i as shown in Figure 3.10, where $i = [1, 2, 3, \dots, 9]$. The grid lines are the vertical and horizontal profiles dividing the whole image into 9 grids and are numbered from $j = [1, 2, 3, \dots, 8]$. The vertical lines from left to right are numbered as 1, 2, 3, 4 and the horizontal lines from top to bottom are numbered as 5, 6, 7, 8. Intersection points of the grid lines are called interest points or the points of focus. Our task is now to properly retarget the objects on a focus point based on background-foreground features to result in the better aesthetically pleasing photographs.

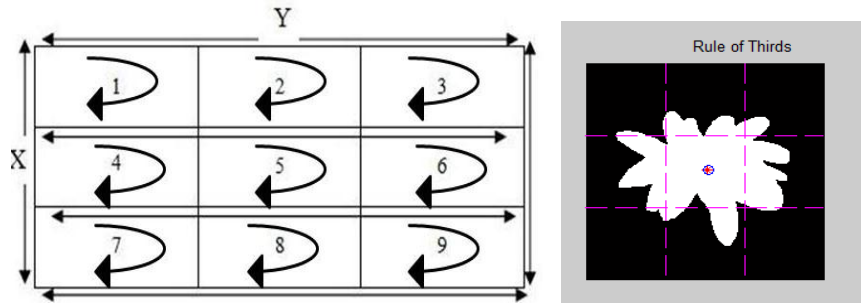


Fig. 3.10: Example of rule of third lines to divide an image into a 3 x 3 gridded mesh.

IV. Retargeting Approach ROI-RT for Better Composition

A. Composition Guidelines

The first goal of better photographic composition is to determine the composition guidelines. In our ROI-RT approach, we make rule-of-thirds (RoT) as a composition guideline, which states that main object of interest in the photograph should be placed along the vertical or horizontal thirds lines or on the focus points (intersections of thirds lines) as shown in Fig. 4.1.

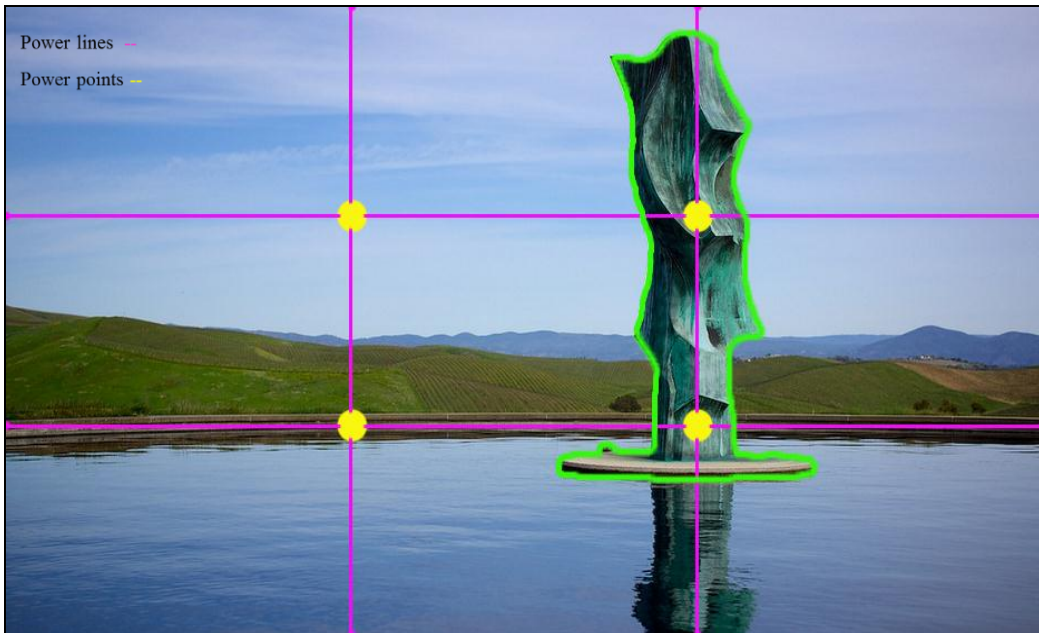


Fig. 4.1: Placement of main object in the photograph according to composition guideline, 'rule-of-third'.

Second composition guideline is balancing elements in our approach. In this guideline, the objects of interest are placed nearby to the center of the photograph to balance the weight as shown in Fig. 4.2 below.



Fig. 4.2: Placement of ROIs in the photograph by composition guideline, ‘balancing elements’.

The third composition guideline in our approach is size ratio, i.e. size of the main object divided by the size of the photograph. This is very important measure that professional photographers follow in their photographs to make high quality and aesthetically pleasing shots. The size ratio composition guideline can be seen in Fig. 4.3.



Fig. 4.3: Visualizing the important composition guideline in photography called ‘size ratio’.

