博　士　論　文

Fast Image Layer Separation by Exploiting Correlation among Multiple Features

(複数特徴間の相関を活用した高速画像レイヤ分離)

東京大学大学院
情報理工学系研究科
電子情報学専攻

48-147407　余　錦澤

指導教員　　佐藤 洋一　教授

平成２８年１２月

Abstract

Image layer separation for decoupling an input image into different component layers is a challenging task in computer vision and has many practical applications. It is inherently an ill-posed problem and cannot be solved without additional information and assumptions. Two types of image layer separation problems are studied in this dissertation: intrinsic image decomposition and structure texture separation, and they were found to be closely related to each other. A particular focus is placed on robustness and the computational efficiency of layer separation as they are important for applications to many problem domains such as virtual reality and gaming. To this end, the correlation between different kinds of features extracted from a given image was exploited to alleviate the ill-posedness of the problem of layer separation. The image formation in the two different layer separation problems was individually analyzed and the useful and robust correlation of features was exploited. Those features, though exist in all the layers of an input image, have different statistical properties and contain different level of information that specify the intrinsic characteristics of the layers.

The main content of this dissertation is divided into three parts. The first part of this thesis proposes a fast method of intrinsic image decomposition based on a single image. Single image-based intrinsic image decomposition is a critical problem since it facilitates many applications in low-level and high-level computer vision domain, as well as those in computer graphics domain. Those applications normally require the understanding of objects' materials, objects' textures and even environmental lighting and the property of the scene where objects are located, which is highly related to the purpose of intrinsic image decomposition. However, single image-based intrinsic image decomposition simultaneously involves a difficult problem with intrinsically existing ill-posedness, which comes from that the solution space has higher dimension than the dimension of input constraints. Therefore, it worths researchers' concerns on proposing effective methods to make it solvable. Previous work has proposed different kinds of prior knowledge or additional information such as user interaction or knowledge learnt from related data to render this problem easier to solve. A novel edge-

based method of intrinsic image decomposition was proposed in this chapter by exploring the correlation between a chromaticity gradient map and an albedo gradient map to exploit another kind of prior knowledge based on color information. This ill-posed problem was specifically addressed in the use of a single image and exploited the use of the chromaticity gradient map to guide recovery of the albedo gradient map; The Retinex assumption was also regularized by using the ℓ_0-norm to piece-wise flatten the albedo layer. The proposed method is simple yet computationally efficient. Experiments were carried out to evaluate the method both qualitatively and quantitatively on public datasets. The experimental results indicated that our method ran much faster than state-of-the-art methods while achieving comparable performance.

The second part of this dissertation addresses the problem of structure-texture separation, which is another common existing layer separation problem. For some applications, structure-texture separation, especially the output structure layer of separation, can be an important preprocessing step for intrinsic image decomposition and enhance the performance of the latter. A method of structure-texture separation using non-gradient-based descriptors is proposed. An alternative image smoothing approach based on the weighted least square (WLS) framework is especially proposed. The proposed approach incorporates the use of Feature Asymmetry (FA) which can better help locate the contours of objects. FA accurately distinguishes structures and textures as it simulates the response of the human perception system to step edges, contours and lines features, and is also sensitive to periodic patterns. WLS can better smooth out images by including FA as weights while preserving structures. Further, such techniques are employed in our phase-based WLS framework. The recent achievement at accelerating the solution of WLS by transforming the 2D optimization problem into two step recursive 1D optimization problems is used in this work to largely reduce the computation. The experimental results indicated that the proposed approach was effective for structure-texture separation and had low computational complexity, compared to state-of-the-art methods.

The third part of this dissertation explains how a fast intrinsic image decomposition method by using a near-infrared (NIR) image in addition to

a RGB image for better decomposition performance was designed when the influence of fine textures exists. Many object's surfaces contain rich textures, especially those made of materials such as fabrics, plastics, metal and leather. These rich textures makes the problem of intrinsic image decomposition difficult. This is because most of the methods for intrinsic image decomposition largely rely on the edge of objects' structure which could be affected by the edge information of textures. In fact, the task of intrinsic image decomposition is to separate an input image into its material-related properties, known as reflectance or albedo, and its light-related properties which are referred to as shading or illumination. An effective algorithm of intrinsic decomposition heavily depends on the prior sparsity of edges with significant magnitudes in the albedo layer, which would be violated when there exist rich textures that contain a large number of edges of significant magnitude. To solve the problem, An NIR image that appears relatively textureless was adopted beside a RGB image, and a fusion strategy to reduce the influence of texture in the RGB image was designed, and then the feasibility of intrinsic decomposition was increased. The correlation between the RGB image and albedo layer was analyzed in our work through a chromaticity map and a pseudo-albedo map was defined based on both the RGB and NIR images that were included in a maximum-a-posteriori (MAP) model for intrinsic image decomposition. Here, a fast solver based on a half-quadratic splitting scheme and Plancherel's theorem was also proposed to efficiently solve our MAP problem using a fast Fourier transform (FFT). The proposed method outperformed methods of the state-of-the-art edge-based intrinsic decomposition and achieved fast computational speed.

Finally, a conclusion part is added for summarizing the main contents and contributions of the research work in this dissertation, as well as proposing the future potential extensions.

Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

I would first like to express the deepest gratitude to my advisor, Prof. Yoichi Sato, for his excellent guidance, valuable suggestions, great tolerance and kind encouragement he gave me through my Ph.D. study in Graduate School of Information Science and Technology at the University of Tokyo. With his help, I learned how to carry out research with suitable direction and solid thinking, then how to find interesting and vivid research problems and solve them and how to write papers and especially how to make a good presentation to express my idea properly.

I would also like to thank Prof. Imari Sato and Prof. Yiqiang Zheng from National Institute of Informatics, Prof. Cewu Lu from Department of Computer Science of Standford University and Prof. Pheng-Ann Heng from the Department of Computer Science and Engineering of The Chinese University of Hong Kong. Their insight and passion for the research teach me how to face difficulties and enjoy research, and help me grow as a researcher with a positive and optimistic view of life. For their consistent and generous support, I thank them from the bottom of my heart.

I also would like to express grateful thank to all members of Sato Laboratory, especially to Prof. Ryo Yonetani, Dr. Minjie Cai for their kind advice, helpful assistance, technical support and countless hours of useful discussion on the direction and many issues regarding my research and life. I would like to thank the secretaries for their support and kindness.

I would appreciate the final defense committee members: Prof. Kiyoharu Aizawa, Prof. Shinichi Satoh, Prof. Takeshi Naemura and Prof. Toshihiko Ya-

masaki, for their valuable comments and suggestions to my work.

This thesis would not have been possible without generous financial support from Japanese Government (Monbukagakusho) Scholarship program. Through its very kind staffs, Monbukagakusho also provides students accommodation and other convenient life supports. These supports are gratefully acknowledged.

I also would like to thank my friends. Without them, my life in Japan will not be so happy and harmonious.

Finally, all this would not have been worthwhile, but for my family. It is impossible to put into words my feelings of loves and gratitude for my parents. It is their understanding, perpetual support, and unconditional love that make me overcome all the difficulties through all the years of my study in Japan.

December 2016

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

This dissertation concerns about two-layer separation problems, intrinsic image decomposition and structure-texture separation, by studying in particular correlation between multiple image features. It is organized as three self-contained chapters of my preliminary work that has addressed these layer separation problems. The three main topics are:

1. Intrinsic image decomposition using a chromaticity gradient: fast and efficient decomposition of the image into albedo and shading components (Chapter 2),

2. Structure-texture separation adopting feature symmetry: decoupling of the image into structure and texture components with the aid of feature symmetry (Chapter 3), and

3. Intrinsic image decomposition of materials with rich textures employing an extra near infrared (NIR) image: intrinsic decomposition of materials with fine and rich textures by using a red, green and blue (RGB) and NIR image pair (Chapter 4).

This chapter provides an introduction to the layer separation problem, with a clear explanation of the challenges in solving its ill-posedness. The three preliminary efforts proposed in this dissertation are also briefly described as guidelines.

It is an overview of the dissertation in Figure 1.1.

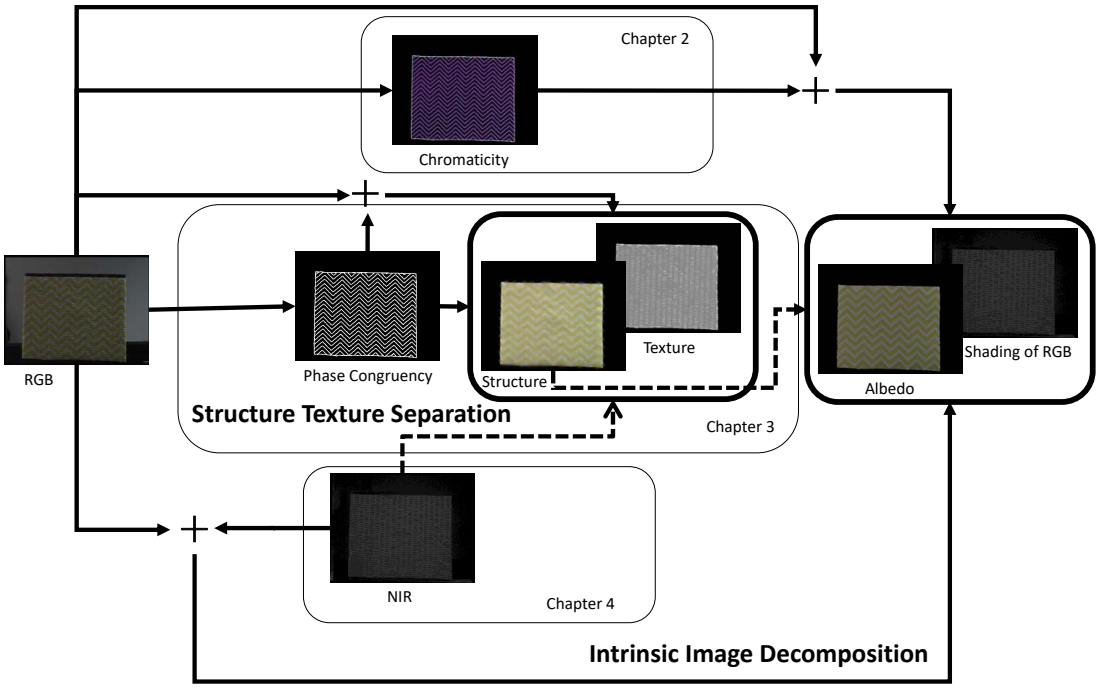


Figure 1.1: Overview of the thesis. Intrinsic image decomposition and structure-texture separation are two problems that are related to each other under the context of images with rich textures.

1.2 The problems with image layer separation

Image layer separation refers to a branch of problems that seek a decomposition of an input image into different layers. The rendering procedure of a photograph was first examined to help describe these problems. The observed color at a given point in a photograph is influenced by many factors, including the shape and the material of objects, their positions, lighting directions, colors of light sources, the position of a viewer, and camera sensor sensitivities. Shadows and specular reflections are effects due to these factors. Material' surfaces can have variations in albedo at different locations, which can be viewed as local textures whereas the whole surface itself is considered as the general structure of objects. These image formation models have been widely researched and are often described by models such as the bidirectional reflectance distribution function (BRDF), the bidirectional scattering-surface reflectance distribution function (BSSRDF) and so on. Image noise often appears as an extra layer added to the original image that reduces its quality. The presence of certain objects in the interactions of

general environments in a scene will occasionally affect the transmission of light rays, which makes the procedure more complicated. For instance, transparent layers (*e.g.* glass windows) in a scene will cause an additional layer of reflection interference. Turbid media (*e.g.* particles like fog, haze and water droplets) in an outdoor atmosphere will lead to low-visibility images. These factors can be modeled with different properties but coexist in the same scene as layers in an image, such that the image is formed as a mixture of multiple layers. Examples of layer separation problems can be seen in Figure 1.2.

The task of image layer separation is to obtain a clear recovery of these layers from the images. Take the two problems in this dissertation, for example, the intrinsic image decomposition in Chapter 2 and 4 are used to separate a given input image into its material-related properties, known as reflectance or albedo, and its light-related properties, such as shading or illumination. Another kind of the image separation problem, discussed in Chapter 3, is structure-texture separation, in which case the structure layer correspond to large major objects or significant edges of corresponding objects in the image, and the textural layer corresponds the fine or quasi-periodic details.

High-quality image layer separation is necessary for various applications. Layer separation can be used as a pre-processing that facilitates other computer vision tasks and computer graphics applications. If we take intrinsic image decomposition for example, there are many image editing applications such as scene relighting, object recoloring and material re-editing based on intrinsic images. Also, a shape-from-shading technique, which infers object geometry from shading changes, requires a clean input image of shading that can also be obtained from intrinsic images. On the other hand, the results of structure-texture separation can be a meaningful input for a lot of low-level vision tasks like tone mapping, image enhancing, inverse halftoning and image retargeting.

The problem of layer separation is inherently ill-posed since it requires more than one layer to be often extracted from fewer input images. There are an infinite number of decomposition solutions to the problem with roughly twice unknown output than known input, as seen in Figure 1.3. Additional information supplied either from user indication, from multiple images, or from the statistical

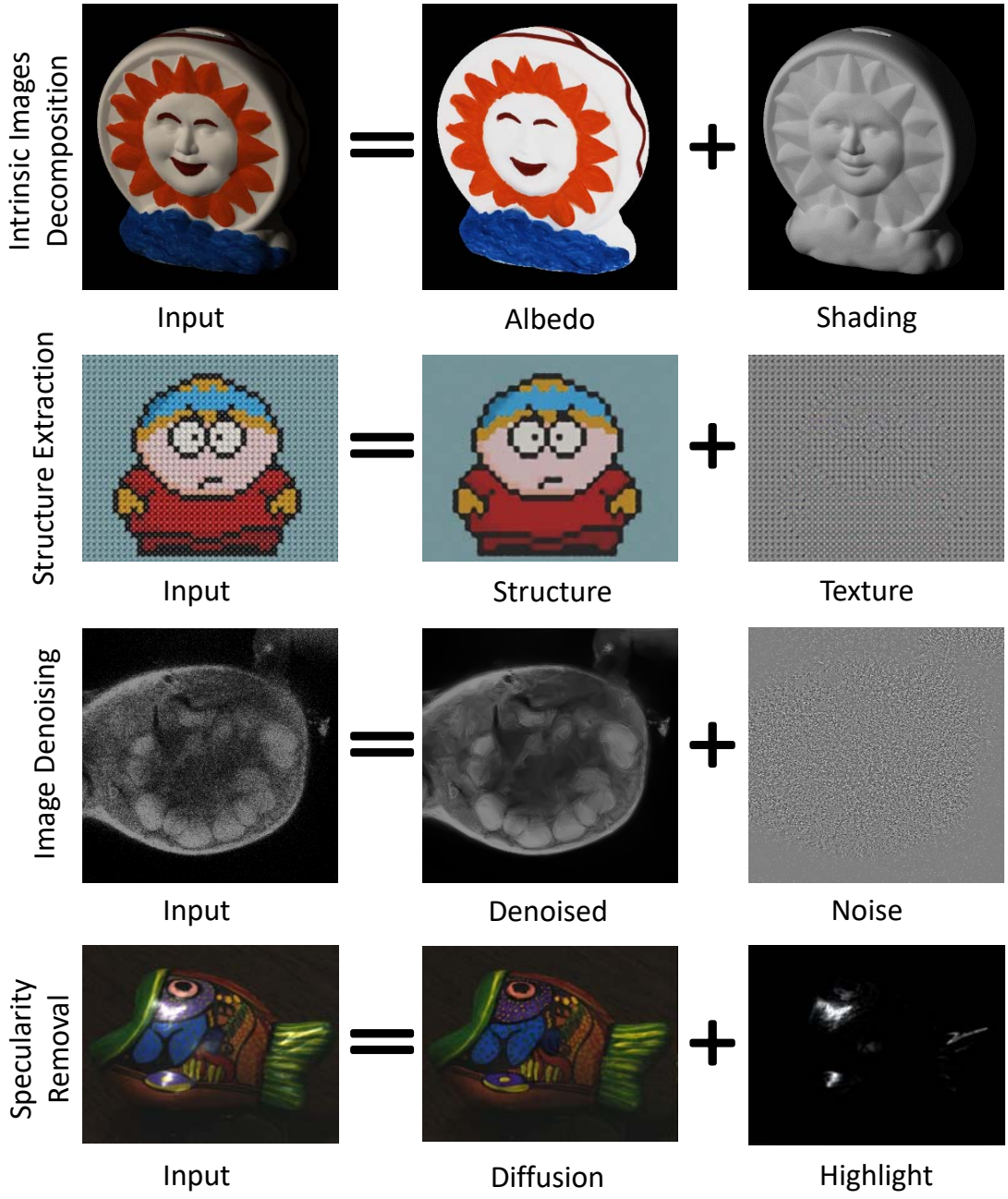


Figure 1.2: Examples of four typical layer separations. The Left side of equal sign indicates inputs and the right indicates separated layers from input. Example images (from up to down) are from [GJAF09], [XYXJ12], [MCB+12], [TNI04b]

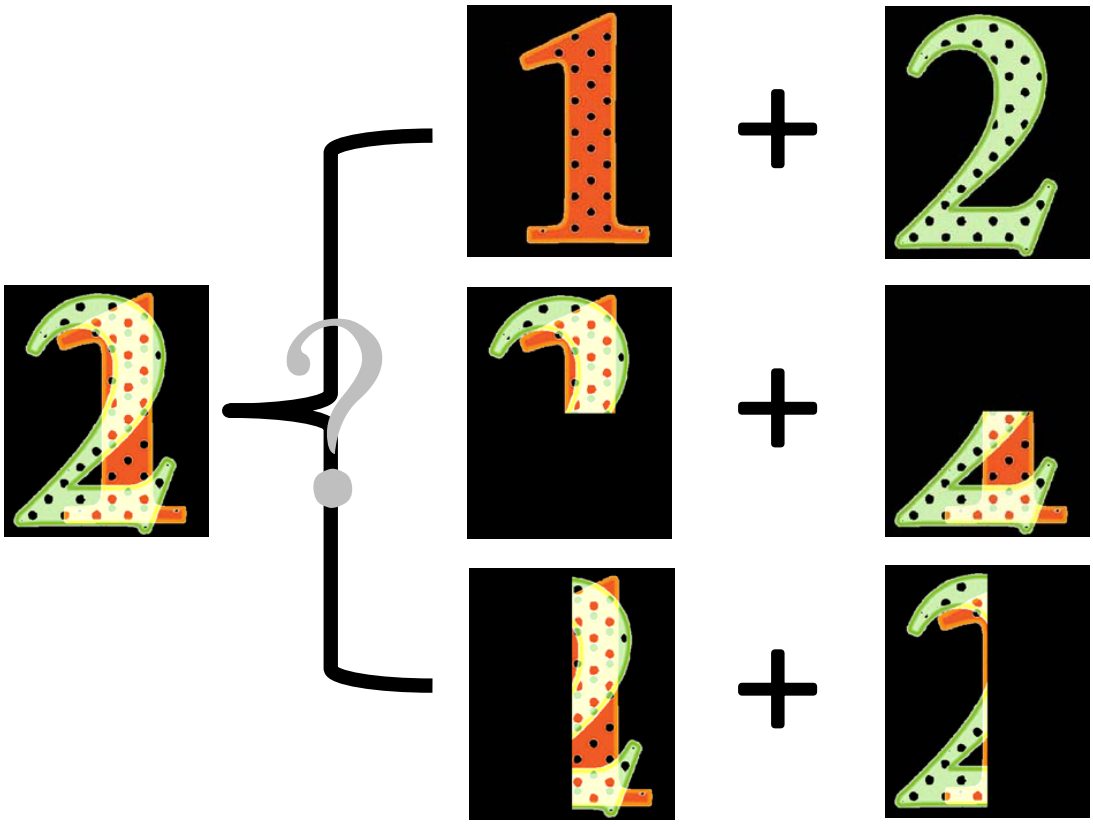


Figure 1.3: This figure shows three possible decompositions to the synthetic mixture image. The ill-posedness of layer separation can result in more possible decompositions besides these three.

information of individual layers is needed to make this problem tractable. Ill-posedness can be eased with this information as constraints, and a most likely explanation for the two layers can be found.