B. Explanation of ROI-RT Algorithm

ROI-RT is used as an acronym for ‘Retargeting Object for Implementation of Rule-of-Thirds’. After extraction of features from the main object and the background in Chapter 3 and

section 4, we apply our computational approaches to retarget the main object on the new textured background. The main steps of ROI-RT algorithm are given below:

Steps in ROI-RT Approach:

Steps in ROI – RT Approach :

1. $IF(M_\beta > G_\beta)$ such that $(W_{r\beta} \leq 2 \times G_\beta)$

2. Update centroid: $\alpha_{p_{new}} = (\alpha_{p_x} + E_{d_x}, \alpha_{p_y} + E_{d_y})$

where, $E_{d_x} = \min(E_{d1_x}, E_{d2_x}, E_{d3_x}, E_{d4_x})$ and

$E_{d_y} = \min(E_{d1_y}, E_{d2_y}, E_{d3_y}, E_{d4_y}) \mid E_{d1} \neq E_{d2} \neq E_{d3} \neq E_{d4}$

3. $IF(E_{d1} = E_{d2} = E_{d3} = E_{d4})$

(i) Change the object size ratio: $S_r = \{X/w, Y/h\}$ by Eq. (14)

and update centroid $(\alpha_{p_{new}})$ in Step 2.

4. $IF(\alpha_{p_{new}} = \alpha_{p_{prev}})$

(ii) Check tilt angle, $\theta > \pm 45^\circ$ and update centroid as given below :

$IF(\theta > 45^\circ \cup E_{d2} \neq E_{d4})$

$\alpha_{p_{new}} = (\alpha_{p_x} + E_{d_x}, \alpha_{p_y} + E_{d_y})$

where, $E_{d_x} = \min(E_{d2_x}, E_{d4_x})$ and $E_{d_y} = \min(E_{d2_y}, E_{d4_y})$

ELSE $(\theta > 45^\circ \cup E_{d2} = E_{d4})$

(iii) Check intensity, I_n and update centroid as given below:

$IF(I_{bg} = \varepsilon + I_n)$, where I_n is given by Eq. (16) and

$(X_k, Y_k) = (R_{\beta k_x}, R_{\beta k_y})$, $k = [1, 2]$

(iv) $IF(\theta \leq 45^\circ \cup E_{d1} \neq E_{d3})$

$$\alpha_{p_{new}} = (\alpha_{p_x} + E_{d_x}, \alpha_{p_y} + E_{d_y})$$

$$\text{where, } E_{d_x} = \min(E_{d1_x}, E_{d3_x}) \text{ and } E_{d_y} = \min(E_{d1_y}, E_{d3_y})$$

$$\text{ELSE}(\theta \leq 45^\circ \cup E_{d1} = E_{d3})$$

Go to step (iii)

5. IF ($M_\beta > G_\beta$) such that ($W_{r\beta} > 2 \times G_\beta$)

Go to Step 3 part (i)

6. Check the result of size ratio, S_r , from above step

$$\text{such that, } \max(R_\beta) \ll \text{Size}\{I(x, y)\} \text{ and}$$

$$\min(R_\beta) \gg \text{Size}\{I(x, y)\} / 3 \times \frac{1}{3}$$

7. Reposition M from position, $P = [v_R, h_R]$ to $\alpha_{p_{new}}(x, y)$.

In this ROI-RT approach given above, β is the size of the foreground in the photograph and M_β is called the size of the main object. G_β is the size of one grid. $W_{r\beta}$ is denoted as $\sqrt{[\min(R_\beta) - \max(R_\beta)]^2}$, is the total width of the main object along horizontal axis. Euclidean distances in vertical or horizontal directions are given by, $E_{d_n} = \sqrt{(\alpha_{p_x} - x_n)^2 + (\alpha_{p_y} - y_n)^2}$, where $n = \{1, 2, 3, 4\}$ for all four focus points. The current position before update of centroid is denoted as $\alpha_{p_{prev}}(x, y)$.

V. Aesthetic Score Assessment

For computation of aesthetic score, we extract visual features and composition features from resultant photographs (outcome of the ROI-RT method) and combine these scores by means of

an aesthetic score function, $ASF = \frac{\sum_{i=1}^n w_i \times f_i}{\sum(w_i)}$ which gives us an aggregate of all aesthetic

scores. Here we mention these visual features as well as composition features which perform well for computation of score.

1. Composition Rule (RoT) Feature

A photograph of width= X , and height= Y , is divided into 9 equal grids with intersection points acting as focus points where important objects are placed e.g. human eyes, mass lines in the photographs etc. There are four intersection points, $P_j = [p_1, p_2, p_3, p_4]$ where $p_1 = (X/3, Y/3)$, $p_2 = (2X/3, Y/3)$, $p_3 = (X/3, 2Y/3)$, and $p_4 = (2X/3, 2Y/3)$. We calculate the distance, $\min(D_j(\alpha_p, P_j))$ for $j=[1,2,3,4]$ and assign weights, $w_j \in (0,1)$. Weight $w_j = 1$ if D_j is minimum for any value of j and otherwise $w_j = 0$. So, RoT feature is calculated as given in Eq. (24) below.

$$\begin{aligned}
 f_{i=1} = & \frac{9}{X \times Y} \frac{1}{\sum_{j=1}^4 w_j} [w_{j=1} \sum_{x=0}^{X/3} \sum_{y=0}^{Y/3} I_S(x, y) \\
 & + w_{j=2} \sum_{x=X/3}^{2X/3} \sum_{y=0}^{Y/3} I_S(x, y) + w_{j=3} \sum_{x=0}^{X/3} \sum_{y=Y/3}^{2Y/3} I_S(x, y) \\
 & + w_{j=4} \sum_{x=X/3}^{2X/3} \sum_{y=Y/3}^{2Y/3} I_S(x, y)] \tag{24}
 \end{aligned}$$

2. Position Feature

The top 3 segments, S_k for $k=[1,2,3]$ are obtained by K-means clustering and position feature for each segment is computed from all focus points, P_j in the photograph to the center of each segment, C_{S_k} , given by f_{i+2} in equation below for $i=[0,1,2]$. Where $D(p1, p2)$ represents the Euclidean distance between two arbitrary points, $p1$ and $p2$.

$$f_{i+2} = \frac{1}{\sum_{k=1}^{k=3} A(S_k)} \sum_{k=1}^{k=3} A(S_k) \times e^{-\frac{D(C_{S_k}(x,y), P_j(x,y))^2}{2\sigma}} \quad (25)$$

where, $\sigma = 0.15$, the values of $j = [1,2,3,4]$, and $A(S_k)$ is the area of the k_{th} segment.

3. Mean Intensity Feature

Intensity feature, $f_{i=5}$ is calculated as the sum of intensity values for, $I_v(x, y)$ in HSV transformation as given below.

$$f_{i=5} = \frac{1}{X \times Y} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_v(x, y) \quad (26)$$

4. Average Saturation Feature

Average saturation is easily computed by taking average of the saturation image, I_s in HSV domain as give in Eq. (27).

$$f_{i=6} = \frac{1}{X \times Y} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_s(x, y) \quad (27)$$

5. Wavelets Feature

Wavelet features for photographs are extracted using 3-level Daubechies decomposition. The nine features are extracted from HSV transformation of RGB images using wavelet coefficients. For example, for saturation image, $I_S(x, y)$ we compute the three level wavelet coefficients at level $k = 1, 2, 3$ to compute the feature f_7 through f_9 as given by Eq. (28).

$$f_{k+6} = \frac{1}{W_k} \left[\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} c_k^{LH}(x, y) + \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} c_k^{HL}(x, y) + \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} c_k^{HH}(x, y) \right] \quad (28)$$

where, $W_k = |c_k^{LH} + c_k^{HL} + c_k^{HH}|$. Similarly, for $I_H(x, y)$, and $I_V(x, y)$ six other features, f_{10} through f_{15} are also computed.

6. Color Feature

Color feature, f_{16} indicates the combinations of colors present in the photographs or in other words, how colorful the photograph is. The dull RGB photographs are not visually appealing, so this feature will help us to measure aesthetics in terms of colorfulness [19]. The image is divided into 64 equal cubic blocks, B_m for $m = [1, 2, 3, \dots, 64]$. For each B_m , sample points are chosen from ideal colorful photographs for evaluating block-wise colorfulness. Earth Mover's Distance is used for evaluating the colorfulness between ideal photographs and input photographs as give below.

$$f_{16} = emd(D1, D2, \{d(a, b) \mid a, b \leq 63\}) \quad (29)$$

where, $d(a, b) = \|I_{LUV}(c_a) - I_{LUV}(c_b)\|$ is the Euclidean distance between the centers of the cubic blocks.

7. Aspect Ratio Feature

Aspect ratio features are photograph's dimensions i.e. area, $f_{17} = X + Y$ and size ratio, $f_{18} = X/Y$. They are simplest features to play an important role for proper understanding of foreground-background ratio's in the photograph.

8. Average Hue Feature

The hue features, f_{19} is computed in HSV color domain for image $I_H(x, y)$ as given in the below equation.

$$f_{19} = \frac{1}{X \times Y} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_H(x, y) \quad (30)$$

9. Aesthetic Score Function

The final aesthetic score is computed by aggregating the scores of individual features, $f_{i=1}$ through $f_{i=19}$ by means of a scoring function, ASF . The scoring function is given in Eq. (31).

$$ASF = \frac{\sum_{i=1}^n w_i \times f_i}{\sum (w_i)} \quad (31)$$

In the above equation, we have given more weights to the features, $f_1, f_2, f_3, f_4, f_{17}$, and f_{18} which are directly related to the composition of the photographs. For, $i = \{1, 2, 3, 4, 17, 18\}$, $w_i = 1$ and for $i = \{5, 6, \dots, 16, 19\}$, $w_i = 0.50$. The final scores are scaled in the range of [0,1] for $ASF = 0.5$ means average aesthetic score of a photograph.

VI. Experimental Results and Discussions

A. Database

In this section, we discuss the experimental results performed on our database. We have taken different sets of digital photographs from online public databases, 1) flicker.com [20], 2) photoshelter [21], and 3) photo.net [22] etc. For simplicity, at first step we have chosen a set of photographs that have blur backgrounds with floral macros (FMB). At later steps, we have included floral macros with difficult backgrounds (FMD). We also have included non-floral photographs (NFP) in our dataset. The original photographs are called positive photographs (PPs). These photographs are further manipulated to make them negative photographs (NPs) i.e. which do not follow the composition rule, rule-of-thirds (RoT). These negative photographs (NPs) are then taken as input to our system. In our database, we have total of 1350 negative photographs. In Figure 6.1 we can see the example set of original positive photographs (PPs) in three categories of FMB, FMD and NPF.