Among all the layer separation problems, two related problems are looked at in this dissertation: intrinsic image decomposition and structure-texture separation, whose imaging model will be given in the following sections. The two problems are related to each other as we work on the images with rich textures. In fact, for a normal object, its surface sometimes contains local variations of albedo which are considered as local textures whereas the whole surface itself is considered as the general structure of objects. Therefore, removing textures from objects' surface help us to obtain a more uniform structure of objects, which can further help understand the albedo of objects' surface. On the other hand, statistically, structure layer normally has the piece-wise smooth property, whereas albedo layer

has piece-wise constant property. Under such situation, the hidden layers for these images will first be layers of structure and texture. Moreover, apart from the structure layer, the following hidden layers will be albedo layer and shading layer. (Figure 1.1)

1.3 Intrinsic Image Decomposition

There are many literatures works addressing layer separation problems. However, most of those problems remain challenging. For instance, intrinsic images problem is first proposed in the 70s [BT78]. Surprisingly besides the inventions done in the 70s and 80s, little research has been done in following decades on intrinsic image decomposition. A comparative study [GJAF09] in 2009 has shown that the color version of simple Retinex algorithm [LM71] from the 70s is still the top performing approach. However, most of the works either require additional equipment or make a strong assumption on the scene's properties like planar geometry, making it difficult to bring them into practical use. More recent work [SYJL11, SY11, GRK+11] showing relatively better results are the first step in this direction. The proposed approaches in solving these layer separation problems focus on improving the quality as well as the efficiency and the practicalness of the solution.

1.3.1 Narrowband Sensitivity

In many algorithms of intrinsic image decomposition as well as other computer vision algorithms, it is common to assume that the camera sensitivity has a narrowband that follows Dirac delta functions. This assumption is an approximation since no commercial camera has such a function. However, this assumption is useful and important since, without it, it is difficult to separate the surface reflectance and illumination color under an integral function. Unfortunately, not all cameras can be approximated to have a narrowband sensor because of the wide distribution of one or more sensors.

The basic image formation model of intrinsic image decomposition is deduced based on the assumption of narrowband camera sensitivity curve. Departing from this assumption, intrinsic image decomposition turns its imaging model from mi-

crophsics optical model into middle-level multiplication of albedo and shading, depending on the practical applications problems that are focused on.

1.3.2 Retinex Method

The earliest model and also the base for intrinsic image decomposition is the Retinex model. Retinex theory was first proposed by Edvin Land in 1964 [LM71]. According to Land [LM71], an image is composed by two parts namely the incident light and the reflectance of the object. The two parts were lately reformed as reflectance or albedo layer and illumination or shading layer. This can be represented by

$$I = RL, \tag{1.1}$$

where L represents the value of incident light, R represents the value of object's reflection, and I represents the value of reflected light. Alternatively, it can be written as following in *log* space:

$$log(I) = log(R) + log(L) \tag{1.2}$$

Retinex is based on color constancy of human visual system. It is called "Retinex" because it is an method bridging the gap between images and the human observation of scenes. Color constancy is a human's ability that properly recognize the object color under even various source of lights such as sunlight or fluorescent light [XLnGhMq09]. If we can removes bias of source lightness from the input image, we can obtain real image without color change. Appling this property, researchers utilized Retinex model for intrinsic image decomposition [Hor74]. The key point of Retinex method is the estimation of the reflectance part. It is suited to handle the images that the local lightness is very low. In the proposed work, we will re-explain the Retinex assumption from another point of view and exploit the piece-wise constant property, and the property of relative smoothness between intrinsic layers deduced from it.

Figure 1.4: Removing bias of source lightness by Retinex method (MSRCR) [HRRR11].

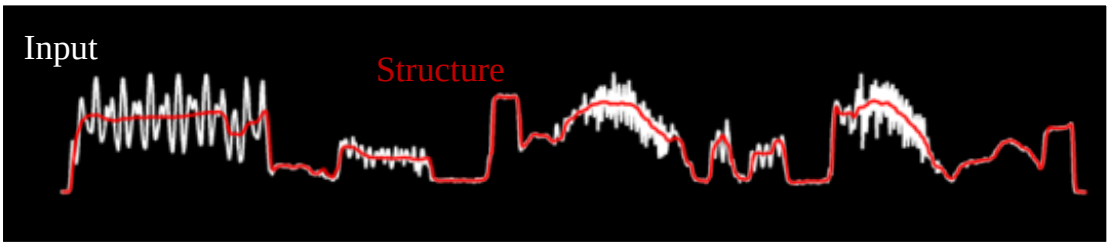


Figure 1.5: 1D signal and its corresponding structure signal.

1.4 Structure-Texture Separation

Structure-texture separation problem, also understood as structure-preserving image smoothing, is an important topic for both computer vision and computer graphics, as such a separation can help better image analysis and image understanding. Many image processing problems can be carried out more effectively and can be well achieved once we separate a scene into two layers possessing different properties of a scene. The definition of structure layer and texture layer is based on their different property in the signal domain. In 1D case, structure signal appears as step edge (Figure 1.5), whereas texture signal appears as small or periodic vibrations. In 2D case, a structure layer represents surface contour or objects' boundaries, whereas texture layer represents small repeated or periodic pattern (Figure 1.6).

For decomposing the original input image into two layers: structure and texture layers, the input image I can be considered as the superimposition of the two

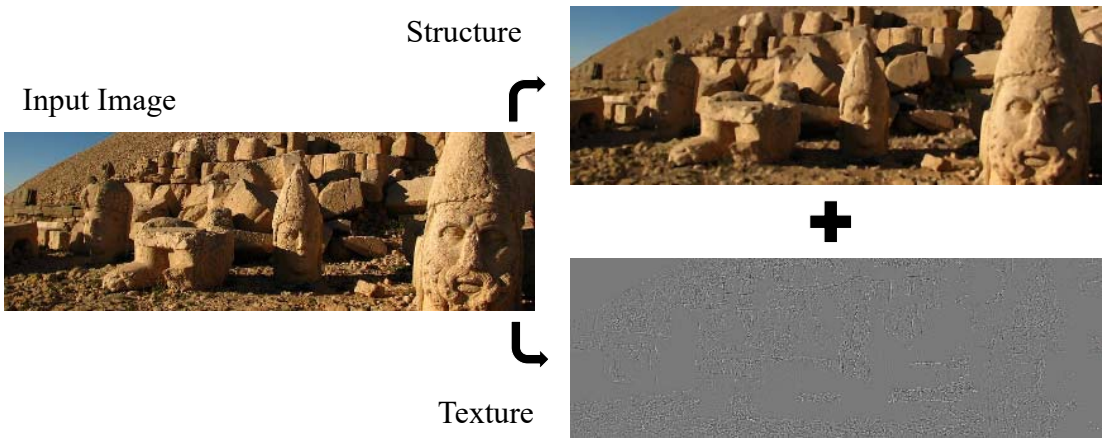


Figure 1.6: Example of 2D image and its corresponding structure and texture layers (image from [KEE13]).

layers:

$$I = S + T, \tag{1.3}$$

where S is the structure layer corresponding to the main large objects in the image, and T is the texture layer corresponding to the fine details [AGCO06]. Example can be seen in Figure 1.2.

Local and global optimization edge-aware smoothing operation (*e.g.* bilateral filter [TM98], weight least square filter [FFLS08]) can be applied to separate an image into a structure layer and texture layer. This procedure produces an image that retains strong structure with details smoothed away. By taking this image as the structure layer S, the texture layer is calculated as $T = I - S$.

This structure-texture separation exploits the fact that most of the structure layer is related to larger but not periodic gradient magnitudes, while the texture layer captures fine and quasi-periodic image details. The most typical and representative method is the total-variation (TV) model based on Rudin-Osher-Fatemi method [ROF92]. The TV model is largely used for image reconstruction, denoising and a few other image processing applications. Based on the TV model, the structure layer S can be obtained by minimizing the following optimization problem:

$$\min_{S} \sum_{p} (S_p - I_p)^2 + \lambda |\nabla S_p|, \tag{1.4}$$

where p means operation is pixel-wise, λ is the regulation parameter. The re-

sulted structure layer from Equation 1.4 captures the significant edges of the scene, whereas the residual $T = I - S$ captures the removed noise layer which contains the texture patterns.

1.5 NIR Imaging

There have been some recent applications showing the benefits of combining NIR imaging with visible photography. These approaches use NIR either to enhance details or to reduce noise/artifacts in the visible image. The key to these applications is the method for combing the NIR and visible image. For example, in the case of image fusion, the low-visibility property of textures in NIR image is exploited for enhancing the visibility of important features in the RGB image.

Normally, the visible spectrum perceived by human visual system ranges from (about) 380 to 780 nanometers in wavelength. The NIR spectrum is located just after the red wavelength and comprises wavelengths that range from 780 to 1100 nanometers. In photography, NIR delivers sharp images with sometimes a dramatic outlook (for example, see some example images from [BS11] in Figure 1.7). Even though the NIR band is located next to the visible one, because the spectral reflectivities of objects in the NIR and visible range vary significantly, there is, in general, almost no correlation between a visible and NIR signal (*i.e.*, knowing the color and brightness of an object gives no information about its NIR response).

NIR spectral imaging has long been used in remote sensing and industrial vision applications. NIR can now be easily used for enhancing photographs and has given rise to methods that combined NIR with conventional visible photography to perform image enhancement, such as contrast boosting, haze removal, blemish reduction, and ISO denoising. Since most of these applications are photographic in nature, it is hard to assign a quantitative value to their results. Most approaches simply show the before-and-after images, where the after-image has been clearly enhanced.

Visible image NIR image

Figure 1.7: Examples of visible and NIR image pair from [BS11].

1.6 Contributions of Dissertation

The main goal of the proposed work was to design robust and efficient methods of solving the layer separation problem. Since image layer separation is a broad topic, our study could not simultaneously address all layer separation problems. Two related problems, intrinsic image decomposition and structure-texture separation, were chosen. Three issues were specifically identified to address:

[Exploiting Correlation of Albedo and Chromaticity Gradients for Fast Intrinsic Image Decomposition]

The correlation was analyzed through normalized cross correlation (NCC) between the chromaticity map and albedo layer through the chromaticity map, and between the albedo and chromaticity gradient. Here, this correlation will be explained via deducing some physics properties of chromaticity and albedo. The chromaticity gradient will be introduced to the intrinsic image decomposition problem for the first time, and be mathematically explained that it contains robust information on the albedo edge. A fast solver based on Plancherel's theorem and a half-quadratic splitting scheme [WYYZ08] is proposed to efficiently solve the proposed intrinsic image decomposition model, which is a problem of ℓ_0-prior-based optimization,

and how the computation was speeded up while maintaining the accuracy of decomposition is explained. The proposed method ran much faster, and the results of the proposed method were close to or better than the state-of-the-art methods that use sophisticated models and inference. Part of this work has been published in ICIP'2015 [2] as a poster presentation.

[Exploiting Feature Asymmetry in Spectral Space for Structure-Texture Separation]

Feature asymmetry (FA) quantity was adopted to structure-texture separation problem. FA is a phase-based quantity that extracts structure edge information in the spectral domain and has better performance than the intensity gradient in the problem being considered. The WLS framework was employed to solve structure-texture separation, whose weight was calculated by including the FA quantity. The proposed FA-aware WLS method was effective in structure-texture separation and could achieve state-of-the-art results. Instead of solving a 2D problem, it was transformed to solving a problem of two 1D recursive iterations when solving the WLS framework by using the existing acceleration framework. The proposed optimization was greatly speeded up and ran much faster than many state-of-the-art methods. Part of this work has been published in the short paper track of PG'2015 [1] as a poster presentation.

[Exploiting Texturelessness of NIR Image for Intrinsic Image Decomposition of Materials with Fine-Textures]

A reasonably plausible ground truth was built composed of shading layers and some albedo layers of materials in the dataset provided by Choe *et al.* [CNK16] using a photometric stereo method. We proposed to include NIR image in the framework of intrinsic image decomposition by applying the textureless properties of some materials for the first time, which helps to largely exclude the effect of texture and yields stronger clues to the Retinex assumption. The correlation between the RGB image and albedo layer through an ℓ_1 chromaticity map, and the correlation between the NIR image and shading layer. were analyzed. A pseudo-albedo map based on RGB and the NIR image was defined and was included in a maximum a posterior (MAP) model for intrinsic image decomposition. A fast solver based on Plancherel's theorem and a half-quadratic splitting scheme [WYYZ08]

was proposed to efficiently solve the proposed MAP problem and speed up the runtime while maintaining the accuracy of decomposition. The proposed method ran much faster, and the results of the proposed method were better than those obtained from state-of-the-art edge-based methods that use sophisticated models inference.

1.7 Organization of Dissertation

This dissertation is organized into five chapters. This chapter 1 provides an introduction to layer separation problems and presents the proposed main goals and contributions. It also provides some background on two-layer separation problems the proposed work involved – intrinsic image decomposition and structure-texture separation, as well as some image features that are related to the proposed works. The next three self-contained chapters (2, 3, 4) describe the proposed work that addresses intrinsic image decomposition and structure-texture separation and give details on prior work that addressed these problems and how these two problems are related to each other. The whole dissertation is concluded in chapter 5 with a discussion on future directions.

Chapter 2

Fast Sparse Edge-based Intrinsic Image Decomposition Guided by Chromaticity Gradient

2.1 Introduction

Intrinsic images estimation is to separate a given input image into its material-related properties, known as reflectance or albedo, moreover, its light-related properties, such as shading, shadows, specular and highlights. This idea of representing distinct scene properties as separated intrinsic images is firstly introduced by Barrow and Tenenbaum [BT78]. Under the Lambertian assumption, intrinsic images decomposition turns out to be decomposing an input image into albedo (or reflectance) layer and shading (or illumination) layer, which will be our focus here.

For the intrinsic image decomposition problem, the observed color at a given point in an image is influenced by many factors, including the shape and the material of objects, the positions and colors of light sources, a viewer's position, and the sensitivities of a camera sensor, *etc.* Successful decomposition of these properties would benefit some computer vision tasks, such as materials re-editing, shadow removal, shape-from-shading, segmentation, scene understanding and object tracking.

The intrinsic image model [BT78] assumes that an image is the product of a scene's albedo and shading at each pixel expressed as $I = R \cdot S$, where R is the

reflective property or albedo at each pixel and S is the shading falling on this pixel. Intrinsic image decomposition's aim is to estimate R and S given an input I.

However, intrinsic image decomposition problem is inherently challenging as it attempts to obtain two unknowns (R and S) from a single input I. Estimating R and S is an ill-posed problem. It requires extracting more layers (unknowns) from one input (knowns). To make the problem tractable prior knowledge on the solutions must be imposed. One of the earliest work addressing intrinsic image decomposition was the Retinex algorithm [LM71] that employed simple heuristics that assumed strong edge gradients belonged to albedo changes in the image. Other intrinsic image decomposition methods using multiple images [Wei01, MLKS04] or using user markup [BPD09] have been proposed and shown to produce good results. For automatic single image intrinsic image estimation, many later works [TFA05, STL08, GJAF09] followed the idea of the Retinex algorithm and focused on separating albedo and shading edges. These methods are referred to as edge-based methods according to a recent survey [GJAF09]. In this survey, the authors proposed the color version Retinex algorithm. Moreover, a ground-truth dataset (the MIT dataset) for intrinsic images containing 16 real objects was also built [GJAF09]. Shen *et al.* [SYJL11] used local information and formulated intrinsic image decomposition as a minimization problem.

The recent progress in probabilistic graphics models and numerical optimizations methods result in more effective intrinsic decomposition methods. Unlike edge-based methods which rely on local edge information, these new approaches use the idea that the set of albedo value present in the scene is sparse, which is usually referred to as global sparsity prior [GRK+11, SY11, SPBV12, BM12, BM15, CCFI14]. These approaches achieved significant results on the MIT intrinsic dataset but tend to be very slow. Another dataset was released by [BBS14] which contains a large number of natural scene images. The authors also proposed a dense CRF-based intrinsic decomposition algorithm which outperforms other state-of-the-art methods on the proposed dataset.

The chromaticity invariance against shading was well known and has been adopted for several applications before [Dre03, FHLD06]. For example, Drew*et*

al. [Dre03] adopted the relationship between R, G, B channels in logarithmic space to recover a shadow free chromaticity map. Finlayson *et al*. [FDL04] included the 2D chromaticity representation of lights and cameras to remove the shadow from images. They [FHLD06] later proposed an intrinsic decomposition method based on entropy minimization in the log-chromaticity space. Using this property of chromaticity, a novel edge-based method was proposed to solve the intrinsic image decomposition by adopting the chromaticity information of the input image based on the Retinex assumption [LM71]. In our case, as can be seen in Figures 2.1, 2.2 and 2.3, the chromaticity map contains similar edges as those in the albedo layer, and rich chromaticity gradients indicate that albedo gradients are also rich at those locations. Such information could be adopted for better recovering an albedo gradient map which is a key step for the Retinex-based method. Besides, it is considered that the albedo layer is piece-wise smooth, and the albedo gradient is reasonably sparse but often has a large value [Bla85] in contrast of the shading layer which is smooth everywhere thus has relatively small gradients. Therefore the idea of ℓ_0-smoothing [XLXJ11] that was originally introduced to create piece-wise smooth artistic effect for image processing was adopted to help recover the albedo layer. Shading layer is obtained by minimizing its gradient and Laplacian values. In this chapter, as most of the research on intrinsic image decomposition, we assume the Lambertian reflectance condition and that the environment is singly-white-colored.

The proposed method is categorized as edge-based since it is based on the gradient-prior of the albedo layer and the chromaticity gradient. It is shown by experiments that the proposed method can achieve comparable or even better performance to the other edge-based methods. The main advantage of the method is its computational efficiency without sacrificing accuracy of decomposition. Briefly, the use of chromaticity edge information helps to keep the accuracy of decomposition by reducing the influence of non-albedo gradients, while decomposition is done very fast by reducing the proposed optimization problem into the alternating minimization problem [WYYZ08] and adopting the fast Fourier transform (FFT) to accelerate the computation. The proposed method runs much faster, and our results are at least close (and in some cases better than) to those obtained by the

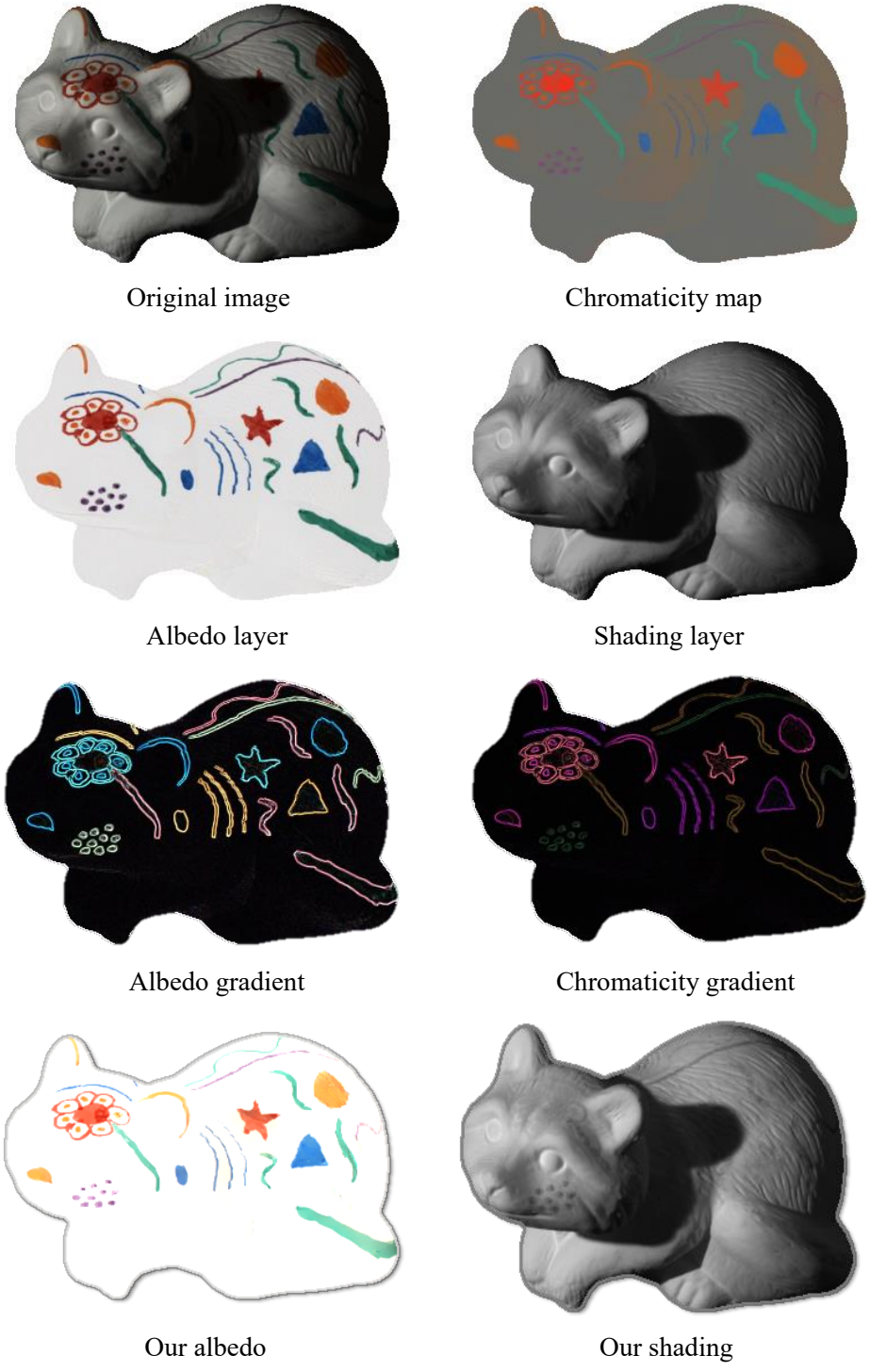


Fig. 2.1: The original input, the chromaticity map, the albedo layer and the shading layer (ground truth) and the obtained decomposition results of raccoon from the MIT dataset [GJAF09].