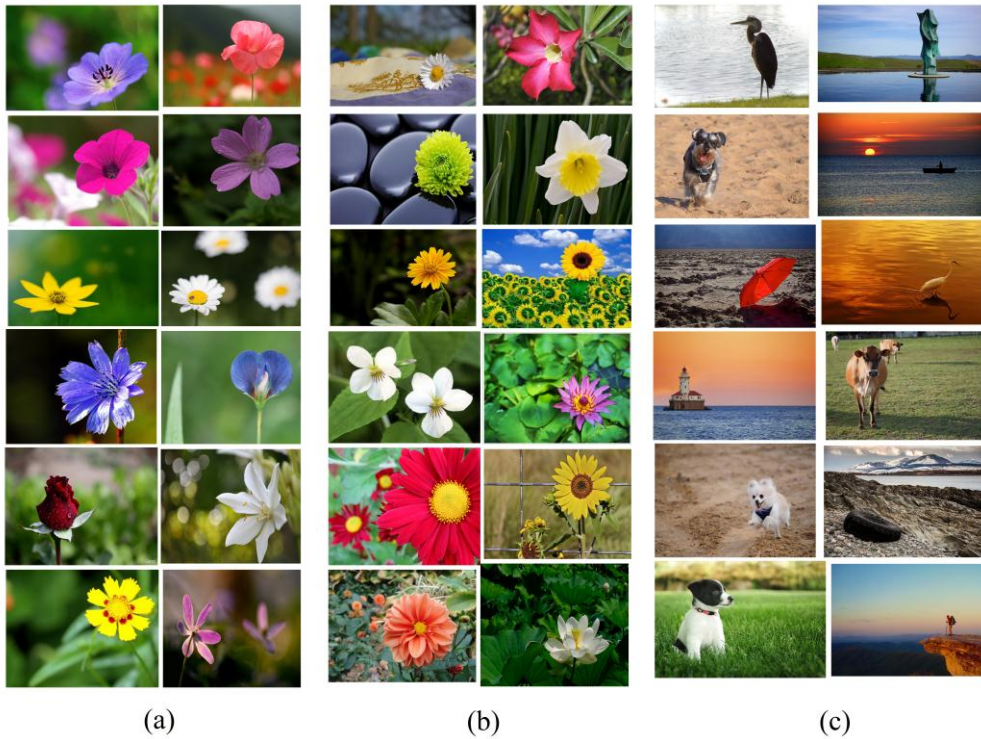


Fig. 6.1: Example set of photographs taken from online public databases [20], [21]-[22].

B. Discussion on Results

The final composition of the photograph is automatically determined by our proposed approach, ROI-RT.

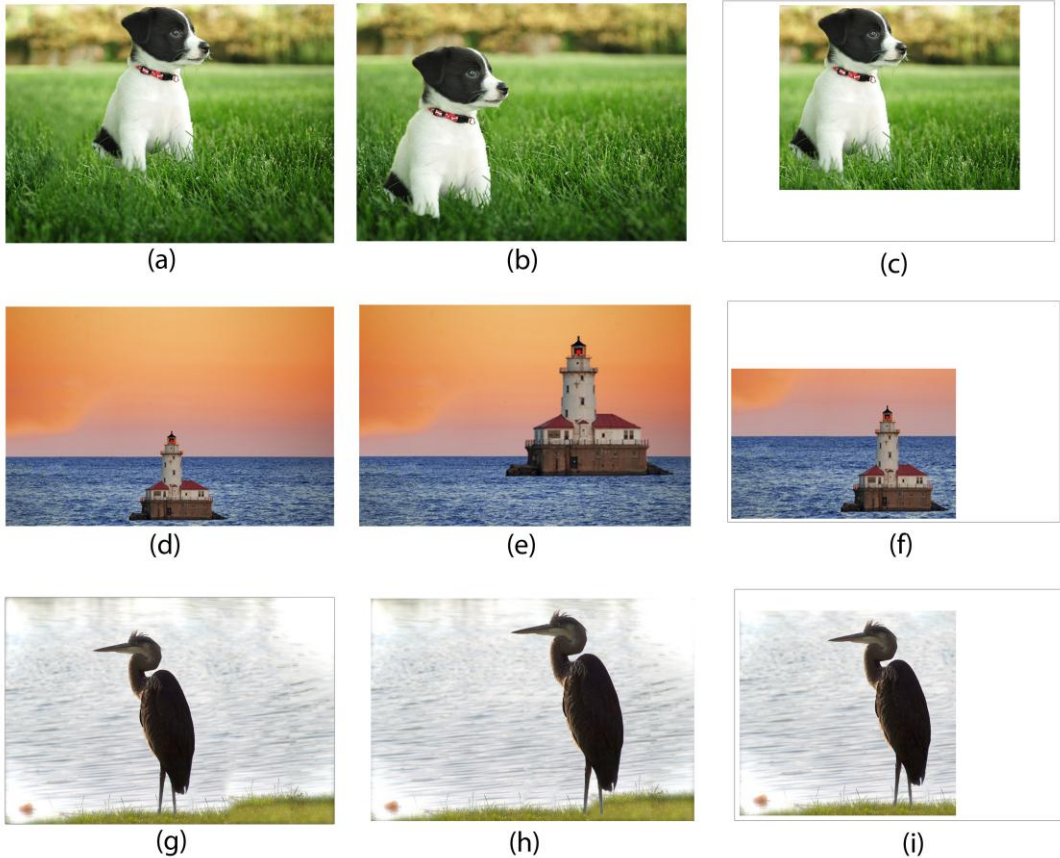


Fig. 6.2: Automatic photo composition by ROI-RT method, (a) – (d) – (g) input negative photographs (NPs) in the first column, (b) – (e) – (h) automatic photo composition by proposed approach in the second column, (c) – (f) – (i) crop and retarget approach [5] given in third column.

In Figure. 6.2, the results of photo-composition are shown. In first column, the manipulated negative images which do not follow rule-of-thirds are given (NPs), in the second column the retargeting results for better composition by proposed approach are given (PR), in the third column we can see the results achieved by crop and retarget (CR) approach presented in [5].

In the approach of Liu *et al* [5], the results are effective only if we have enlarged the pictures from all borders, so the excessive borders are cropped out and the resultant cropped image is retargeted. In the dog image in (c), the result of CR approach is acceptable because there is enough room for cropping the dog in appropriate frame which gives optimal performance. In second row, where light house is near the bottom line as shown in (d), there is not enough room for cropping it, so the CR technique presented in [5] is not effective. It crops the main object i.e. lighthouse and a portion of the sea and a smaller portion of sky with clouds by keeping the size ratio of the object with the size of the cropping window. The picture loses it's inherited semantics by CR approach. In Fig. 6.2, the images shown in the third column are cropped by optimal windows size only but not scaled for better visual understanding of the crop window in the CR approach.

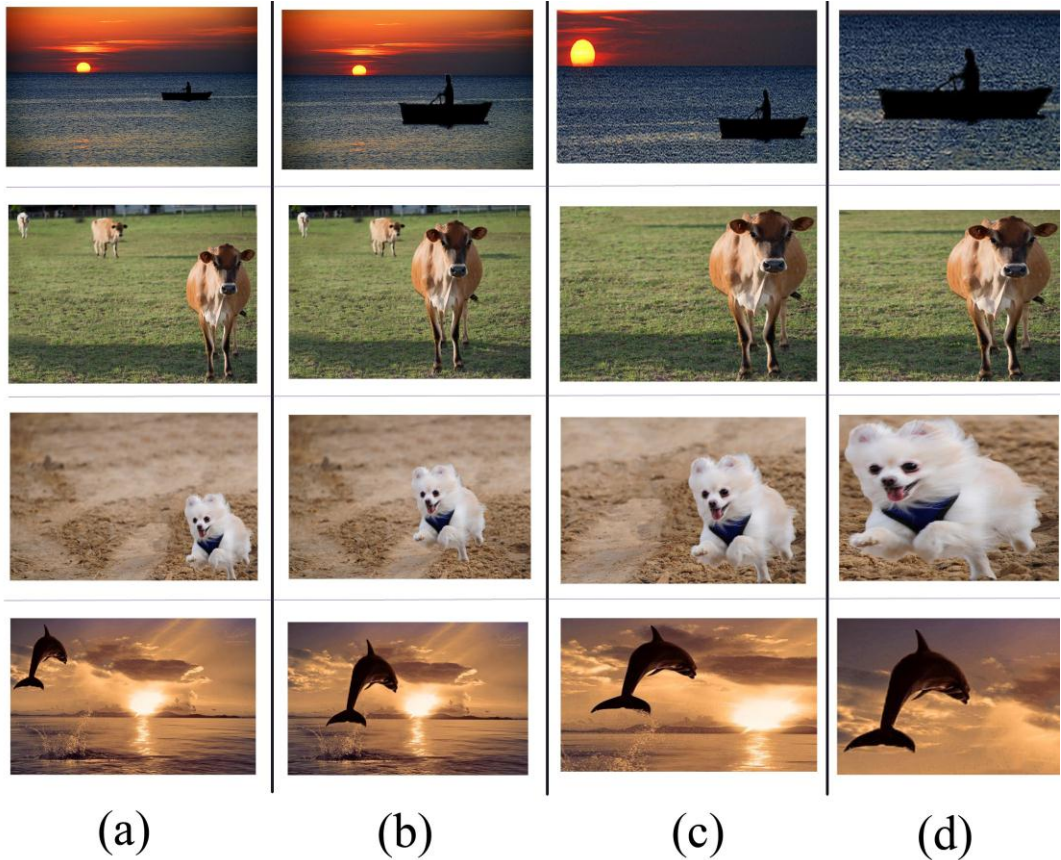


Fig. 6.3: Automatic photo-composition optimization by ROI-RT method, (a) input images (NPs), (b) composition optimization by ROI-RT (our results), (c) results of CR approach [5], (d) results by SOAR approach [23].

In Figure 6.3, some more results of ROI-RT scheme are presented where our approach outperforms then the existing work and other conventional methods. In Fig. 6.3, second column, the boat with the sunset and shades in the sky gives us better visual aesthetics. With the CR approach much of the clouds are cropped out, that is not an optimal result. Although the horizontal lines, thirds lines and size ratio's are kept under considerations. Results by [23] scale and object aware retargeting, (SOAR) crops much of the meaningful portion from the photograph. The semantics of the photographs are not the same both in columns (c) and (d). The cows in the background are cropped out because of the optimal window size in CR approach and also in SOAR approach as shown in Figure 6.3, second row. The third row of the same the figure, where dogs must have some space in the running direction, our approach re-

targets it to one of the thirds position providing optimal composition. In fourth row second column, the dolphin is shifted to proper visually aesthetic point by following the composition rule proposed in ROI-RT. The semantics of the dolphin photograph are not same by CR approach and similarly by SOAR approach.

The results of aesthetic score measurement are shown in Figure 6.4. Some more example of ROI-RT method with their aesthetic score measures are also provided in the Figure 6.5.