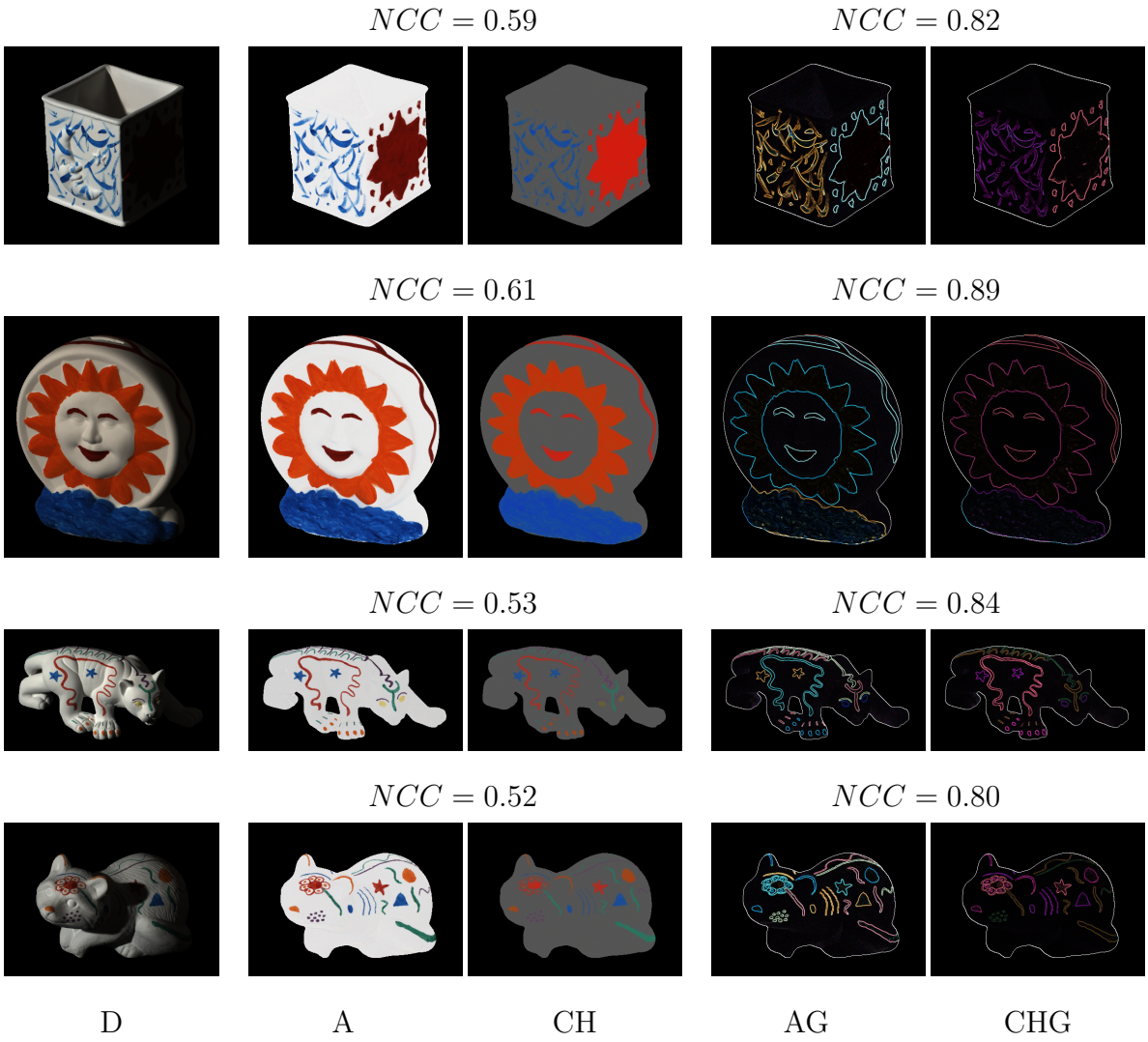


Figure 2.2: NCC between the albedo layer and the chromaticity map, and between the albedo gradient and the chromaticity gradient. It can be observed that an albedo gradient map and a chromaticity gradient map has stronger correlation than an albedo layer and a chromaticity map does. (D: diffuse, A: albedo, CH: chromaticity, AG: albedo gradient, CHG: chromaticity gradient)

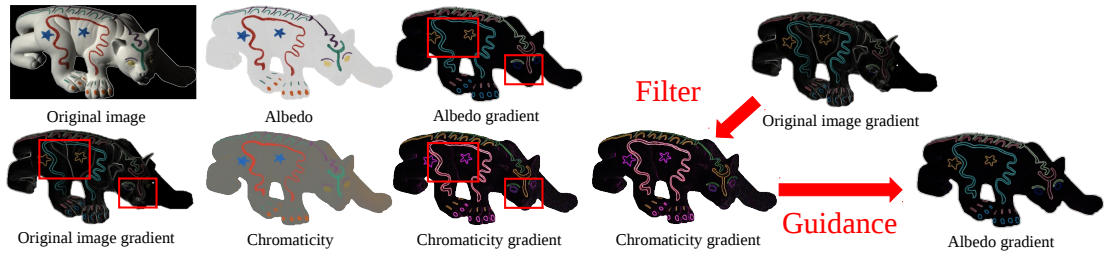


Fig. 2.3: Example of the image gradient, the albedo map, the albedo gradient, the chromaticity map and the chromaticity gradient. The Chromaticity gradient is employed to filter the image gradient, and then to recover the albedo gradient.

state-of-the-art methods that use sophisticated models and inference. Part of this work has been published in conference and can be found in [2].

To sum up, the highlights of the work described in this chapter include:

- We analyze the correlation through normalized cross correlation (NCC) between a chromaticity map and an albedo layer, and between a chromaticity edge and an albedo edge. At the same time, this correlation is explained via deducing some physics property of chromaticity and albedo.

- We firstly introduce the use of chromaticity gradients in the problem of intrinsic image decomposition and show mathematically that chromaticity gradients contain location information of albedo edges.

- We propose a fast solver based on Plancherel's theorem and half-quadratic splitting scheme [WYYZ08] to efficiently solve the proposed intrinsic image decomposition model, and keep the accuracy of decomposition while speeding up the computation. The proposed method runs much faster, and the results of the proposed method are close or better than the state-of-the-art methods that use sophisticated models and complicated optimization methods.

The remainder of the chapter is organized as follows. The related work is presented in Section 2.2. Section 2.3 overviews the proposed approach; Section 2.4 provides experimental comparisons with prior approaches. A discussion and summary conclude the paper in Section 2.5.

2.2 Related Work

2.2.1 Chromaticity

The property that a chromaticity map is free of shading influence has been widely applied by previous researchers for a wide domain of computer vision tasks. Due to its property, chromaticity is also used for high-level vision tasks such skin detection under changing illumination [SMHL00] and color object recognition [DWL98]. In this work, the utility of chromaticity map was extended, and the chromaticity edge for intrinsic image decomposition was adopted. The relationship between a chromaticity edge map and an albedo edge map was analyzed in Section 2.3.

Tan and Ikeuchi [Tan08] introduced a method to separate reflection components of textured surfaces by shifting a pixel's intensity and chromaticity nonlinearly without explicit color segmentation which becomes ineffective for complex textured images. They [TNI04b] showed that a direct correlation between illumination chromaticity and image chromaticity could be obtained in a newly introduced two-dimensional inverse-intensity chromaticity space to analyze highlights. Tan *et al*. [TNI04a] analyzed the noise that most real images suffer from to identify the diffuse maximum chromaticity correctly for robust separation of reflection components. Different from the previous work, the correlation between a chromaticity gradient and the corresponding underneath albedo edge is analyzed.

2.2.2 Intrinsic Image Decomposition

In the early stage, the problem of intrinsic image decomposition was first understood as shape-from-shading [Hor70], which addressed a subset of the intrinsic image problem in which all objects were assumed to be relatively smooth and environment light is white. In this case, shading was assumed to be known, and all materials share the same albedo. It was not until Land and McCanns's Retinex algorithm [LM71] that the first algorithm for the task of intrinsic decomposition was proposed. Horn [Hor74] later on effectively used the Retinex algorithm by analyzing more the edge information. Despite its simplicity and long history, Retinex remained as the most effective intrinsic image algorithm for a considerable length of time [GJAF09].

Following the idea underlying the Retinex algorithm, researchers [TFA05, STL08] focused on separating albedo and shading edges using different kinds of priors based on a single image. By explaining model the Retinex algorithm in a new optimization problem, Grosse *et al.* [GJAF09] later proposed a color version Retinex algorithm. Shen *et al.* [SYJL11] used local information and formulated intrinsic image decomposition as a minimization problem. These methods are referred to as edge-based methods. In Grosse *et al.* [GJAF09]'s work, they also presented a ground truth dataset (the so-called MIT dataset) for intrinsic images containing 20 real simple scenes with one object in each scene. Another dataset was released by Bell *et al.* [BBS14]. It contained a large number of indoor scene images with manually labeled information on relative reflectances between different objects or different parts of an object, via online crowdsourcing. They also proposed a dense conditional random field (CRF)-based algorithm combining several generally used priors for intrinsic decomposition. Their method outperformed the state-of-the-art methods on the proposed dataset.

Multi-images-based decomposition approaches, apart from single-image-based methods, cope with the complexity of natural scenes by exploiting multiple views or multiple lighting conditions of the same scene [Wei01] [MLKS04] [LBD13] [DRC+15] [Yu16]. Laffont *et al.* [LBD13] adopted multiple images to reconstruct the geometry of 3D scenes to better separate albedo from shading and from decomposing the shading. Yu [Yu16] proposed a multi-images-based decomposition method by using a modified version of the robust principle component analysis (RPCA) and achieved a common reflection layer for all the input images. However, multi-view data are not always available and are difficult to capture for video.

The work by Vineet *et al.* [VRT13] combined scene-shape, illumination, and reflectance from shading (SIRFS) [BM13] with semantic reasoning, and demonstrated that improved reasoning about intrinsics could improve semantic segmentation. Recent work on an efficient optimization-based approach has introduced a heterogeneous shading model [CK13, JCTL14]. Besides the input of color images, Chen and Koltun [CK13] applied the information from depth images and improved the performance of estimating albedo and shading fields. Jeon *et al.* [JCTL14] preprocessed an image by separating the structure-texture of a scene, and then

combined surface normals from depth images to obtain good decomposition results from actual complicated scenes.

The latest research branch in machine learning / deep learning techniques was introduced to yield satisfactory results on real complicated scenes [NMY15] [ZKE15] [SBD15]. Zhou *et al.* adopted a convolutional neural network (CNN) model to predict relative albedo ordering between image patches and they included this information to enable intrinsic image decomposition. Shelhamer *et al.* first learned the depth information of a scene via CNN, and then used the learned depth as input to the method in Vineet *et al.* [VRT13] to obtain beneficial decomposition results. Narihira *et al.* built a CNN model based on synthetic computer graphics ground truth with albedo and shading information, which they used for directly separating the intrinsic layers.

2.3 Our Approach

2.3.1 Imaging Formation

For Lambertian materials, a physically plausible relation for the measured color value I_k ($k \in \{R, G, B\}$) reads:

$$\begin{aligned} I_k(p) &= \int c_k(\lambda)\rho(p,\lambda)\overrightarrow{n}(p)\cdot\overrightarrow{l}\ s(\lambda)d\lambda \\ &= \overrightarrow{n}(p)\cdot\overrightarrow{l}\int c_k(\lambda)\rho(p,\lambda)s(\lambda)d\lambda \\ &= \overrightarrow{n}(p)\cdot\overrightarrow{l}\ c_k(\lambda_k)\rho(p,\lambda_k)s(\lambda_k) \\ &= R_k(p)S(p) \end{aligned} \tag{2.1}$$

where λ is wavelength, ρ is the albedo coefficient, c_k describes the camera spectral response of the k-th channel, $\overrightarrow{n}$ is a surface normal direction and $\overrightarrow{l}$ is a lighting direction, p is an image pixel, $s(\lambda)$ is an illumination spectra. R_k are the albedo layer of k-th channels and S is the shading layer, which are defined as $R_k(p) = c_k(\lambda_k)\rho(p,\lambda_k)s(\lambda_k)$ and $S(p) = \overrightarrow{n}(p)\cdot\overrightarrow{l}$. The second last equal sign is valid under the narrowband camera assumption. In the following, we use ∇ and Δ to indicate the gradient and Laplacian operator, and ∂_x and ∂_y as gradient operations through x and y directions, respectively. The same symbols are adopted also in

the following context.

2.3.2 Chromaticity Gradient

The chromaticity map (*e.g.* Fig. 2.3) is obtained by removing the intensity inhomogeneity due to the effect of illumination from the original image I [Dre03], which is defined by $CH = \frac{(I_r, I_g, I_r)}{\frac{1}{3}(I_r + I_g + I_b)}$, and

$$CH_k = \frac{3R_k S}{(R_r + R_g + R_b)S} = \frac{3R_k}{R_r + R_g + R_b} = \frac{R_k}{R_{mean}} \tag{2.2}$$

where R_{mean} is the mean of R_k for all channels. Chromaticity edges are then defined by gradient as

$$\nabla CH_k = \nabla(\frac{R_k}{R_{mean}}) = \frac{\nabla R_k R_{mean} - R_k \nabla R_{mean}}{R_{mean}^2} \tag{2.3}$$

Under the assumption that albedo is piece-wise constant, which means that albedo edges exist only in the border of two neighboring constant regions, we can see that the location of a chromaticity edge corresponds well to the location of an albedo edge (R_k and R_{mean} are piece-wise constant in the whole image). In the interior of a surface region with a uniform albedo, there is no chromaticity edge or albedo edge since R_k and R_{mean} are constant. Chromaticity edges can have a non-zero value only where the albedo of the surface changes. On the contrary, the albedo edge may exist even when there is no chromaticity edge. This happens in a special case where $\frac{\nabla R_k}{R_k} = \frac{\nabla R_{mean}}{R_{mean}}$. This can be seen from 2.3 because such condition makes $\nabla R_k R_{mean} - R_k \nabla R_{mean}$ become zero.

Further, we consider the normalized cross correlation (NCC) between the chromaticity map and the albedo layer, and between the chromaticity gradient and the albedo gradient. Note that the property of gradient consider here is magnitude instead of the orientation. If not indicated, in following the gradient also means its magnitude. Some results are listed in Figure 2.2. The ground truth of the albedo layer is provided in the dataset of [GJAF09]. The mean NCC among 20 images of the dataset is listed in Table 2.1. According to Table 2.1, It can be seen that the chromaticity gradient and the albedo gradient has a high correlation, compared to the correlation between a chromaticity map and an albedo layer. It is reasonable as we have given the analysis of the albedo gradient and the chromaticity gradient above.

Table 2.1: Quantitative comparison with previous methods

Comparison	**NCC**
Albedo vs. Chromaticity	0.56
Albedo Gradient vs. Chromaticity Gradient	0.79
Albedo Gradient vs. Masked Chromaticity Gradient	0.86

Further, compared to the magnitude of a chromaticity gradient, its non-zero location provides more information of an albedo gradient. That is because the location of a chromaticity edge corresponds to the location of an albedo edge, according to our discussion above (Equation 2.3). Therefore, to find a better descriptor of the location of the albedo gradient, the chromaticity gradient is thresholded to obtain a bi-value mask with 1 and 0 indicating the gradient magnitude larger than 90% the maximum of the chromaticity gradients of the image, or not. The NCC between the albedo gradient and the masked chromaticity gradient can be seen in Table 2.1. Here it can be seen that the non-zero location of the chromaticity gradient provides more information than its magnitude for the non-zero location of the albedo gradient.

Therefore, according to the analysis with NCC, the location mask obtained from the chromaticity gradient can serve as a strong hint to acquire the albedo edge thus can be used as the guidance map. The mask based chromaticity gradient is then defined as:

$$T_t(p) = \begin{cases} 1, & max(||\nabla CH_k(p)||) > \frac{max(\nabla Chrom)}{t} \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

where subindex k denotes channels (R, G, B), t is a parameter to adjust the threshold. T_t has non-zero magnitudes (value 1) at pixels where chromaticity gradient magnitudes have a sufficiently large value, max operator is with respect to all pixels in an image. On the other hand, in the region where there should be no albedo edge, which is indicated that no chromaticity edge exists, T_t has zero magnitudes.

2.3.3 Edge-based Decomposition Model

Intrinsic decomposition problem can be considered as a layers separation problem, whose objective is to obtain a solution maximizing the joint probability

$$p(R, S|I) = p(R|I)p(S|I), \tag{2.5}$$

where the equality is valid under the assumption that the albedo layer R and shading layer S are independent. In this work, we seek the likelihoods $p(R|I)$ and $p(S|I)$ by their statistical property on edges.

Some symbols are used and denote $\log(I)$, $\log(R)$, and $\log(S)$ as $\hat{I}$, $\hat{R}$, $\hat{S}$ respectively and write as $\hat{I} = \hat{R} + \hat{S}$ in the *log* domain. The original images are normalized to $[0, 1]$. $\hat{R}$ should fall in the range $[\hat{I}, 0]$. A previous study indicates that for a simple scene composed of simple objects or a single object, the albedo layer satisfies the piece-wise constant assumption [WO04]. Accordingly, the optimization model under the framework with ℓ_0-norm is as follow (Equation 2.6). The intuition of the proposed mathematical model for the edge-based model or the Retinex assumption can be seen from the statistics phenomenon of the two layers R and S (see Fig. 2.4). In the following, $||\cdot||$ without a subscript means sum of ℓ_2-norm for all pixels in the image, subscript p means value at a pixel, and the *sum* operator runs over the whole image.

$$\begin{aligned} &\min_{\hat{R},\hat{S}} ||\nabla\hat{R}||_0 + \lambda_1||\nabla\hat{S}|| + \lambda_2||\Delta\hat{S}|| \\ &\text{s.t.} \quad \hat{I}_p \leq \hat{R}_p \leq 0,\ max(\hat{S}) = 0, \end{aligned} \tag{2.6}$$

where ∇R, ∇S and ΔS are the gradient of R, the gradient of S and the laplacian of S, respectively. All of them have the same dimension as input image I. Here $max(\hat{S}) = 0$ is the largest value among all pixels. It comes from the white patch assumption for white-balancing and under the assumption in this work that the illumination of the scene is of white color.

Noticing that $\hat{S} = \hat{I} - \hat{R}$, Equation 2.6 is written into the pixel-wise form as:

$$\begin{aligned} &\min_{\hat{R}} \Big\{ C(\nabla\hat{R}) + \sum_p (\lambda_1(\nabla\hat{I}_p - \nabla\hat{R}_p)^2 + \lambda_2(\Delta\hat{I}_p - \Delta\hat{R}_p)^2) \Big\} \\ &\text{s.t.} \quad \hat{I}_i \leq \hat{R}_i \leq 0. \end{aligned} \tag{2.7}$$

where $C(I) = \#\{p||\partial_x I_p| + |\partial_y I_p| \neq 0\}$, $\#$ is the symbols indicating the number of elements in a set.