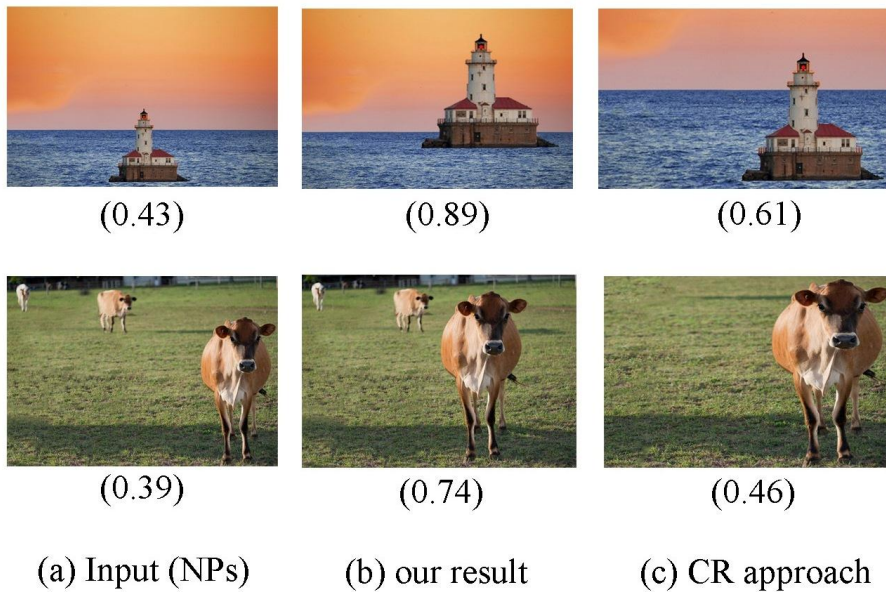
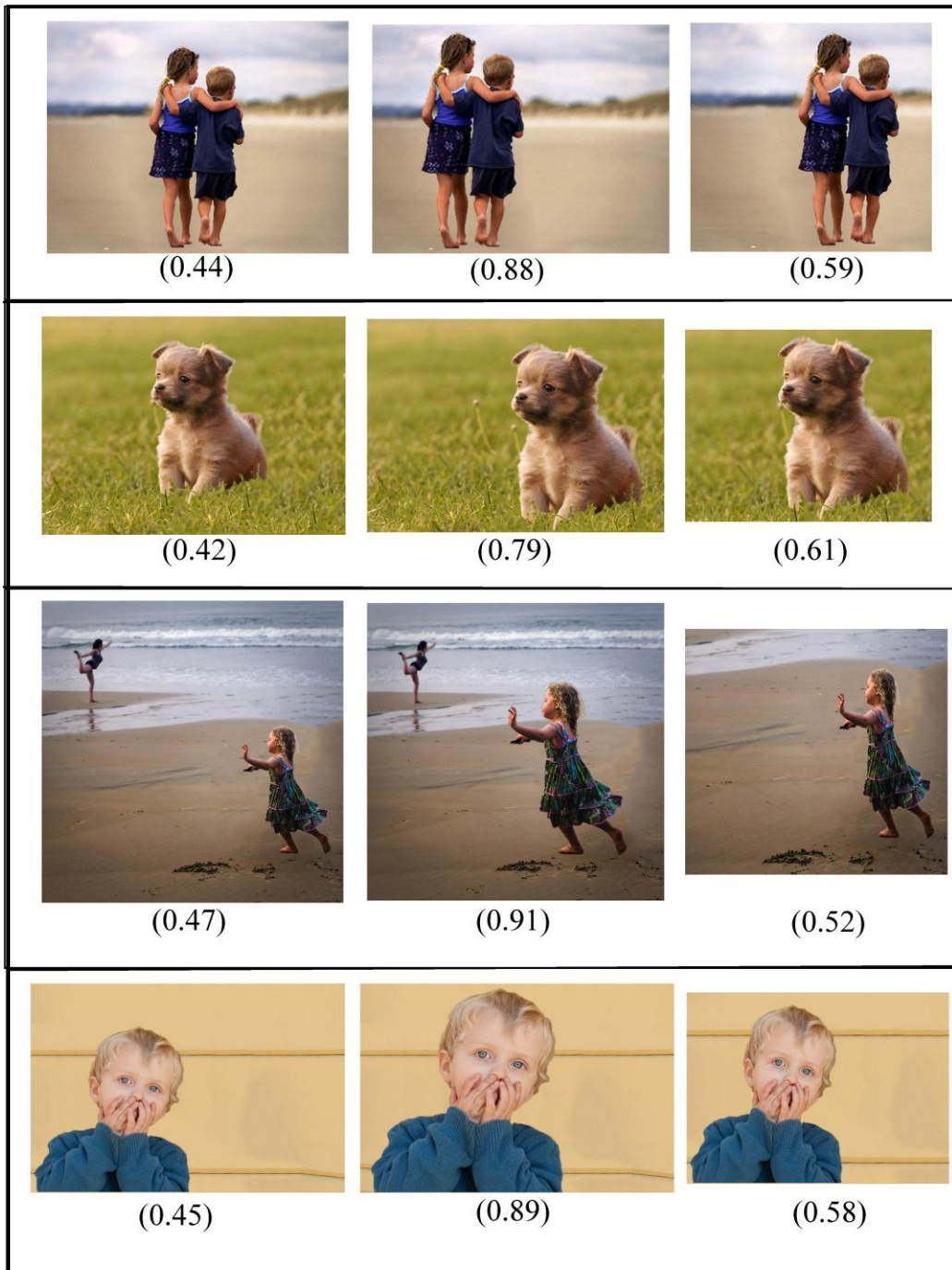


Fig. 6.4: Aesthetic score results, (a) input photographs (NPs), (b) composition optimization results by ROI-RT (our approach), (c) results of CR approach [5].



(a)

(b)

(c)

Fig. 6.5: Photographic composition optimization results together with aesthetic scores, (a) negative photographs (NPs), (b) ROI-RT method (our approach), (c) CR approach [5].

The system accuracy rate in terms of photographs that follow the rule of thirds is given in Table 1 for all three categories of photographs in our database.

Table 1: System accuracy for automatic photographic composition optimization.

Category	Number of Photographs	System Accuracy
FPB	448	98.21%
FPD	453	90.94%
NFP	449	93.76%

Table 2: Efficiency comparison of composition based on aesthetic scores.

Photograph	Original Aesthetic Score	Aesthetic Score for Proposed Composition	Compositional Aesthetic Score by Sun et al., [23]	Compositional Aesthetic Score by Liu et al., [5]
Two Kids	0.44/1	0.88/1	0.32/1	0.59/1
Girl	0.47/1	0.91/1	0.31/1	0.52/1
Boy	0.45/1	0.89/1	0.39/1	0.58/1
Brown Dog	0.42/1	0.79/1	0.36/1	0.61/1
White Dog	0.43/1	0.82/1	0.33/1	0.59/1
Crane Bird	0.41/1	0.79/1	0.25/1	0.62/1
Cow	0.39/1	0.74/1	0.35/1	0.46/1
Light House	0.43/1	0.89/1	0.39/1	0.61/1
Boat	0.49/1	0.81/1	0.33/1	0.58/1
Dolphin	0.34/1	0.83/1	0.23/1	0.55/1

VII. Conclusion and Future Enhancement

In this thesis, we have presented a computational approach for automatic photo-composition optimization of photographs. The technique we presented is named as ‘Retarget Object for Implementation of Rule-of-Thirds’, (ROI-RT). This method automatically improves the composition of photographs resulting highly aesthetically pleasing photographs. This technique keeps the semantics of the photographs by not cropping any background portions out of the photos. The steps we adopted are background-foreground estimation, main object segmentation, features extraction, background best patch detection, texture synthesis by patch match, and retargeting the alpha matte extracted foreground on synthesized background based on ROI-RT to regenerate a photograph which follows photographic composition rules giving better visual aesthetics. This overall system gives good performance on blur backgrounds as well as on the complex backgrounds. The foreground is retargeted on background without boundary visibility errors by parameter-free spectral matting in our approach. The overall system accuracy in terms of photographs which follow the *rule – of – thirds* is 94.30 % which is very effective for achieving the better photographic composition. In the future, we would like to progress our research on automating photo-composition for other famous rules in photography, i.e., view point, depth-of-field, and diagonal dominance etc.

Bibliography

- [1] S. Riaz, K. H. Lee, and S.-W. Lee, “Aesthetic Score Assessment based on Generic Features in Digital Photography,” in Proc. of 5th AUN/SEED-Net Regional Conference on Information and Communications Technology, Manila, Philippine, October 2012, pp. 76-79.
- [2] B. Krages, “The Importance of Visual Skills,” in Photograph: The Art of Composition, New York (NY), Allworth Communication, 2005, ch. 1, sec name: ‘The Importance of Composition’, pp. 7-11.
- [3] A. Levin, A. Rav-Acha, D. Lischinski, “Spectral Matting,” in Proc. of Computer Vision and Pattern Recognition, IEEE Conference on CVPR’07, 2007, pp. 1–8. doi:10.1109/CVPR.2007.383147.
- [4] L. Mai, H. Le, Y. Niu, F. Liu, “Rule of Thirds Detection From Photograph,” in Proc. of IEEE International Symposium on Multimedia, 2011, pp. 91–96. doi:10.1109/ISM.2011.23 .
- [5] L. Liu, R. Chen, L. Wolf, D. Cohen-Or, “Optimizing Photo Composition,” in Proc. of Computer Graphic Forum (Proc. of Eurographics), vol. 29, no. 2, 2010, pp. 469–478. URL: <http://www.math.zju.edu.cn/ligangliu/cagd/projects/composition/>
- [6] T. Ren, Y. Liu, G. Wu, “Image Retargeting Based on Global Energy Optimization,” in Proc. of IEEE International Conference on Multimedia and Expo, (ICME 2009), 2009, pp. 406–409. doi:10.1109/ICME.2009.5202520 .
- [7] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, B. Gooch, “Automatic Image Retargeting,” in Proc. of the 4th International Conference on Mobile and Ubiquitous Multimedia, MUM ’05, ACM, New York, NY, USA, 2005, pp. 59–68. URL: <http://doi.acm.org/10.1145/1149488.1149499>.
- [8] A. Mansfield, P. Gehler, L. Van Gool, C. Rother, “Scene Carving: scene consistent image retargeting,” in Proc. of the 11th European Conference on Computer Vision, ECCV’10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 143–156. URL: <http://dl.acm.org/citation.cfm?id=1886063.1886076>.