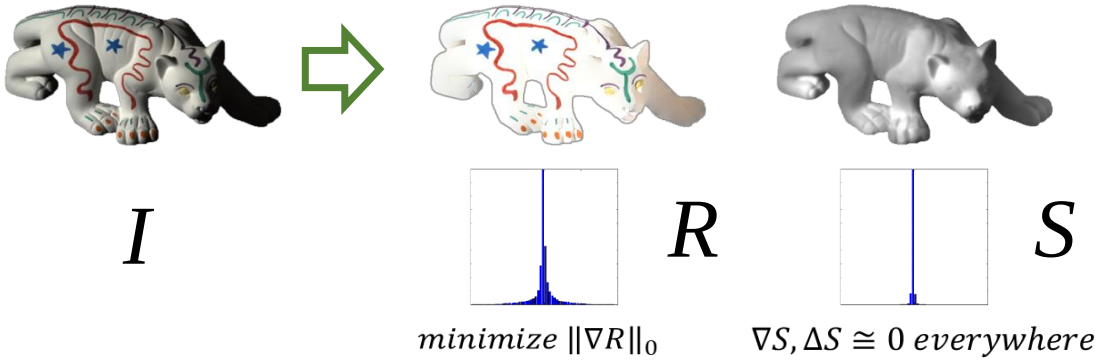


Fig. 2.4: Our mathematical model for the Retinex assumption.

In the following subsection, we will solve the optimization problem (Equation 2.7) by the alternating strategy and function T_λ will be included in each iteration to guide recovering the gradient of the albedo layer $\hat{R}$.

2.3.4 Optimization

The objective function we defined is non-convex because of $C(I)$. The half-quadratic separation scheme [WYYZ08, XLXJ11] is used to solve this problem.

As in [XLXJ11], we introduce auxiliary variables h_p^x and h_p^y to make the problem solvable and yield the objective function

$$\begin{aligned}\min_{\hat{R},(h^x,h^y)} \Big\{ & C(h^x,h^y) + \sum_p (\beta((\partial_x \hat{R}_p - h_p^x)^2 + (\partial_y \hat{R}_p - h_p^y)^2) \\ & + \lambda_1((\partial_x \hat{I}_p - \partial_x \hat{R}_p)^2 + (\partial_y \hat{I}_p - \partial_y \hat{R}_p)^2) + \lambda_2(\Delta \hat{I}_p - \Delta \hat{R}_p)^2) \Big\}\end{aligned} \tag{2.8}$$

where $C(h^x,h^y) = \#\{p||h_p^x| + |h_p^y| \neq 0\}$, β is a weight that is increased during the optimization (in our implementation, starting from around 10 and multiplied by η each time, and η is used to accelerate the iteration and set to be between 1 and 2.). h_p^x and h_p^y has the same dimensions as the input image I. Minimizing Equation 2.8 for a fixed β can be performed by alternating between computing $\hat{R}$ and (h^x,h^y) as following.

Computing (h^x,h^y) Keeping $\hat{R}$ fixed, similar as in [SJA08] and [XLXJ11], the following problem is minimized:

$$\min_{(h^x,h^y)} \Big\{ C(h^x,h^y) + \sum_p (\beta((\partial_x \hat{R}_p - h_p^x)^2 + (\partial_y \hat{R}_p - h_p^y)^2) \Big\} \tag{2.9}$$

As it is easy to find the two local minimum points of this equation, which are $(\partial_x \hat{R}_p, \partial_y \hat{R}_p)$ and $(0,0)$. When $(h^x, h^y) = (\partial_x \hat{R}_p, \partial_y \hat{R}_p)$ at pixel p, the function value is 1 for the individual term of second half of Equation 2.9; whereas when $(h^x, h^y) = (0,0)$ the function value is $\beta((\partial_x \hat{R}_p)^2 + (\partial_x \hat{R}_p)^2)$. Thus when $1 < \beta((\partial_x \hat{R}_p)^2 + (\partial_x \hat{R}_p)^2)$, namely $(\partial_x \hat{R}_p)^2 + (\partial_x \hat{R}_p)^2 > \frac{1}{\beta}$, the global minimum is at $(h^x, h^y) = (\partial_x \hat{R}_p, \partial_y \hat{R}_p)$. Otherwise, the global minimum is at $(h^x, h^y) = (0,0)$. Therefore, the closed-form solution for Equation 2.9 is:

$$(h^x, h^y) = \begin{cases} (\partial_x \hat{R}_p, \partial_y \hat{R}_p), & (\partial_x \hat{R}_p)^2 + (\partial_x \hat{R}_p)^2 > \frac{1}{\beta} \\ (0,0), & \text{otherwise.} \end{cases} \tag{2.10}$$

Computing $\hat{R}$ Fixing (h^x, h^y), the subproblem of Equation 2.8 w.r.t.$\hat{R}$ is to be solved. At the same time, considering the relation between the chromaticity map and the albedo layer discussed in Sec. 2.1, to avoid the influence of the gradients caused by the shading as much as possible, we use the information provided by the chromaticity map as a filter. Therefore we include the chromaticity gradient map $Chrom$ as the a guider using T_λ in Equation 2.4. Then the $\hat{R}$ estimation subproblem corresponds to

$$\min_{\hat{R}} \left\{ E_{\hat{R}} \right\} \tag{2.11}$$

$$\begin{aligned} E_{\hat{R}} &= \textstyle\sum_p (\beta((\partial_x \hat{R}_p - T_\beta(p) h_p^x)^2 + (\partial_y \hat{R}_p - T_\beta(p) h_p^y)^2) \\ &+ \lambda_1((T_\beta(p)\partial_x \hat{I}_p - \partial_x \hat{R}_p)^2 + (T_\beta(p)\partial_y \hat{I}_p - \partial_y \hat{R}_p))^2 + \lambda_2(\Delta \hat{I}_p - \Delta \hat{R}_p)^2) \end{aligned} \tag{2.12}$$

$$\begin{aligned} E_{\mathcal{F}(\hat{R})} &= \textstyle\sum_p (\beta((\mathcal{F}(\partial_x \hat{R}_p) - \mathcal{F}(T_\beta(p) h_p^x))^2 + (\mathcal{F}(\partial_y \hat{R}_p) - \mathcal{F}(T_\beta(p) h_p^y))^2) \\ &+ \lambda_1((\mathcal{F}(T_\beta(p)\partial_x \hat{I}_p) - \mathcal{F}(\partial_x \hat{R}_p))^2 + (\mathcal{F}(T_\beta(p)\partial_y \hat{I}_p) - \mathcal{F}(\partial_y \hat{R}_p)))^2 \\ &+ \lambda_2(\mathcal{F}(\Delta \hat{I}_p) - \mathcal{F}(\Delta \hat{R}_p))^2) \end{aligned} \tag{2.13}$$

If we consider ∇ and Δ as convolution operations (since the ∂_x, ∂_y and Δ can be considered as matrices $[-1\ 1]$, $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ that are operated on input image I), from the property of Fourier transform, we have $\mathcal{F}(\partial_x \hat{R}) = \mathcal{F}(\partial_x)\mathcal{F}(\hat{R})$, $\mathcal{F}(\partial_y \hat{R}) = \mathcal{F}(\partial_y)\mathcal{F}(\hat{R})$, $\mathcal{F}(\Delta \hat{R}) = \mathcal{F}(\Delta)\mathcal{F}(\hat{R})$ and $\mathcal{F}(\Delta \hat{I}) = \mathcal{F}(\Delta)\mathcal{F}(\hat{I})$ here the multiplication is pixel-wise multiplication.

$$\begin{aligned} E_{\mathcal{F}(\hat{R})} &= \textstyle\sum_p (\beta((\mathcal{F}(\partial_x)\mathcal{F}(\hat{R}_p) - \mathcal{F}(T_\beta(p) h_p^x))^2 + (\mathcal{F}(\partial_y)\mathcal{F}(\hat{R}_p) - \mathcal{F}(T_\beta(p) h_p^y))^2) \\ &+ \lambda_1((\mathcal{F}(T_\beta(p)\partial_x \hat{I}_p) - \mathcal{F}(\partial_x)\mathcal{F}(\hat{R}_p))^2 + (\mathcal{F}(T_\beta(p)\partial_y \hat{I}_p) - \mathcal{F}(\partial_y)\mathcal{F}(\hat{R}_p)))^2 \\ &+ \lambda_2(\mathcal{F}(\Delta)\mathcal{F}(\hat{I}_p) - \mathcal{F}(\Delta)\mathcal{F}(\hat{R}_p))^2) \end{aligned} \tag{2.14}$$

The 2D Fourier transform is applied on the analytical solution of Equation 2.12 to obtain Equation 2.13 according to Plancherel's theorem [Bra00]. In fact, Plancherel's theorem states that the quadratic norm of a function equals the quadratic norm of its Fourier transform. For all values of $\hat{R}$, the energy equivalence $E_{\hat{R}} = E_{\mathcal{F}(\hat{R})}$ is valid. It further follows that the optimal values of variable $\hat{R}^*$ that minimize $\hat{R}$ correspond to minimizing $\mathcal{F}(\hat{R})$ in the frequency domain. Thus, optimal R^* can be obtained by:

$$R^* = \mathcal{F}^{-1}(\underset{\mathcal{F}(R)}{\operatorname{argmin}}\{E'_{\mathcal{F}(R)}\}). \tag{2.15}$$

As $E'_{\mathcal{F}(R)}$ is a quadratic sum of unknown $\mathcal{F}(R)$, it is a convex problem and can be solved by setting $\frac{\partial E'_{\mathcal{F}(R)}}{\partial \mathcal{F}(R)}$ to zero. By further calculation, $\hat{R}$ can be found by

$$\begin{aligned}
\hat{R} &= \mathcal{F}^{-1}(A/B), \\
A &= \beta(\overline{\mathcal{F}(\partial_x)}\mathcal{F}(T_\beta \circ h^x) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(T_\beta \circ h^y)) \\
&\quad +\lambda_1(\overline{\mathcal{F}(\partial_x)}\mathcal{F}(T_\beta \circ \partial_x \hat{I}) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(T_\beta \circ \partial_y \hat{I})) \\
&\quad +\lambda_2\overline{\mathcal{F}(\Delta)}\mathcal{F}(\Delta)\mathcal{F}(\hat{I}), \\
B &= (\beta + \lambda_1)(\overline{\mathcal{F}(\partial_x)}\mathcal{F}(\partial_x) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(\partial_y)) \\
&\quad +\lambda_2\overline{\mathcal{F}(\Delta)}\mathcal{F}(\Delta) + \tau,
\end{aligned} \tag{2.16}$$

where $\mathcal{F}$ is the Fourier transform, $\circ$ is the pixel-wise product and / is pixel-wise division (p is omitted without losing the generality), $\overline{\mathcal{F}(\cdot)}$ is the complex conjugate (however, we use the Plancherel's theorem in the case of real number), T_β is defined as in Equation 2.12 where β is the same as in Equation 2.8, the parameter τ added to the denominator is a small number necessary for the stability of division. Both the multiplication and division are pixel-wise operation.

The whole process is sketched in Algorithm 3.

2.4 Experiments

Experiments were carried to test the convergence behavior and the optimization speed of the proposed algorithm. Comparisons with the state-of-the-art methods were also done to demonstrate the effectiveness of the proposed method. Evaluations were done on the MIT dataset [GJAF09] and the dataset provided by Bousseau [BPD09].

Algorithm 1 Chromaticity Edge-based Intrinsic Decomposition

Input: input image I; smoothness weight λ_1, λ_2; initial β_0; iterations number $i_{\max}$; increasing rate η;

Initialization: $\hat{R} \leftarrow I$; $\beta \leftarrow \beta_0$; $i \leftarrow 0$.

while $i < i_{\max}$ **do**

fixing $\hat{R}$, computing (h^x, h^y) using Equation 2.10;

fixing (h^x, h^y), compute $\hat{R}$ using Equation 2.16;

$\beta = \eta * \beta$;

$i++$;

end while

Output: $\hat{R}$ and $\hat{S} = \hat{I} - \hat{R}$;

The experiments were done on a PC with Intel i7-3612M CPU (quad-core 2.1 Hz) and 8GB RAM. The implementation was done using C++. The proposed algorithm have been tested on the MIT intrinsic image dataset [GJAF09] and some images provided by Bousseau *et al.* [BPD09]. In most of the experiments in this work, it was found that λ_1 is better to be small than $0.01\lambda_2$. In this chapter, we fixed λ_1 to be $0.005\lambda_2$ for all the given results.

2.4.1 Behaviour of convergence

From Equation 2.16, It can be seen that when λ_2 (thus λ_1) tends to ∞, A/B tends to $\mathcal{F}(I)$ (as T_β becomes 1 everywhere after some iterations) thus $\hat{R}$ equals I after some iterations. In this case the (h^x, h^y) have effect. Otherwise, when λ_1 tends to 0, A/B approaches to

$$\frac{\overline{\mathcal{F}(\partial_x)}\mathcal{F}(T_\beta \circ h^x) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(T_\beta \circ h^y))}{\overline{\mathcal{F}(\partial_x)}\mathcal{F}(\partial_x) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(\partial_y) + \tau/\beta}. \tag{2.17}$$

In this case the behaviour of the proposed algorithm is almost independent of the input image $\hat{I}$. Actually, for λ_1 decreasing from ∞ to 0, the details of intermediate result of iterations move gradually from R to S. The proposed method distributes the gradient and Laplacian values between the two layers. R obtains the significant values, whereas S obtains the relatively smoother details. The effect of different λ_1 setting is shown in Fig. 2.5. It can be qualitatively seen how λ_1 behaves to distribute the gradients between R and S. In most of the experiments, setting

λ_1 equal to 0.01 can help the proposed method to obtain good enough results. However, if the shading variation or the shading effect is large, small λ_1 is needed.

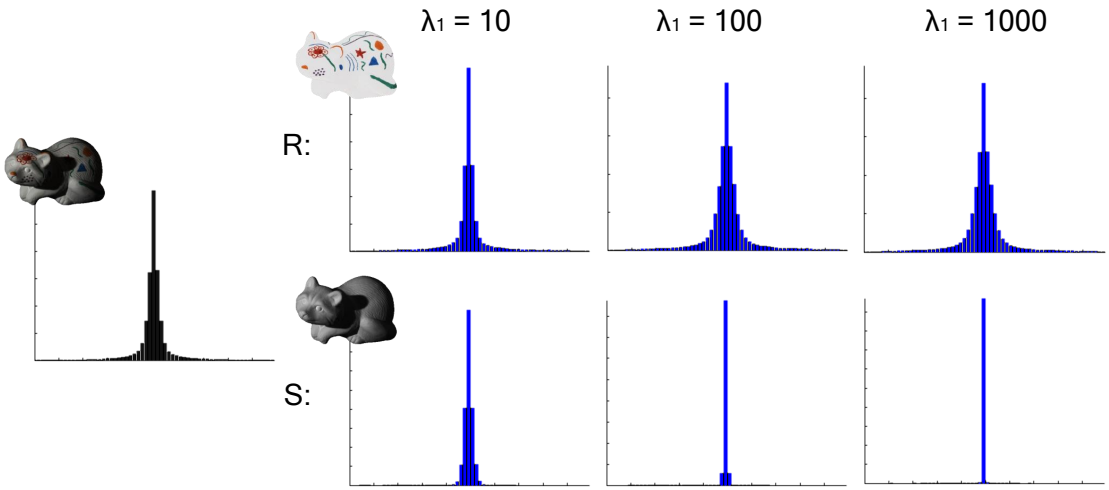


Fig. 2.5: This figure shows the effect of different λ_1 setting on the final composition results on the example of raccoon. (x-axe represents the value of gradient and y-axe represents the number of pixels for each gradient value.)

2.4.2 Fast Optimization

Experiments showed that the proposed optimization framework converged fast to a good solution, normally within 5 iterations. In fact, for each iteration, the proposed method can return results quickly. Although the obtained results are local-optima due to the non-convexity of the proposed model, most of the cases, the results of the proposed method are of good quality and usable for further processing task such as re-texturing, changing shading, *etc.*

For the calculation of FFT and IFFT, the FFTW[1] library was adopted which can give more than twice higher computational performance than FFT of OpenCV. According to Equation 2.16, it can be known that $\overline{\mathcal{F}(\partial_x)}\mathcal{F}(\partial_x)$, $\overline{\mathcal{F}(\partial_y)}\mathcal{F}(\partial_y)$ and $\overline{\mathcal{F}(\Delta)}\mathcal{F}(\Delta)$ can be calculated at initialization and be used for each iteration. Therefore for each iteration when computing $\hat{R}$, 9 FFT and 1 IFFT need to be

[1]http://www.fftw.org/

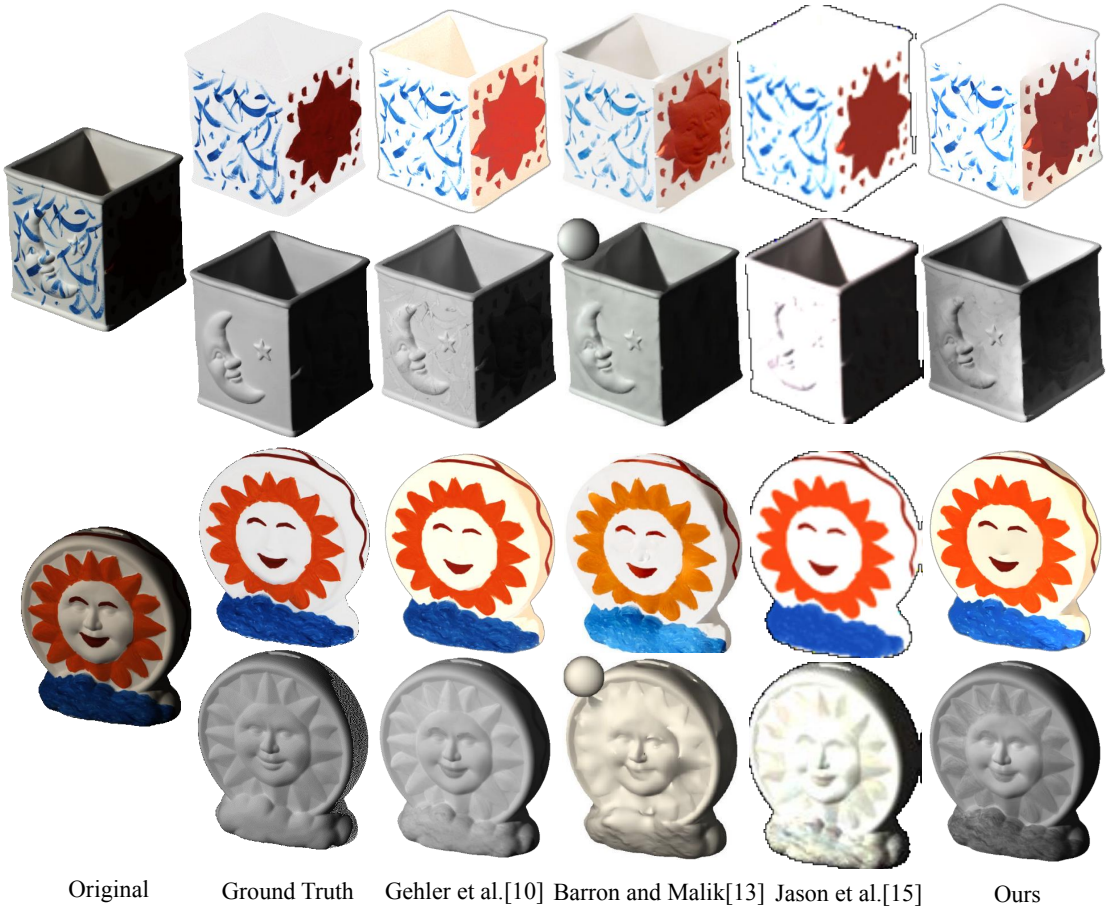


Figure 2.6: This figure shows the decomposition results by Gehler *et al.*'s [GRK+11], Barron & Malik's. [BM12], Chang *et al.*'s [CCFI14] and the proposed method on two images from the MIT intrinsic dataset. Ground Truth is shown for reference.

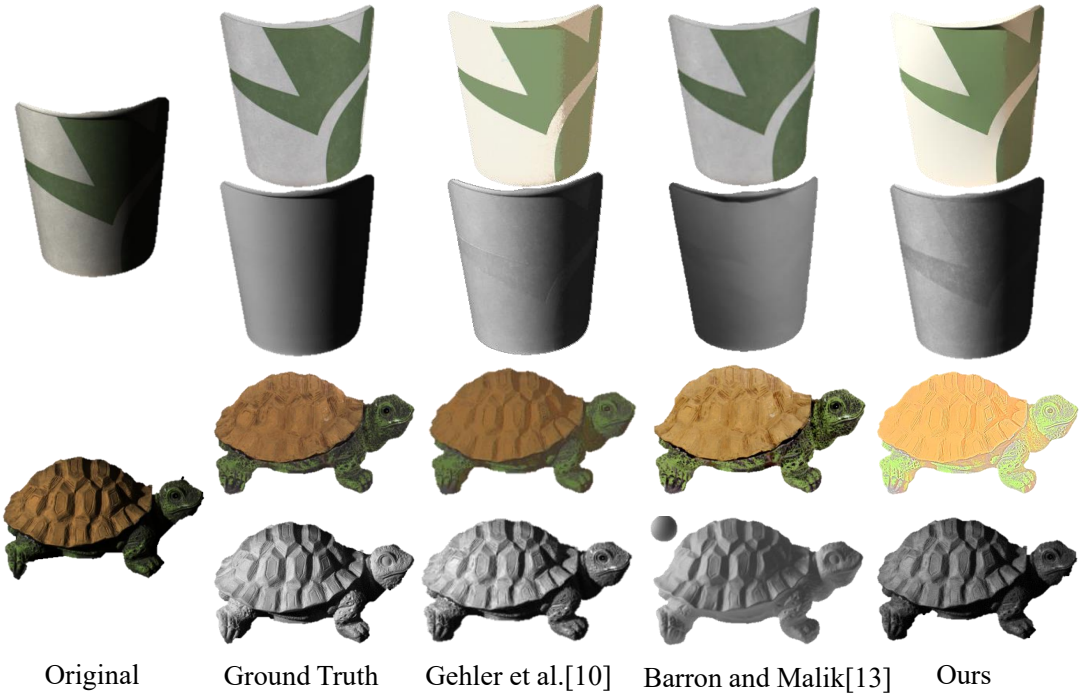


Figure 2.7: This figure shows the decomposition results by Gehler *et al.*'s [GRK+11], Barron & Malik's. [BM12], and the proposed method on two images from the MIT intrinsic dataset. Ground Truth is shown for reference.

calculated, which means low computation is needed for each iteration. Together with a small number of iterations, the proposed method can achieve low computational time and is much faster compared to the state-of-the-art methods for intrinsic decomposition.

2.4.3 Comparisons

The performance of the proposed method was compared with several representative intrinsic image estimation methods and reported the runtime per image as well as two quantitative measurements: the local mean squared error (LMSE) [GJAF09]. LMSE is the local mean-squared error. It is the average of the scale-invariant MSE errors but computed on overlapping square windows of size 10% of the image along its larger dimension.

The four methods in [SY11, GRK+11, BM15, CCFI14] are considered as state-of-the-art methods depending on their performance on the dataset of [GJAF09]. Qualitative and quantitative results of these methods are either generated by the

source code provided the authors or taken from the corresponding published papers. The quantitative comparison results are listed in Table 4.2. The proposed

Table 2.2: Quantitative comparison with state-of-the-art methods

Method	**Runtime**	**LMSE**
Shen & Yeo [SY11]	>5 m	0.0149
Gehler *et al.* [GRK+11]	>9 m	0.0131
Barron & Malik [BM15]	>3 m	0.0125
Chang *et al.* [CCFI14]	—	0.0111
Proposed (unique para.)	∼0.1s	0.0141
Proposed (fine tuned para.)	∼0.1s	0.0132

method can achieve decomposition performance close to methods employing the complexity models and learning-based techniques (e.g. [GRK+11, BM15]), and runs much faster. Note that two LMSE results of the proposed method were given, the lower one (0.0147) was obtained with the unique parameters that is mentioned above for all the images from MIT dataset; whereas the higher one (0.0132) was achieved by fine tuning the parameters for each image from MIT dataset. However, the fine tuning process is not large brute force workload but can be tuned by observing the behaviour of convergence that was mentioned above (Figure 2.5). There is a lack of timing report for [CCFI14] due to the lack of available source code and that the runtime was not reported in the paper. However, due to the inference and the design of algorithm, the method of Chang *et al.* [CCFI14] could have large computational complexity and runtime could be significantly larger than the proposed method.

Two examples of results (box and sun from MIT dataset) with the comparison with [GRK+11, BM12, CCFI14] are shown in Fig. 2.6. The proposed method gives visually comparable results with the others, which illustrates its effectiveness. The obtained albedo and shading layer also look close to the ground truth. Another two examples of results (turtle and cup1 from MIT dataset) with the comparison with [GRK+11, BM12] are shown in Fig. 2.7. Barron *et al.*, as well as Chang *et al.*'s results, were copied directly from their paper. According to Fig. 2.6 and Fig. 2.7, the proposed method gives visually comparable results with the others,

which illustrates its effectiveness. The obtained albedo and shading layer also look close to the ground truth.

The proposed method was also tested on the input images from the dataset provided by Bousseau *et al.* [BPD09] and compared our result with it and also Tappen *et al.*'s method [TFA05]. Bousseau *et al.*'s dataset contains images of artificially made scenes without the ground truth of albedo and shading layers. It is normally used for comparisons between intrinsic image decomposition methods with user indications. Bousseau *et al.*'s method in [BPD09] is a user-assisted one that can generate more piece-wise constant albedo with user's labeling of regions sharing same albedo or same shading. Tappen *et al.*'s method [TFA05] is an edge-based method that learns a classifier to distinguish albedo edges and shading edges. As seen in Fig. 2.8, the proposed method shows arguably better albedo and shading decomposition results, considering the piece-wise smooth albedo, clear edges corresponding to the chromaticity map.

2.5 Discussion

A chromaticity map is considered to be an important clue to obtain an albedo layer in intrinsic image decomposition. In this work, the usefulness of chromaticity gradients was demonstrated, which indicate the existence of albedo edges. A novel edge-based intrinsic image decomposition method was proposed by employing the chromaticity map and by adopting the Retinex assumption under the ℓ_0-norm based optimization model. A fast solver by using FFT and IFFT was designed. The proposed method runs much faster than the state-of-the-art methods with comparable performance.

One limitation of the use of chromaticity gradients is that there is ambiguity between strong variations of reflectance (*i.e.*. monochromatic variation between black, gray and white) and variations caused by shadow. As it can be seen in Figure 2.9. In the red box, the color variation is due to the difference of reflectance, which cannot be indicated by a chromaticity map. As it has been in Section 2.3, this happens because a chromaticity map is obtained by dividing R, B, G channels by the mean of the three channels. The mean operation itself brings the ambi-

guity between variations of reflectance and variations caused by shadow. In the equation 2.4, T_λ is defined by tuning parameter which renders it dynamic change during the optimization. The strategy in this work handle in some terms the limitation case that is mentioned here. Finding an extra clue to distinguish the variation of reflectance and the variation due to shadow is our future work.

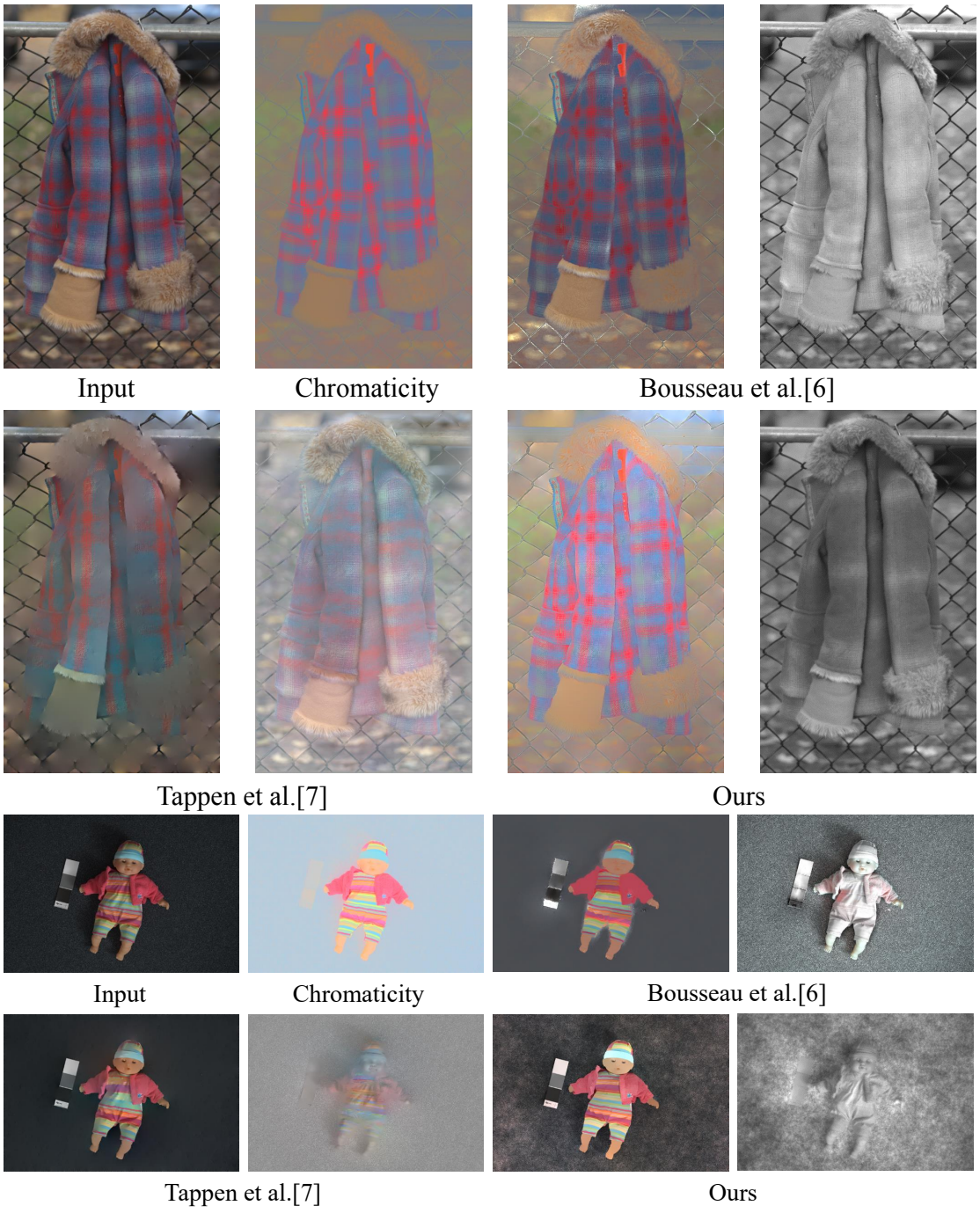


Fig. 2.8: Comparison of decomposition results on a photo with the user-assisted approach [BPD09] and Tappen et al.'s [TFA05].

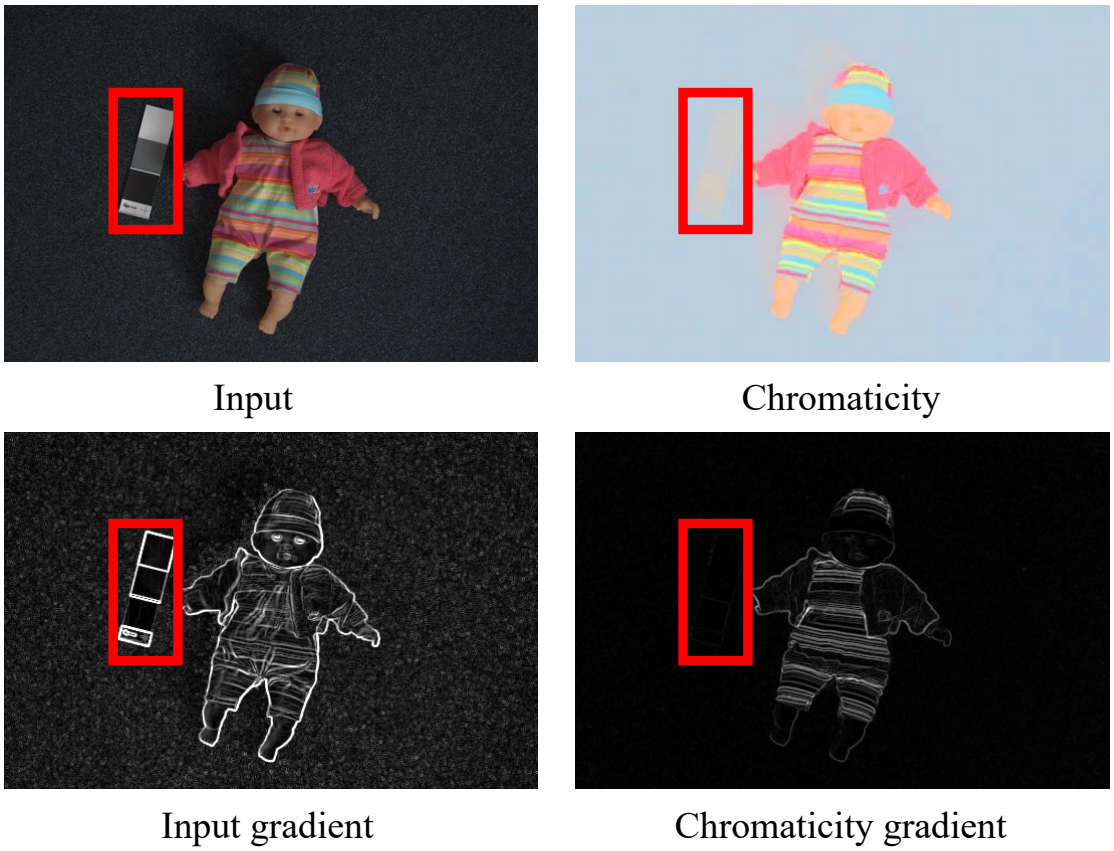


Fig. 2.9: Limitation of the chromaticity map. In red box, the variation of color is due to difference of reflectance, which cannot be indicated by the chromaticity map.

Chapter 3

Structure-Preserving Image Smoothing via Feature Asymmetry-aware WLS

3.1 Introduction

Natural scenes and human-created art pieces typically contain rich textures as can be seen on the handkerchief with a cartoon figure that is shown in Figure 3.1. Some natural scenes contain even more complicated patterns of texture (see Figures 3.6 and 3.10). While the human perception system can easily distinguish structures and textures inside images, understanding and then separating structures and textures poses a great challenge. For example, it is difficult to discriminate fine-scale edges and details. Both of them appear as small variations even though the details often appear as quasi-periodic.

The feature asymmetry (FA) has been found to be a very effective feature detector for natural image processing [Kov97, Bel16] in the frequency domain, as well as for the medical image processing with complicated noise patterns such as those with speckle noise in ultrasound images [MPN98, BBML11]. Moreover, FA is invariant to changes in image brightness or contrast due to analyzing the phase information in the frequency domain. This makes the FA an effective indicator of local image features such as edges and corners. FA has been proved to be useful for detecting non-periodic patterns from all kinds of strong edges ig-

Input Structure Texture

Figure 3.1: Example of structure-texture separation result for handkerchief with cartoon figure.

noring largely local periodic variations, which is a strong property to distinguish structures from textures within images. Consequently, presenting an optimization framework in which structures and textures are penalized differently by defining an FA based regularization term. Textures with large penalties are smoothed out whereas structures are well preserved due to the small penalties imposed on them, especially for low contrast features. The experimental results in this work revealed FA could accurately extract structure features that were observed by the human eyes.

The weighted least square (WLS) framework has been found to posses the nice property of being able to smooth image details on different scales without blurring the edges. However, it was found that the original gradient-based WLS was not suitable for texture removal, while the proposed method based on FA turned out to be a better alternative. This is because FA can better represent the human visual system's response to contours and can detect periodic patterns since it works in the frequency domain [MRBO86, MO87, PP11]. A local phase-based measure based on this observation was employed to extract the structure map from images. The map was then incorporated into the WLS framework as a weighting function to guide the optimization during smoothing. Further, the WLS framework has recently been widely accelerated by researchers [MCL+14], who have transformed the 2-dimension (2D) optimization into recursive 1-dimension (1D) optimization problems. Such 1D recursive iteration scheme guarantees the convergence of optimization while preventing the "streaking artifacts" that often occur with 1D

separable algorithms [GO11]. Since the proposed FA-based framework is modeled within the framework of WLS, we could similarly apply such techniques to the FA-based framework and greatly accelerate its solving process.

Experiments were carried out to compare the proposed method with some state-of-the-art structure-texture separation methods. The results obtained from the experiments revealed that the proposed method achieved better or comparable performance and had low computational complexity, compared to the state-of-the-art approaches. As can be seen from Figures 3.1 and 3.10, the proposed model could effectively eliminate texture without distorting structures.

To sum up, the three main highlights of the proposed work reported in this chapter are the followings:

- FA was employed to structure-texture separation problem. FA is a phase-based quantity that extracts structure edge information in the frequency domain. It has been proved through experiments that FA can serve as a better structure edge descriptor than image gradients.

- The WLS framework was employed to solve structure-texture separation. The weight of WLS was calculated based on FA. The proposed FA-aware WLS method was effective in structure-texture separation and could achieve state-of-the-art results.

- When solving the WLS framework, the 2D optimization problem was transformed to solving two problems with 1D recursive iterations by using the existing acceleration framework. This leads to significant speed up. Moreover, the proposed method ran much faster than many state-of-the-art methods.

Part of this work has been published in a conference paper and can be found in Yu and Sato [1].

This chapter is organized as follows. Some related work will be reviewed in Section 3.2. Details of the proposed method are then presented in Section 3.3 and experimental results on synthetic and clinical ultrasound images are discussed in Section 3.4.

3.2 Related Work

Some researchers made great efforts in tackling this challenging problem of structure-texture separation [Mey01, SSD09, FFL10, BLMV10, XYXJ12, KEE13]. Structure-texture separation can be formulated as an estimation problem in which a given image is separated into two components that correspond to coarse and fine scale image details. The Gaussian filter is the earliest and the most commonly used isotropic smoothing operator [Wit84, BA87]. The smoothed image was taken as the structure layer whereas the difference between the input image and smoothed image was taken to be the texture layer. Edge-aware smoothing approaches such as uses of the anisotropic diffusion filter [PM90], the total variation model [ROF92], the bilateral filter [TM98, DD02], the non-local (NL) means filter [BCM05], the WLS filter [FFLS08] and ℓ_0 smoothing [XLXJ11] have utilized differences in intensity or color values or gradient magnitudes for predicting the existence of edges, and then used the edge information to guide the smoothing process. Such intensity variations or gradient-based definitions of edges, however, might fail to capture high-frequency or periodic patterns that are related to fine image details or textures. Therefore, these approaches cannot fully separate textured regions from the main structures as the edge indicators will consider such textures as being part of the structure to be retained due to their large gradient magnitudes.

Subr *et al.* [SSD09] framed the separation problem regarding local extrema modulation based on the fact that edges are determined by intensity oscillations between local extrema. Subsequently, Xu *et al.* [XYXJ12] proposed a relative total variation descriptor to better classify structure and texture elements, and they then proposed including this information into the total variation framework to obtain better separation results. Karacan *et al.* [KEE13] adopted the region covariances to the NL-means filter and used it for image smoothing. By using region covariances commonly used for representing textures, their method was able to remove small-scale textures from images while preserving structures. Zang *et al.* [ZHZ14] used local extrema for feature characterization as Subr *et al.* [SSD09] had done, but introduced curvalization techniques to represent the 2D regions' properties into 1D curves. They reduced the 2D computation into 1D processing to achieve faster processing. Min *et al.* [MCL+14] proposed the 1D separable

implementation of the WLS method to obtain an efficient global filter with a comparable runtime to fast edge-preserving filters.

This chapter presents a novel approach for structure-texture separation based on the WLS framework [FFLS08] with the weight calculated from FA. The computation is accelerated by using the 1D separable implementation in [MCL+14]. Our method is fast and performs well on separating structures and textures.

3.3 The proposed method

3.3.1 Edge Detection Measure

A previous study has found that phase information provides evidence of an object's contours [PP11]. In fact, the local energy model developed by Morrone *et al.* [MRBO86] and Morrone and Owens [MO87] postulates that features are perceived at points where the Fourier components are maximally in phase (see Figure. 3.2), which comes from the concept of the phase congruency (PC) model.