- [9] Y.-Y. Chang, H.-T. Chen, “Finding Good Composition in Panoramic Scenes,” in Proc. of IEEE 12th International Conference on Computer Vision, 2009, pp. 2225–2231. doi:10.1109/ICCV.2009.5459470.
- [10] S. Banerjee, B. Evans, “In-Camera Automation of Photographic Composition Rules,” in Proc. of IEEE Transactions on Image Processing, vol. 16, no. 7, 2007, pp. 1807–1820. doi:10.1109/TIP.2007.898992.
- [11] Y.-Y. Chuang, B. Curless, D. H. Salesin, R. Szeliski, “A Bayesian Approach to Digital Matting,” in Proc. of IEEE CVPR, vol. 2, IEEE Computer Society, 2001, pp. 264–271.
- [12] A. Levin, R.-A. Alex, L. Dani, “Spectral Matting (2008),” in Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 10, 2008, pp. 1699-1712. doi: <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2008.168>
- [13] Paint.net Documentation
URL: <http://www.getpaint.net/doc/latest/MagicWand.html>
- [14] E.-N. Mortensen, W.-A. Barrett, “Intelligent Scissors for Image Composition,” in Proc. of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '95, ACM, New York, NY, USA, 1995, pp. 191-198. doi:10.1145/218380.218442 .
URL: <http://doi.acm.org/10.1145/218380.218442>
- [15] S.-N. Sinha, “Graph Cut Algorithms in Vision, Graphics and Machine Learning,” An Integrative Paper, November 2004, UNC Chapel Hill.
- [16] C. Rother, V. Kolmogorov, A. Blake, “‘grabcut’: Interactive Foreground Extraction using Iterated Graph Cuts,” ACM Trans. Graph, vol. 23, no. 3, 2004, pp. 309-314. doi:10.1145/1015706.1015720.
- [17] R. Sidra, S.-W. Lee, “Main Object Segmentation in Marco Photography,” in Proc. of the International Conference on Smart Media and Applications, 2012, pp. 103-104. URL: <http://sidrariaz.webs.com/documents/Final-SMA2012-2012.pdf>
- [18] C. Barnes, E. Shechtman, A. Finkelstein, D.-B. Goldman, “PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing,” in Proc. of the ACM Transactions on Graphics, vol. 28, no. 3, 2009, pp. 1-8.

- [19] R. Datta, D. Joshi, J. Li, J. Z. Wang, "Studying Aesthetics in Photographic Images using a Computational Approach," in Proc. of the 9th European Conference on Computer Vision, ECCV'06, Springer-Verlag, Berlin, Heidelberg, 2006, pp. 288–301.
- [20] Online Public Database @online: Flickr.com
URL: www.flicker.com
- [21] Online Public Database @online: Photoshelter.com
URL: www.photoshelter.com
- [22] Online Public Database @online:photo.net
URL: www.photo.net
- [23] J. Sun, H. Ling, "Scale and Object Aware Image Retargeting for Thumb-nail Browsing," in Proc. of the IEEE International Conference on Computer Vision (ICCV), 2011, pp. 1511-1518. doi:10.1109/ICCV.2011.6126409.

ABSTRACT

A Photograph Reconstruction Method Based on Parameter-Free Spectral Matting for Better Composition

Sidra Riaz

Advisor: Prof. Sang-Woong Lee

Department of Computer Engineering

Graduate School of Chosun University

In digital photography, composition rules are essential for capturing highly aesthetic photographs. Aesthetic images create a response of visual appreciation to the viewers. Rule of thirds (RoT) is the most important and basic rule accepted by photographers for taking aesthetically pleasing shots. The important object is placed along the thirds-lines or intersections of thirds lines. In many cases, the aesthetic quality of the photograph taken from a traditional camera is improved if the photograph is produced to respects the rule of thirds. In this thesis, a novel computational approach, "Retarget Object for Implementation of Rule-of-Thirds", (ROI-RT) is presented. The technique automatically improves the composition of photographs to follow the composition rule, rule-of-thirds (RoT). For achieving the said task, the main adopted steps are background-foreground estimation, main object segmentation, features extraction, background best patch detection, texture synthesis for occluded background, and retargeting the main object on synthesized background to reproduce a photograph which respect composition rule RoT, in photography. Experimental results show an overall acceptable performance of 94.30% for our method. A comparison with existing techniques shows that ROI-RT performs better in terms of visually aesthetic photographs.

ACKNOWLEDGMENT

I would first of all thank Allah Almighty Who enabled me to carry out this project with full devotion and consistency. It is only because of His blessings that I could find my way up to the completion of this task.

Next, I would like to express my immense regards and honest gratitude to my supervisor, Prof. Sang-Woong Lee. His valuable support, guidance, appreciation and supervision, throughout the course of this novel research, motivationally steered me to accomplish this task successfully. I would also like to thank him for supporting me financially to gain a professional experience here in Korea.

I also appreciate the time and efforts of the reviewing committee members, Prof. Moon, In-Kyu and Prof. Yang, Hee-Deok for their critical analysis and precious advices during the preparation of this thesis.

Finally, I also pay my esteem regards to MKE (Ministry of Knowledge Economy), Korea under the Global IT Talents Program supervised by NIPA (National IT Industry Promotion Agency), for its financial support during the period of my Master studies.