The phase congruency model is a model of visual processing built on frequency. It supposes that the visual system can perform calculations using amplitude and phase of frequency components from a singal instead of processing visual data spatially. We assume a 1D signal $f(x)$ that is reconstructed from its Fourier transform by

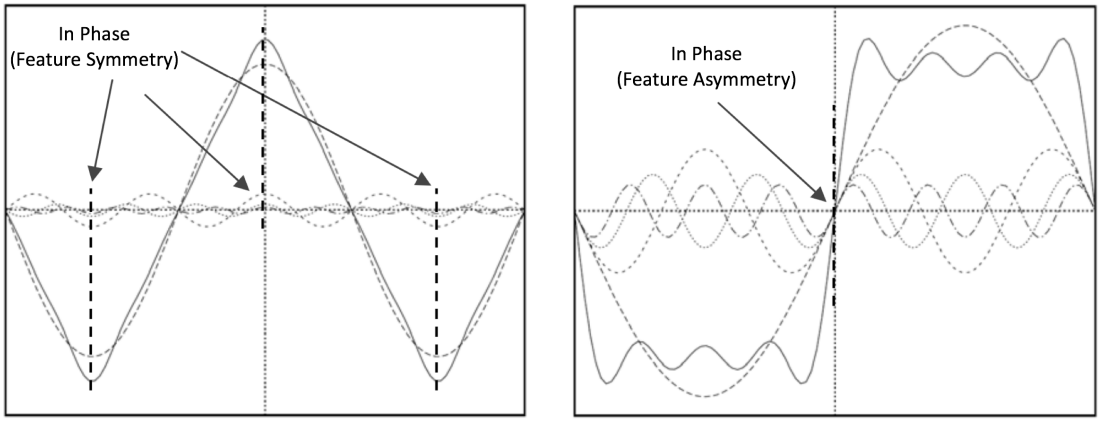
$$f(x) = \int_{-\infty}^{+\infty} a_\omega cos(\omega x + \phi_\omega) d\omega, \tag{3.1}$$

where, for each frequency ω, a_ω is the amplitude of the cosine wave and $\omega x + \phi_\omega$ is the phase offset of that wave. Correspondingly, The phase congruency function at each point x in the signal is defined as

$$PC(x) = \max_{\theta \in [0,2\pi)} \frac{\int a_\omega cos(\omega x + \phi_\omega - \theta)}{\int a_\omega d\omega}. \tag{3.2}$$

According to this definition, PC can be considered as a weighted average of local amplitude with weights obtained from phase information (calculated as $cos(\omega x + \phi_\omega - \theta)$). Examples of points in phase can be seen in Figure 3.2. Based on $f(x)$ (Equation 3.1), we define

$$h(x) = -\int_{-\infty}^{+\infty} a_\omega sin(\omega x + \phi_\omega) d\omega. \tag{3.3}$$



(Square wave) (Triangular wave)

Figure 3.2: Points in phase, examples of square wave and triangular wave.

To extract the local amplitude and local phase of the 1D signal $f(x)$, we need to represent it in its analytic form f_a as

$$f_a(x) = f(x) - ih(x), \tag{3.4}$$

where $i = \sqrt{-1}$. The local amplitude is defined as

$$E(x) = \sqrt{f(x)^2 + h(x)^2}. \tag{3.5}$$

The local phase

$$Arg(x) = arctan(h(x)/f(x)) \tag{3.6}$$

represents the angle or the orientation at which the phase congruency occurs. It can be used to define the feature type. To understand this definition and its relationship with phase congruency, one can think of a signal $f(x)$ at any point x as being made up of the sum of various sine waves at different amplitudes and phase angles. (see the vector map in Figure. 3.3). Morrone [MO87] proved that local peaks in the local amplitude correspond to local peaks in the phase congruency. Therefore, searching for local maxima in the phase congruency function is equivalent to searching for local maxima in the local amplitude function. These local maxima normally appear at lines, step edges and bar edges, and other types of features such as corners and curves (2D features).

The human visual system has the capacity to simulate convolution by odd and even symmetric filters in quadrature. The definition that filters are in quadrature

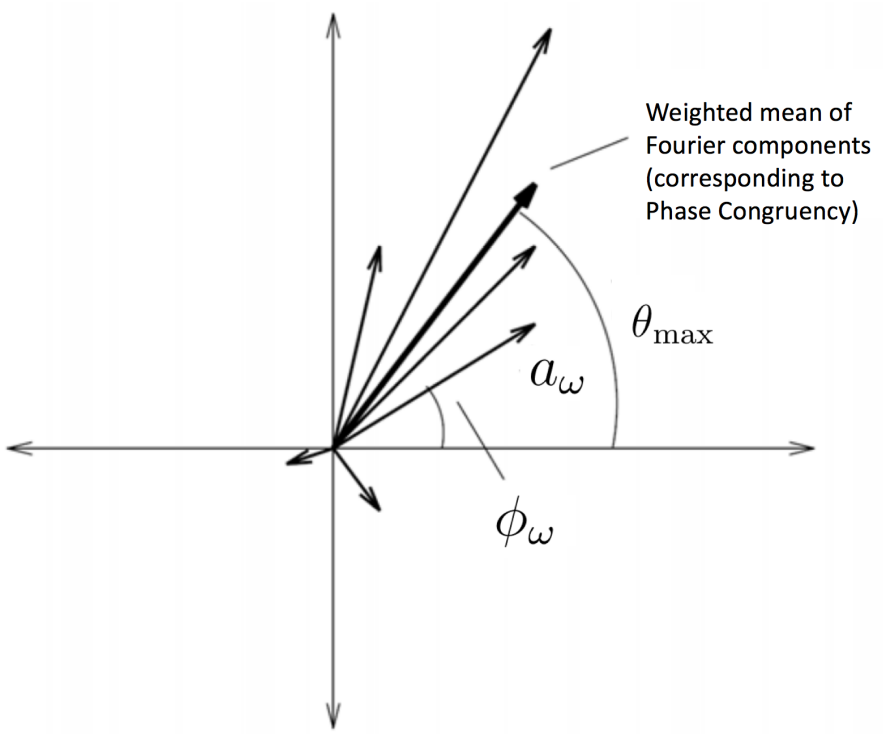


Figure 3.3: Signal $f(x)$ at any point x as being made up of the sum of various sine waves at different amplitudes and phase angles.

is that not only that the filters form an odd and even symmetric pair (which means the output of convolution by one filter is a 90° phase shift of the output of the other), but also that both of them have the same quadratic sum value and a zero mean. More in details, if g_e represents the even symmetric filter and g_o represents the odd symmetric filter, then

$$\int g_e(x)dx = \int g_o(x)dx = 0, \tag{3.7}$$

$$\int g_e^2(x)dx = \int g_o^2(x)dx. \tag{3.8}$$

Thus, the human visual system is capable of computing the local energy of a general 1D singal $f(x)$ by computing the sum of squares of the output from convolution with the odd and even symmetric filters, which has the similar mechanism as combining the output of the simple and complex cell responses in human eyes:

$$E = \sqrt{(g_e(x) * f(x))^2 + (g_o(x) * f(x))^2}. \tag{3.9}$$

The even symmetric filter g_e is chosen so that it covers as much of the frequency spectrum as possible, and at the same time eliminating the D.C. term. The odd symmetric filter g_o is then a 90° phase shift filter of g_e.

For more details about the local phase, the case that points are in phase can be one of the two occasions: feature symmetry and feature asymmetry (FA), according to Kovesi [Kov97]. As can be seen with the phases of a step function in Figure 3.4, blue dotted lines indicate the location of feature symmetry and red dotted lines indicate the location of feature asymmetry. The plot in bottom illustrate the phases of step edge function in different frequencies. Compared to feature asymmetry, according to Kovesi [Kov97], the results obtained from the feature symmetry measure can sometimes be counter-intuitive. It is because feature symmetry measures local symmetry to the exclusion of everything else. The measure is invariant to the magnitude of the local contrast, and so features that might be considered to be of little significance can be marked as having strong symmetry (see the flat region of the step edge function in Figure 3.4). Therefore, feature asymmetry (FA) is used as the phase-based detector to extract structure edges in this chapter.

Locally, the 1D signal of the step edge function is illustrated in Figure 3.5. One can see that the axis of asymmetry corresponds to the point where all the frequency components are at the most asymmetric points in their cycles, *i.e.*, the inflection point (the steps on the square wave). Here it should be noted that only the asymmetry of intensity values in images, *i.e.*, a low-level view of symmetry and asymmetry, is being considered. It is assumed that asymmetry represents a generalization of step edges or structure edges in this work [Kov97].

In [FS01] and [FS04], Felsberg and Sommer proposed a 2D isotropic analytic signal which is called monogenic signal. This 2D signal preserves the main property of 1D analytic signal that decomposes a local amplitude and local phase. In their research, the monogenic signal is defined as 2D analytic singal by convolving the input 2D signal with its Riesz transform. Based on the monogenic signal [FS01] (which is defined for the whole image), a local monogenic phase-based measure is defined for a given image to detect asymmetric features. The monogenic signal is defined by combining the 2D signal, f, with its Riesz transform (details can be found in [FS01]) $\overrightarrow{f}_R$ to form:

$$\begin{aligned} f_M(x,y) &= (f(x,y), \overrightarrow{f}_R(x,y)) \\ &= (f(x,y), h_1 * f(x,y), h_2 * f(x,y)), \end{aligned} \tag{3.10}$$

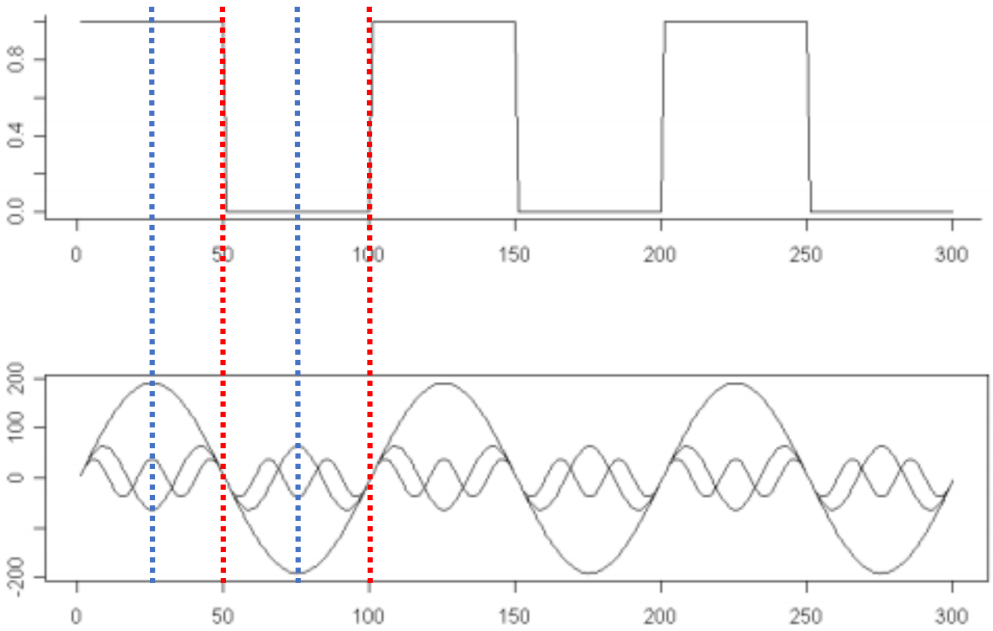


Figure 3.4: Feature symmetry and feature asymmetry (FA). Blue dotted lines indicate the location of feature symmetry and red dotted lines indicate the location of feature asymmetry. The bottom graph illustrate the phases of step edge function in different frequency.

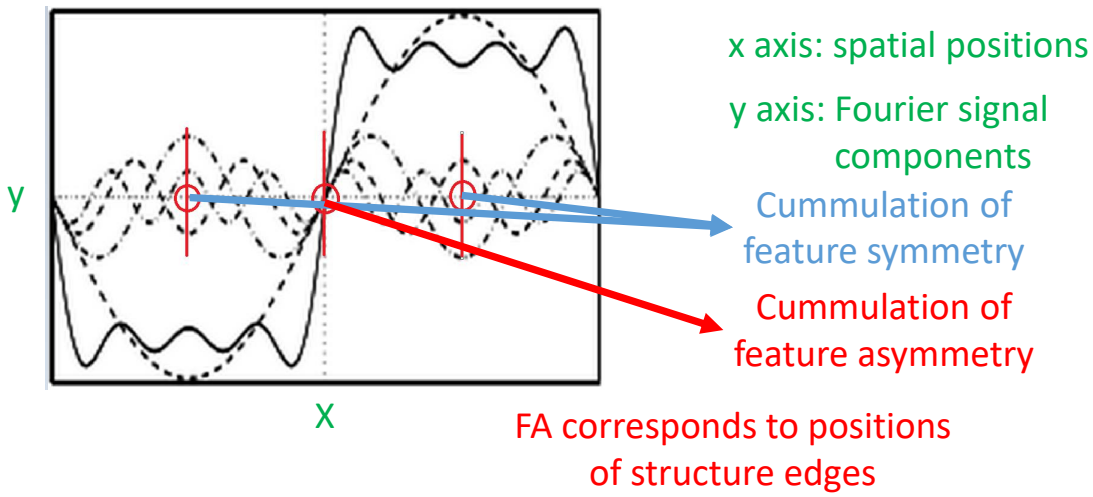


Figure 3.5: Feature symmetry and feature asymmetry (FA) in a local view.

h_1 and h_2 are the Riesz filters, and $*$ is the convolution operator. The definitions of h_1 and h_2 are:

$$h_1(x,y) = -\frac{x}{2\pi(x^2+y^2)^{\frac{3}{2}}}, \quad h_2(x,y) = -\frac{y}{2\pi(x^2+y^2)^{\frac{3}{2}}} \tag{3.11}$$

Since real images generally contain a wide range of frequences, the local properties are analyzed in practical applications via a bunch of bandpass quadrature filters tuned according to various spatial frequencies. The monogenic signal is combined with a set of bandpass filters $c(x,y;s)$ (s is the scale parameter):

$$f_{M,s} = (c*f, c*h_1*f, c*h_2*f) = (even, \overrightarrow{odd}) \tag{3.12}$$

where $even = c*f$ and $\overrightarrow{odd} = (c*h_1*f, c*h_2*f)$. In other words, $c*f$ represents the scalar-valued even and $(c*h_1*f, c*h_2*f)$ represents the vector-valued odd responses of the quadrature filters. In this work, Cauchy kernels are adopted as the bandpass filters due to their exceptional behaviour in localization. A 2D isotropic Cauchy kernel in the frequency domain is defined as:

$$C(k) = |k|^a \exp(-s|k|), \tag{3.13}$$

where $a \geq 1$, $k = (u,v)$, s is the scale parameter of different levels of bandpass filters (the same as in Equation 3.11). Here C is the Fourier transform of function c (equation in the spatial domain can be found in [BNB04]) and in Equation 3.12. More details on the parameters can be found in [BNB04]. The s in all the experiments in this work was taken to be 13, and a was taken to be 1.5.

As Kovesi [Kov97] mentioned, the absolute values of the $\overrightarrow{odd}$ symmetric filter responses are large while the absolute values of the even symmetric filter responses are small at FA points (see Figure 3.4 and 3.5). A way to define FA is to use the differences between the $\overrightarrow{odd}$ and the even symmetric filter responses to detect asymmetric features:

$$FA = \sum_s \frac{\lfloor |\overrightarrow{odd}|_s - |even|_s - T_s \rfloor}{\sqrt{\overrightarrow{odd}_s^2 + even_s^2 + \varepsilon}}, \tag{3.14}$$

where s indicates different scales, ε is a small constant to avoid division by zero, T_s is the scale specific noise threshold, $|\cdot|$ is the $\ell_1 - norm$ for $\overrightarrow{odd}$ and absolute value for $even$, and $\lfloor \cdot \rfloor$ denotes the zeroing of negative values.

From the definition of FA, it first takes values in [0, 1], and is close to 0 in smooth regions and close to 1 near boundaries or structure edges. Second, due to the subtraction of noise level T_s, FA is robust to noise and thus can help detector edges in a noisy image.

Some examples of FA edge maps can be seen in Figure 3.6. It can be observed that edge maps generated using FA correspond well to manually create ground truth (GT) of structure maps provided by Xu *et al.* [XYXJ12]. Unlike the gradient edge maps that contain many edges due to texture, the edge maps from FA capture the object contours, while effectively suppressing edges from periodic patterns.

To further assess the relationship between the structure layer and FA map, the correlation between GT of the structure edge provided in the dataset of Xu *et al.* [XYXJ12], an FA map, and the corresponding image gradient were examined. Here, a normalized cross correlation (NCC) index was adopted as in Chapter 2 for comparing correlation.

The NCC between the GT, FA, and image gradient are compared. Some of the results are listed in Figure 3.8. The mean NCC of 20 images in the dataset is listed in Table 3.1. It can be seen from Table 3.1 that besides qualitative similarity, the FA map also has a higher correlation to the structure edge than the image gradient, which indicates that FA is a better clue for seeking the structure layer.

Table 3.1: Mean NCC of 200 images from dataset of Xu *et al.* [XYXJ12]

	NCC
GT vs. FA	0.66
GT vs. Gradient	0.34

Although the FA map is highly correlated with the structure edge, its edge location (non-zero location) shifts one to three pixels compared to its corresponding gradient location. Thus, a *dilated* morphological operation on the FA map was carried out to obtain a better structure edge descriptor. We compared the dilated FA map with the structure edge and found that NCC was reasonably enhanced due to this operation. Further, this dilated FA map was used as a mask to filter the gradient map of the input image such that the filtered gradient had a reasonably correct location as the GT of structure edges, with an edge magnitude

Figure 3.6: Comparisons between ground truth (GT), feature asymmetry (FA) and gradient mask. Here gradient mask is generated by filtering gradient map with threshold equal to 0.9 multiplies the maximum value of gradient map.

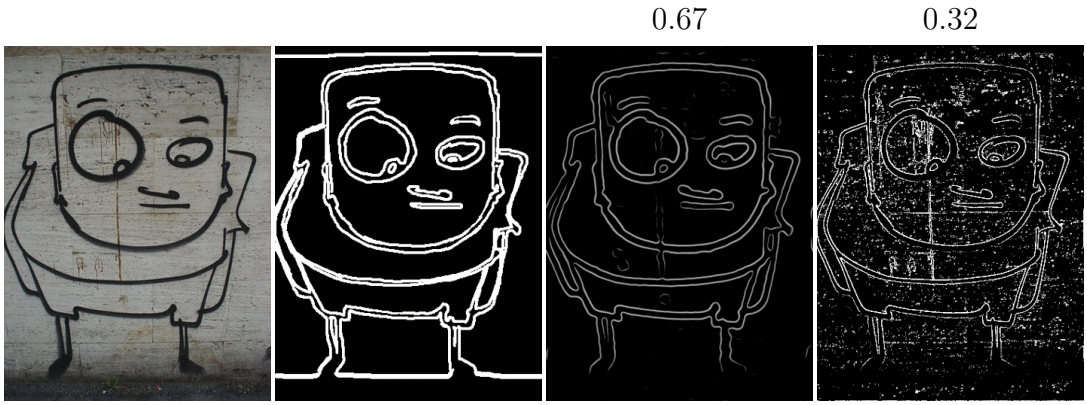


Figure 3.7: Comparisons between FA, GT and gradient map.

that strongly corresponded to that of the original image. Some of the results are listed in Figure 3.8. The mean NCC is listed in Table 3.1. It can be seen that the dilated operator rendered the FA map with better descriptive power for the structure edge. The following explains how this filtered gradient was applied to the WLS framework that enabled us to propose the proposed FA-aware WLS.

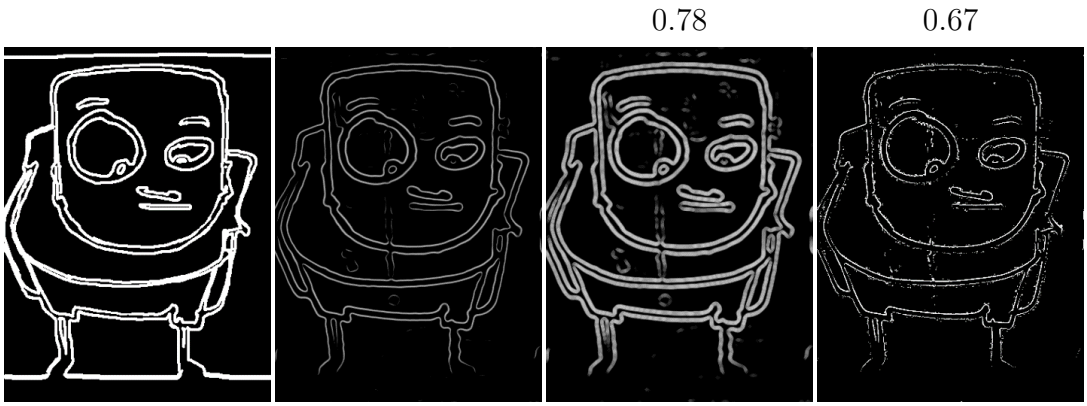


Figure 3.8: Comparisons between GT, dilated FA and gradient map.

3.3.2 Feature Asymmetry Weighted Least Squares (FAWLS)

We aimed at finding a new image S which was as close to I as possible but was also as smooth as possible everywhere, except when passing across significant features to extract a structure layer from a natural image, I. Note $M = H \times W$ as the image size of S and I, where H is the height and W is the width of an image.

The WLS framework has been found to perform well in smoothing image details while preserving edge features [FFLS08]. WLS carries out edge-aware smoothing by minimizing the following energy function:

$$\sum_{p} \Big((I_p - S_p)^2 + \lambda \Big(w_{h,p}(I) \Big(\frac{\partial S}{\partial x} \Big)_p^2 + w_{v,p}(I) \Big(\frac{\partial S}{\partial y} \Big)_p^2 \Big) \Big), \tag{3.15}$$

where p represents image pixels and λ influences the smoothness of the optimized result, S; here, $w_{h,p}$ and $w_{v,p}$ are weights at pixel p for the given image I in terms of horizontal and vertical gradients.

However, WLS does not work well when is directly applied to the structure-texture separation task because of the influences of textures. One main reason is that the weighting functions are usually defined based on image gradients, are rather weak in indicating structures or contours, and are easily influenced by textures, as shown in Figure 3.6. this problem was solved by incorporating FA (Equation 3.14) into the WLS framework. FA first serves as a mask to filter the gradient of the input image, and then the filtered gradient is included in the WLS framework to enhance the effect of the plausible structure edge. A feature asymmetry weighted least squares (FAWLS) framework was specifically designed by setting the weighting functions, $w_{h,p}$ and $w_{v,p}$ as

$$\begin{aligned} w_{h,p}(I) &= ((FA_p(I) + 1)\partial_x I)^{\alpha} + \varepsilon')^{-1}, \\ w_{v,p}(I) &= ((FA_p(I) + 1)\partial_y I)^{\alpha} + \varepsilon')^{-1}, \end{aligned} \tag{3.16}$$

where α controls the sensitivity of the FA edge map (see Equation 3.14) and gradient ∂_x, ∂_y; ε' is added to avoid dividing by zero. The value of FA can be set between one and two. α is set to 1.5 in the experiments in this work. As was demonstrated in the experiments in this work, $w_{h,p}$ and $w_{v,p}$ in Equation 3.16 had a large effect in adapting WLS to structure-preserving image smoothing. The FA measure is sensitive to structure edges and less sensitive to texture edges, which is different from gradient-based operators. The smoothing process of WLS is prevented near the structure but is greatly encouraged in homogeneous regions or regions with rich textures, which results in smoothed images with structures being preserved. The pipeline of the proposed method can be seen in Figure 3.9.

For the acceleration, by arranging the summing up of pixel-wise quadratic sums into a vectorization version, first we see that Equation 3.15 can be rewritten into

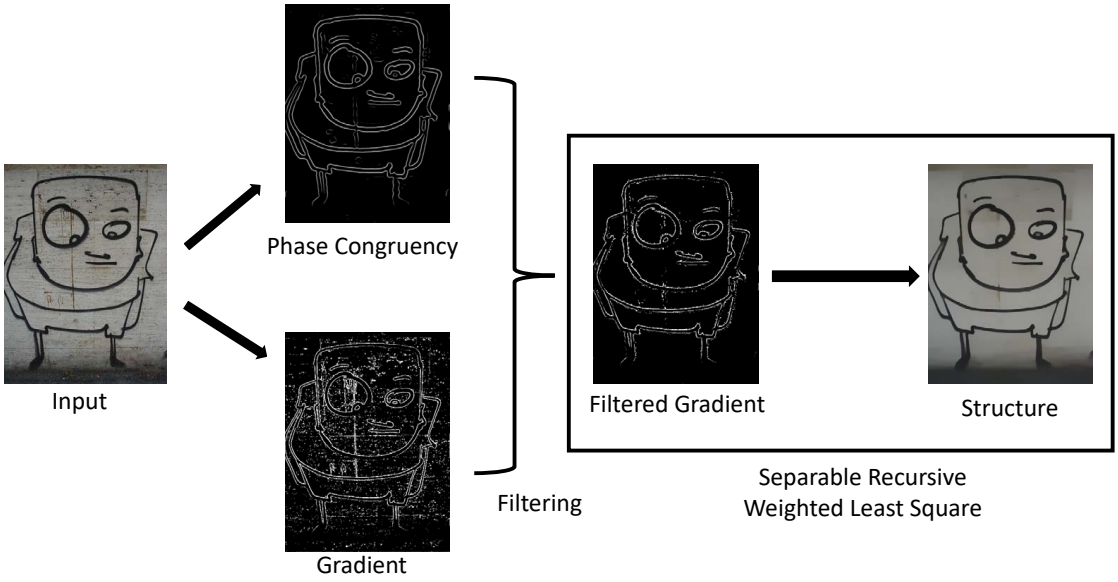


Figure 3.9: Pipeline of the proposed method.

the matrix form:

$$(q^I - q^S)^T(q^I - q^S) + \lambda(q^{S^T} D_h^T W_h D_h q^S + q^{S^T} D_v^T W_v D_v q^S), \tag{3.17}$$

where q^I and q^S are the 1D vectorization of I and S, D_h and D_v are the Toeplitz matrices formed by arranging the discrete gradient operators with a forward difference according to Xu *et al.* [XYXJ12]. The W_h and W_v are diagonal matrices whose values are set at $W_h(i;i) = w_{h,p_i}(I)$ and $W_v(i;i) = w_{v,p_i}(I)$, where i is the order of pixel p_i in the vector formed from I.

To be in details, q^I and q^S have dimension of $M \times 1$ (remembering that M is the size of image pixels). D_h, D_v, W_h and W_v have the dimension of $M \times M$. $D_h q^S$ is the vectorization of $\left(\frac{\partial S}{\partial x}\right)_p^2$ and $D_v q^S$ is the vectorization of $\left(\frac{\partial S}{\partial y}\right)_p^2$ obtained from S.

Solving Equation 3.15 can be done iteratively by the following two steps:

- From the estimated structure image S in the second iteration, D_h^T, D_q^T, W_h and W_v can be calculated based on Equation 3.16 and property of discrete gradient operators with a forward difference.

- To update the structure image S, Equation 3.17(so as 3.15) is solved. Its optimization can be written into the following analytical sparse linear system

$$[J + \lambda(D_h^T W_h D_h + D_q^T W_v D_v)]q^S = q^I, \tag{3.18}$$

or can be written as

$$(J + \lambda L)q^S = q^I, \tag{3.19}$$

where J is the identity matrix. $L = (D_h^T W_h D_h + D_q^T W_v D_v)$ is the symmetric positive definite Laplacian matrix, whose dimension is $M \times M$.

Further, according to Min *et al.* [MCL+14], Equation 3.19 can be solved by iteratively solving two linear systems:

$$(J_h + \lambda_t L_h)q^{S,h} = q^{I,h} \; and \tag{3.20}$$

$$(J_v + \lambda_t L_v)q^{S,v} = q^{I,v}, \tag{3.21}$$

where subindices h and v indicate horizontal and vertical operations, J_h, J_v are identity matrices with sizes of $W \times W$ and $H \times H$. And L_h and L_v are Laplacian matrices with sizes of $W \times W$ and $H \times H$. The $q^{I,h}$ and $q^{S,h}$ represent the vector notations of I_h and S_h, which are the horizontal vectors of I and S (rows of an image), and $q^{I,v}$ and $q^{S,v}$ represent the vector notations of I_v and S_v (colum of an image), which are the vertical vectors of I and S. Note that here λ_t was set to be $\frac{3}{2}\frac{4^{T-t}}{4^T-1}\lambda$, representing the step size of each iteration for the recursive optimization. In this work λ_t is defined the same way as in [MCL+14]. By this 1D separable implementation, we reduce the problem of solving a complex equation system to solving two much simpler equation system.

The solution, $q^{S,h}$, for Equation 3.20 is obtained by recursively forward-backward computing:

$$\begin{aligned} a_x &= \lambda_t L_h(x, x-1), \\ b_x &= 1 + \lambda_t L_h(x, x), \\ c_x &= \lambda_t L_h(x, x+1), \\ \tilde{c}_x &= c_x/(b_x - \tilde{c}_{x-1}a_x), \\ \tilde{q}_x^{S,h} &= (q_x^{S,h} - \tilde{q}_{x-1}^{S,h})/(b_x - \tilde{c}_{x-1}a_x), \end{aligned} \tag{3.22}$$

with $\tilde{c}_0 = c_0/b_0$ and $\tilde{q}_0^{S,h} = q_0^{S,h}/b_0$ and with $x = 1, \ldots, W-1$.

$$q_x^{I,h} = \tilde{q}_x^{S,h} - \tilde{c}_x q_{x+1}^{S,I}, \tag{3.23}$$

with $q_{W-1}^{I,h} = q_{W-1}^{S,h}$ and with $x = W-1, \ldots, 0$. Similar recursive computation is carried out for the solution, $q^{S,v}$, of Equation 3.21. The whole process is summarized in Algorithm 2.

Algorithm 2 Separable Global Smoother for FAWLS

Parameters:

T: iteration number, $M = H \times W$: image size (H: height and W: width)

λ: smoothing parameter

L_h (or L_v): $W \times W$ (or $H \times H$) three-point Laplacian matrix

I^h, FA^h, S^h: 1D horizontal signal of I, FA, S;

$q^{I,h}$, $q^{S,h}$: of dimension $W \times 1$, 1D vectorization of rows of I, S

I^v, FA^v, S^v: 1D vertical signal of I, FA, S

$q^{I,v}$, $q^{S,v}$: of dimension $H \times 1$, 1D vectorization of colums of I, S

Input:

2D image $I(x, y)$; q^I ($M \times 1$ vector)

2D FA map $FA(x, y)$;

Initialization: $S \leftarrow I$

for $t = 1 : T$ **do**

compute $\lambda_t = \frac{3}{2}\frac{4^{T-t}}{4^T-1}\lambda$ (λ_t is defined as in [MCL+14])

for $y = 0 : H - 1$ **do**

$q^{S,h}(x) \leftarrow S(x, y)$ for all $x = 0, \ldots, W - 1$

compute $w_{h,x}$ using $FA^h(x, y)$ for all $x = 0, \ldots, W - 1$

build a tridiagonal L_h

solve $(q^{I,h} + \lambda_t L_h)q^{S,h} = q^{S,h}$ using 3.22

$S(x, y) \leftarrow q^{S,h}(x)$ for all $x = 0, \ldots, W - 1$

end for

for $x = 0 : W - 1$ **do**

$q^{S,v}(y) \leftarrow S(x, y)$ for all $y = 0, \ldots, H - 1$

compute $w_{v,y}$ using $FA^v(x, y)$ for all $y = 0, \ldots, H - 1$

build a tridiagonal L_v

solve $(q^{I,v} + \lambda_t L_v)q^{S,v} = q^{S,v}$ using 3.23

$S(x, y) \leftarrow q^{S,v}(y)$ for all $y = 0, \ldots, H - 1$

end for

end for

Output: 2D image $S(x, y)$; q^S ($M \times 1$ vector)

3.4 Experiments

The proposed approach was compared with state-of-the-art structure-texture separation methods in the experiments: the WLS [FFLS08], the relative total variation (RTV) [XYXJ12], and the region covariances-based NL means (RCNLM) filter [KEE13] on two images from Xu *et al.* [XYXJ12] (the first and third rows in Figure 3.10) and one image from Karacan *et al.* [KEE13] (the second row in Figure 3.10). Experiments were carried out on a PC equipped with an Intel i7-3612QM 2.10-GHz central processing unit (CPU) with 8 GB of memory. The code was written in Matlab 2013b. For the 1D recursive optimization part of WLS, the C-Matlab executable (MEX) version from Min *et al.* [MCL+14] was used. The source codes of the methods that were compared were obtained from websites provided by the authors of [FFLS08, XYXJ12, KEE13]. The parameters of all the tested methods were either set as mentioned in the corresponding papers or carefully tuned to achieve good results. All the methods were qualitatively evaluated on the basis that an efficient method should only smooth fine details and textures and preserve the structure. Also, the extracted textures or so-called detail components should be devoid of any information regarding the structure. Computational complexities were also compared, and the results are listed in Table 3.2. The parameter, λ, was set to 0.01 in all the experiments in this work on the three images.

Table 3.2: Computational Complexity of Images in Figure 3.10 for different methods. (numbers in parentheses indicate the sizes).

Method	Row1 (1024×768)	Row2 (710×511)	Row3 (495×536)
WLS [FFLS08]	7.0s	4.2s	2.9s
RTV [XYXJ12]	8.8s	4.9s	3.6s
RCNLM [KEE13]	1124.2s	553.4s	358.6s
Proposed	1.6s	1.1s	1.1s

It can be seen that from Figure 3.10 and Table 3.2 the proposed method achieved comparable or better separation of structure layers at less computational cost than the other methods. To be more specific, WLS yielded large color bleeding and blurring effects. Although RCNLM gave pleasing qualitative results, its

runtime was significantly longer than that of the other methods. Finally, although both RTV and the proposed method could achieve exceptional smoothed images, the proposed method ran largely faster. Also, as can be seen from Row 2 in Figure 3.10, some artifacts could be enhanced by RTV.

Figure 3.10: Comparisons between the proposed method (FAWLS), WLS [FFLS08], RTV [XYXJ12] and RCNLM [KEE13].

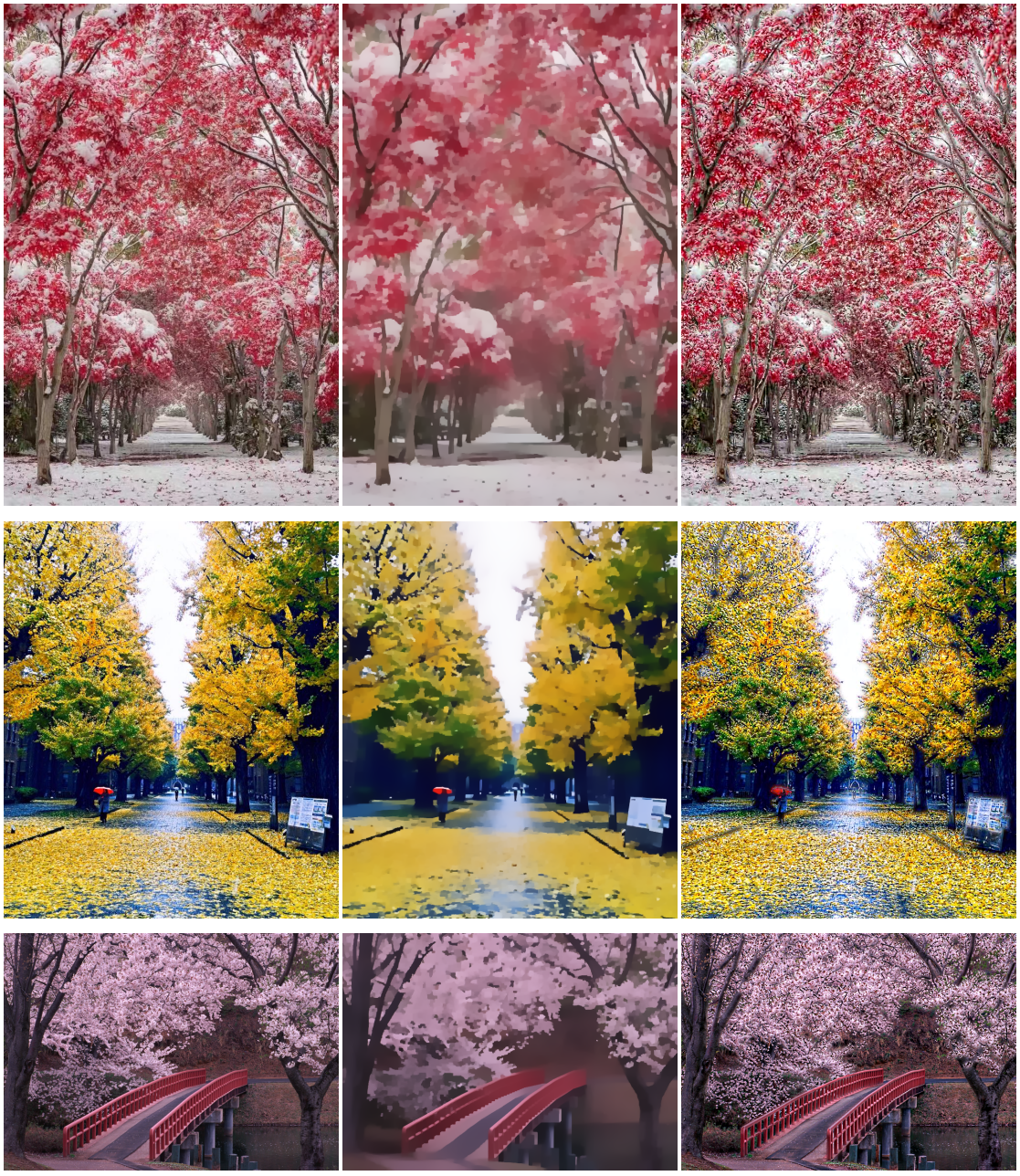
All the images were resized in the dataset provided by Xu *et al.* [XYXJ12] to 200×200, 400×400 and 800×800 to further compare the computational efficiency of the proposed methods with those of others and have listed the computational times in Table 3.3. It can be seen that the proposed method ran faster than all the state-of-the-art approaches for all the tested resolutions of images.

Table 3.3: Mean running time of different methods on dataset [XYXJ12]

Method	200×200	400×400	800×800
WLS [FFLS08]	1.2s	3.4s	7.4s
RTV [XYXJ12]	1.3s	3.9s	9.4s
RCNLM [KEE13]	261.5s	517.3s	1454.0s
Proposed	0.4s	0.6s	1.8s

3.5 Applications

3.5.1 Image Processing



Input image Structure Enhanced image

Figure 3.11: Results from detail enhancements.

One typical application for structure texture separation is image enhancement. To obtain the fine details of an image which is represented by texture layer, it is crucial to achieving high-quality of structure-texture separation result. By correctly detecting the structure edge with the guide of FA, the proposed method

can smooth away highly contrasting textures without distorting the gradients of structure edges. Examples can be seen in Figure 3.11.

The proposed method can be used for high quality of inverse halftoning. Figure 3.12 shows the results for inverse halftoning that aims to remove stipple dots from the halftone images. While the proposed method is not tailored to solve this particular problem, it shows good performance regarding removing dots while keeping important structure edges without the need for post-processing such as shock filtering.

Input [KL12] Proposed

Figure 3.12: Inverse halftoning. Kopf and Liscinski's method [KL12] was dedicated to this problem and produced the best results. Compared to [KL12], the proposed method can give comparably good shape of the original black lines. (A crop from Iron Man ©Marvel Comics).

3.5.2 Preprocessing for Intrinsic Image Decomposition

As was explained in Chapter 2, the existence of rich and fine textures would violate the Retinex assumption and make many methods of edge-based intrinsic image

decomposition less effective. By removing the textures from the objects' surfaces or from the scene by preprocessing via structure-texture separation techniques, we could attain a piece-wise smooth structure layer that could serve as an excellent input for intrinsic image decomposition. Intrinsic image decomposition was adopted in Figure 3.13 on the structure layer that was obtained in Figure 3.10. It can be seen that by sequentially processing the layer separation methods (structure texture separation then intrinsic image decomposition), we can obtain a sequential decomposition of layers: albedo layer, shading layer and texture layer, which can be input for different post-processing and further applications. Also, texture layer here represents local variations of albedo, thus can be considered as a complement of albedo layer.

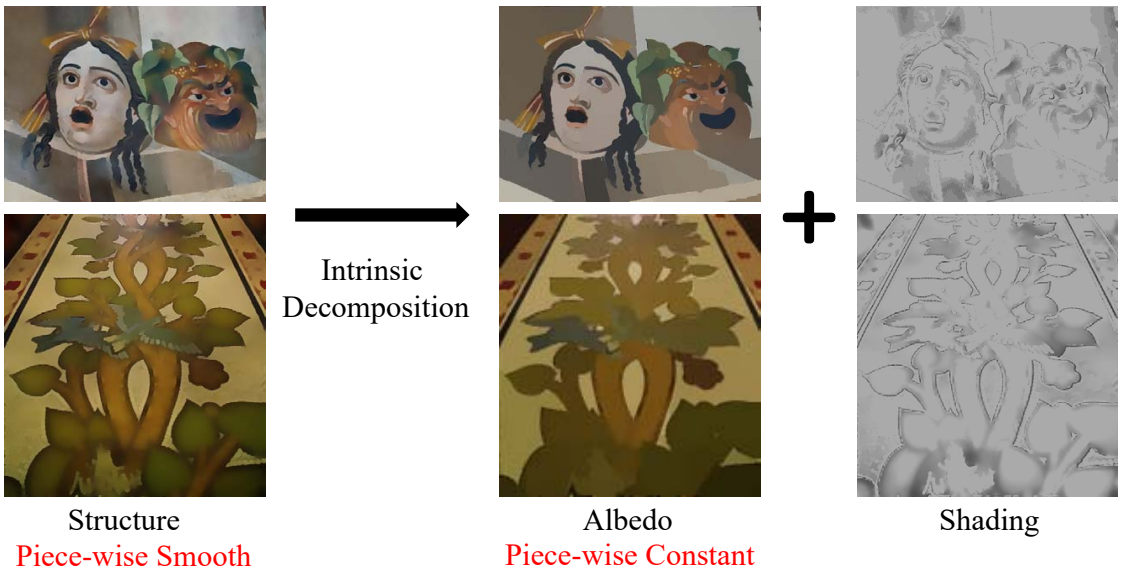


Figure 3.13: Structure texture separation as preprocessing for intrinsic image decomposition.

Apart from these discussed applications, the proposed method can also be used for other applications such as noise removal, tone mapping, and texture replacement.

Chapter 4

Intrinsic Image Decomposition for Materials with Fine-Textures via RGB-NIR Images

4.1 Introduction

Many object's surfaces contain rich textures, especially those of materials such as fabrics, plastics, metals and leathers. These rich and fine textures will cause problems in certain cases for some low-level vision algorithms such as intrinsic image decomposition [BT78] and segmentation. We deal with Intrinsic image decomposition of objects with rich texture in this work. It tackles the problem of separating an input image into an albedo layer that reflects the material's color, and a shading layer that indicates the geometry of object's surfaces and the lighting of the scene in which objects may contain rich textures.

The use of local features, such as those in gradient and Laplacian information is a well-known way of performing intrinsic decomposition. However, using only a single RGB image with local differential information would be less effective for complicated scenes such as those with rich textures or ones in which materials had complicated surfaces. Problems, such as color leaks in the albedo layer or remnant texture in the shading layer after decomposition, can occur in these cases, and the decomposition results would be less useful for further applications. Also, efficient algorithms for intrinsic decomposition greatly depend on the prior sparsity of

edges with significant magnitudes in the albedo layer, which would be violated by the existence of rich and fine textures that contain some edges of significant magnitude.

The range of NIR light is between visible red light and long infra-red (LIR) light in the electromagnetic spectrum, and NIR has wavelengths in the range of 750–1400 nm, which is longer than that of visible light (380–750 nm). The human eye cannot see NIR light, but most digital cameras can sense it very well. As can be observed in our daily lives, NIR images of many materials exhibit far fewer texture (or albedo) variations. For example, textures on mugs, towels, and fabrics appear to have a uniform color in the infra-red (IR) spectrum, whereas, black ink is visible in both visible and IR spectra. Salamati *et al.* [SFS09a] further analyzed this NIR light's properties. They captured images of many different types of materials in the visible and NIR spectra, and then analyzed the luma, intensity, and color information of images in the two different spectra and found that many pigments used to colorize materials appeared to be transparent in the NIR spectrum.

It is still difficult to understand the texture of complex scenes due to the difficulty of using only a single RGB image or to further process intrinsic decomposition. However, as the properties of NIR images can help to avoid the effect of textures and obtain rough information on the shading layer, the use of the NIR layer as a guide map is proposed to efficiently carry out intrinsic decomposition.

This property of NIR images makes them suitable for the task of intrinsic image decomposition. The albedo (or reflectance) layer normally contains a small number of structure edges with relatively large magnitude, according to the basic assumption of intrinsic image decomposition, which is the Lambertian surface condition as well as the Retinex assumption. The existence of rich texture will violate these assumptions and make the Retinex-based method less effective. The NIR layer in certain wavelengths is not just a complementary channel for the R, G and B of color images, but it also provides different physics behaviors that can hint to enable layer separation. The reflectance spectral curves of most materials will become largely flat above wavelengths from 800 to 1100 nm, which is approximately the interval of NIR spectra. The spectral curve in such intervals will even

be constant value, especially for fabric materials. This means that images that have a similar appearance to the actual shading layer of the scene can be obtained by capturing an NIR image of the materials in a scene (Figure 4.6). This chapter attempts to prove this with a discussion in subsection 4.3.

We propose a novel edge-based method of solving intrinsic image decomposition with an extra NIR image by incorporating it, as well as the chromaticity information on the input RGB image, beyond the Retinex assumption [LM71]. In fact, chromaticity invariance against shading is well known and has previously been adopted for several applications [Dre03, FHLD06]. Yu and Sato claimed that a chromaticity map contains similar edges as those in the albedo layer, and the chromaticity gradient indicates the location of the albedo gradient. Such information could be adopted to enable better recovery the albedo gradient. Here, we define a pseudo-albedo map, which is the division of the pixel values in RGB and NIR images, whose gradient is filtered by chromaticity gradient information to obtain prior information on the albedo gradient. All this information is included in a maximizing-a-posterior (MAP) framework to build the proposed RGB-NIR intrinsic image decomposition model. Decomposition is done very quickly by reducing the optimization problem here into an alternating minimization problem [WYYZ08] and adopting Plancherel's theorem and a fast Fourier transform (FFT) to accelerate computation. The results obtained from experiments revealed that the proposed method ran much faster and outperformed state-of-the-art edge-based methods that used sophisticated models and inference.

To sum up, the five main highlights of the proposed work include:

- We built a dataset with plausible ground truths of shading layers and some albedo layers of materials in the dataset provided by Choe *et al.* [CNK16] using a method of photometric stereo.

- We proposed to include an NIR image into the framework of intrinsic image decomposition by applying the properties of some materials that they are almost textureless in the NIR channel. This property helped to largely remove the effect of textures from the RGB image using certain fusion method and thus yielded stronger clues to the Retinex assumption.

- We analyzed the correlation between an RGB image and an albedo layer through the ℓ_1-norm based chromaticity map, and the correlation between an NIR image and a shading layer.

- We defined a pseudo-albedo map based on RGB and NIR images and incorporated it into a MAP model for intrinsic image decomposition.

- We proposed a fast solver based on Plancherel's theorem and a half-quadratic splitting scheme [WYYZ08] to efficiently solve the proposed MAP problem and speed up the runtime while maintaining the accuracy of decomposition. The proposed method ran much faster, and the results of the proposed method were better than those obtained from state-of-the-art edge-based methods that used sophisticated models and inference.

The remaining parts of the chapter is organized in this way: related work is presented in subsection 4.2. subsection 4.3 overviews the proposed approach and subsection 4.4 provides experimental comparisons with prior approaches.

4.2 Related Work

4.2.1 NIR Imaging

Multispectral data captured in the NIR band of the spectrum have been conventionally used in specialized areas such as security and surveillance, the food industry, medical applications, and remote sensing [ESA12, Sch, YR10], where imaging systems that have been specially designed for industrial and scientific purposes have been employed. The additional information provided by the differences between color and NIR images has recently been used in several tasks in the domain of computational photography and computer vision such as high-dynamic-range imaging [ZSM08], image denoising [QY13] [MSH14], dehazing [SFS09b], shadow detection [RFS14], material-based segmentation [SFS09a], semantic segmentation [SLCS12], and scene understanding [BS11]. Fredembach and Sustrunk [FS09] estimated environmental illuminants by using ratios of RGB and NIR images. They pinpointed the location of diverse illuminants and recovered a lighting map by using an RGB and NIR pair. Kerl *et al.* [KSSC14] used NIR and RGB images

to compute a shading invariant map. NIR images have also been used in various 3D applications such as enhancing depth quality or refining the geometry of 3D meshes [CPTK14] [HCG14]. These applications have indicated that NIR images are not affected by general indoor (ambient) lighting, which simplifies the lighting model. Choe *et al*. built an NIR dataset (NISAR dataset) containing 100 materials for bidirectional reflection distribution function (BRDF) estimation using NIR images. They also provided the corresponding RGB images acquired under the single-view and same environment as their NIR image pair.

This chapter explains how plausible GT was built for the shading layer of an RGB image and selected some materials that had the most diffuse appearance for building albedo-shading layers based on the NISAR dataset. This will be described in detail in the next subsection.

4.3 The Proposed Method

4.3.1 Image Formation Model

The intrinsic image decomposition problem in general form is highly underconstrained because each pixel corresponds to unknown shading and albedo. Our technique focused on a scenario described by four hypotheses:

- The cameras employed for acquiring RGB and NIR images have narrowband sensitivity curves.
- As the surfaces are Lambertian, an image color is the product of shading and albedo in each channel.
- NIR images have almost uniform albedo values except for a sparse number of locations. (*e.g*. Figure 4.4 and 4.5)
- Illumination is white. Thus, white balance effect does not need to be considered otherwise the preprocessing for white balance is necessary.

The radiance (which is also the captured image), I, toward the camera at each non-emissive visible point corresponding to a pixel is formally given by:

$$\begin{aligned} I_k(x,y) &= \int c_k(\lambda)\rho(x,y,\lambda)\overrightarrow{n}(x,y)\cdot\overrightarrow{l}\ s(\lambda)d\lambda \\ &= \overrightarrow{n}(x,y)\cdot\overrightarrow{l}\int c_k(\lambda)\rho(x,y,\lambda)s(\lambda)d\lambda \\ &= \overrightarrow{n}(x,y)\cdot\overrightarrow{l}\ \ c_k(\lambda_k)\rho(x,y,\lambda_k)s(\lambda_k) \\ &= L(x,y)R_k(x,y) \end{aligned} \tag{4.1}$$

where λ is wavelength, ρ is the albedo coefficient, c_k describes the camera spectral response depending on k-th channel (k is R, G, B and NIR), and $\overrightarrow{n}$ is the surface normal. Here, $\overrightarrow{l}$ is the lighting direction, (x, y) is a pixel coordinate in the image, and $s(\lambda)$ represents the illumination spectra. R_k are the albedo layer of k-th channels and L is the shading layer ($L(x,y) = \overrightarrow{n}(x,y)\cdot\overrightarrow{l}$ and $R_k(x,y) = c_k(\lambda_k)\rho(x,y,\lambda_k)s(\lambda_k)$). The second last equal sign is valid under the narrowband camera assumption.

4.3.2 Building Albedo-Shading Ground Trutch (GT)

This subsection describes how the GT of albedo and shading layers of NIR images were built based on 100 materials from the NISAR dataset [CNK16]. The photometric stereo information on each material was given in this dataset under the condition that: each material sample was tilted to 9 different angles with respect to (w.r.t) the camera and imaged under 12 different lighting directions, giving a total of 108 images with sizes of $\sim 1600 \times 1200$ pixels. Finally, 12 warped images were obtained under different lighting from the 108 images, all of which faced the same camera position.

Cheo *et al.* [CNK16] compared the albedo features of NIR images to RGB images, and they also captured an RGB image using a visible-light camera with the same resolution as an NIR camera. For the capture of NIR images, they used an IR pass filter that blocked visible light under wavelengths of 760 nm. They included a variety of fabrics (knits, weaves, cotton, satin, and leather), organic materials (skin, leaves, jute, tree trunk, and fur) and inorganic materials (plastic, concrete, and carpet). Figure 4.6 shows a representative set of 27 materials from

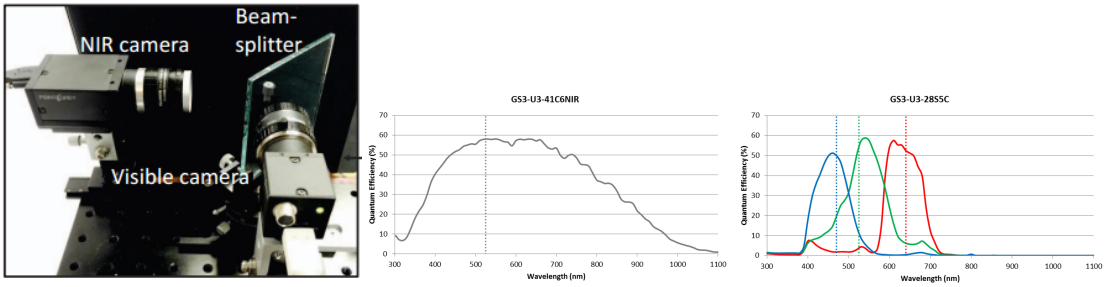


Figure 4.1: Acquisition setup (left, the Choe's setup). NIR and visible light cameras are co-located via beam-splitter. Camera spectral sensitivity curve of pointgrey "GS3-U3-41C6NIR-C" (middle) and "GS3-U3-28S5C" (right) for capturing NIR and RGB images is illustrated.

the database (images have been taken from Choe *et al.* [CNK16]). It can be observed from this dataset that the NIR images had less texture than the RGB images.

The camera setup for the RGB and NIR cameras in Choe *et al.*'s work [CNK16] is the same as that in Figure 4.1. They used a pointgrey camera "GS3-U3-41C6NIR-C" with an IR pass filter that blocked visible light under a wavelength of 760 nm. The spectral sensitivity of the camera was the same as that in Figure 4.1. Although they did not describe what kind of Pointgrey they used for acquiring visible light, it normally has camera sensitivity as shown in Figure 4.1. Incorporating the effect of the filters used for the image acquirement, it is reasonable to make the narrow-band assumption with the NISAR dataset.

The shading layer of each material surface can be obtained under the Lambertian condition by:

$$L_{nir}(x,y) = \max(< \overrightarrow{n}(x,y), \overrightarrow{l}_{nir} >, 0), \tag{4.2}$$

where $p = (x, y)$ means the operation is pixel-wise, $\overrightarrow{n}$ is the 3D surface normal, and $\overrightarrow{l}_{nir}$ is the direction of light. According to the image acquisition process, the light source of the RGB image can be considered as ambient light (diffusion), whereas the light sources of the NIR image are direction light. Instead of calculating the shading from individual light source direction, the shading that shares a similar view direction of the RGB image need to be calculated. Viewing the spatial distribution of 12 light sources, by averaging the lighting of all the light sources,

a mean shading layer that has the same camera setting can be obtained for the averaged NIR image and the captured RGB image. We then obtain the GT of albedo for RGB and NIR images by dividing the input image value pixel-wise by using the obtained shading value. Examples of shading layers recovered in this way can be found in Figure 4.3.

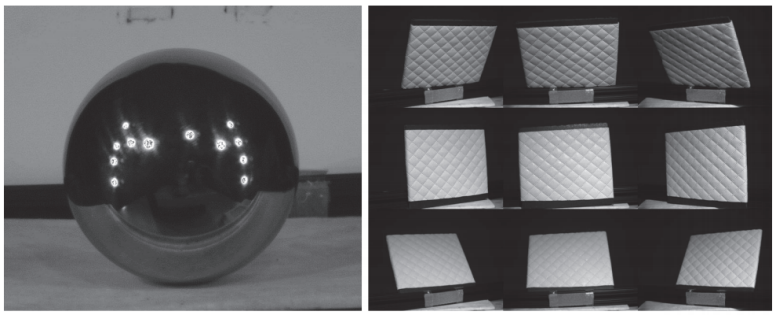


Figure 4.2: Calibration data using mirror sphere in [CNK16].

By actually dividing the estimated GT shading layer of the material images, the NIR albedo can roughly be obtained by:

$$R_{nir} = \frac{I_{nir}}{L_{nir}}. \tag{4.3}$$

Some examples of the NIR albedo layer are shown in Figure 4.4 and 4.5.

Apart from the GT of the shading layer, the GT of the albedo layer is needed to attain better results in the quantitative comparison experiments. With RGB image I_k (k=R, G, B) and the shading layer obtained above, the albedo layer can be obtained by:

$$R_k = \frac{I_k}{L}, \tag{4.4}$$

where L is the average shading of L_{nir}. However, of the 100 give materials, not all of them satisfy the Lambertian reflectance assumption. Therefore, 27 materials are manually selected which is listed in Figure 4.6 to build our ground truth of albedo layers. Besides the diffuse reflectance property of these selected materials, their RGB-NIR image pairs contain as well general cases of appearance of textures (as in Figure 4.9) for testing the effectiveness of the proposed method.

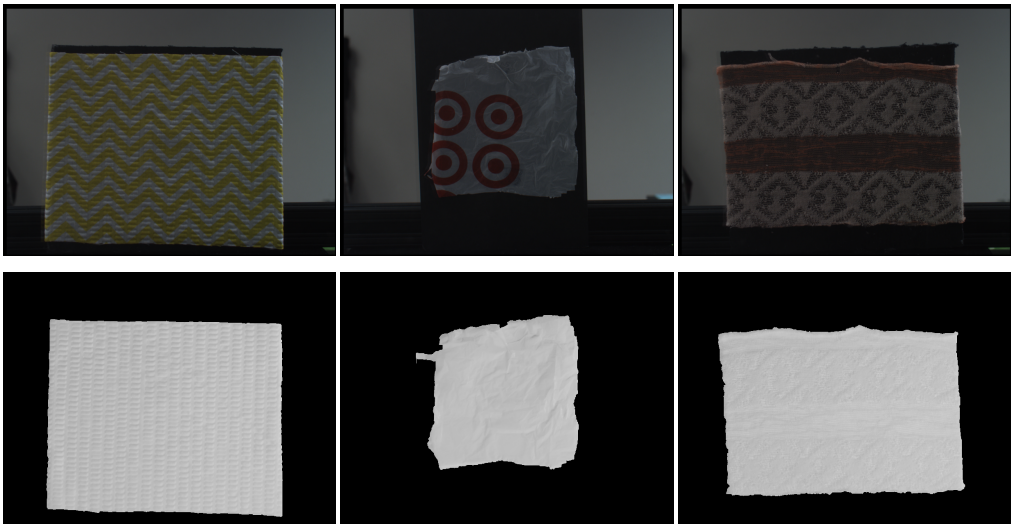


Figure 4.3: Examples of shading layers of some materials.

4.3.3 Albedo Edges, Shading Edges, and Pseudo-albedo

This section analyzes the relationship between the shading edge and NIR gradient, as well as defines the pseudo-albedo, which is an important clue for the proposed RGB-NIR intrinsic image decomposition model. The relationship between an albedo layer and a chromaticity map has been discussed in Chapter 2. Some examples can be seen in Figures 4.7.

The correlation between the NIR intensity edge and shading edge (∇L) will now be discussed. The NIR intensity edge is defined by the gradient as

$$\nabla N = \nabla(R_{nir}L) = \nabla R_{nir}L + R_{nir}\nabla L \tag{4.5}$$

Under the assumption that NIR has almost a uniform albedo (which is valid with the NISAR dataset), which means ∇R_{nir} equals zero almost everywhere, Equation 4.5 is $\nabla N = R_{nir}\nabla L$. Therefore, the NIR intensity edge and the shading edge have the same value except for the ratio of albedo. Without loss of generality, we assume the constant albedo value R_{nir} to be of value 1, as in the following.

We define the pseudo-albedo (PA) to be the division of RGB and NIR images, which is (Figure 4.8):

$$PA_k \triangleq \frac{I_k}{I_{nir}} = \frac{R_k L}{R_{nir}L} = \frac{R_k}{R_{nir}},\ k = r, g, b. \tag{4.6}$$

Under the assumption that NIR has a uniform albedo, the pseudo-albedo has the same value as the albedo layer except for the ratio of the NIR albedo (R_{nir}).

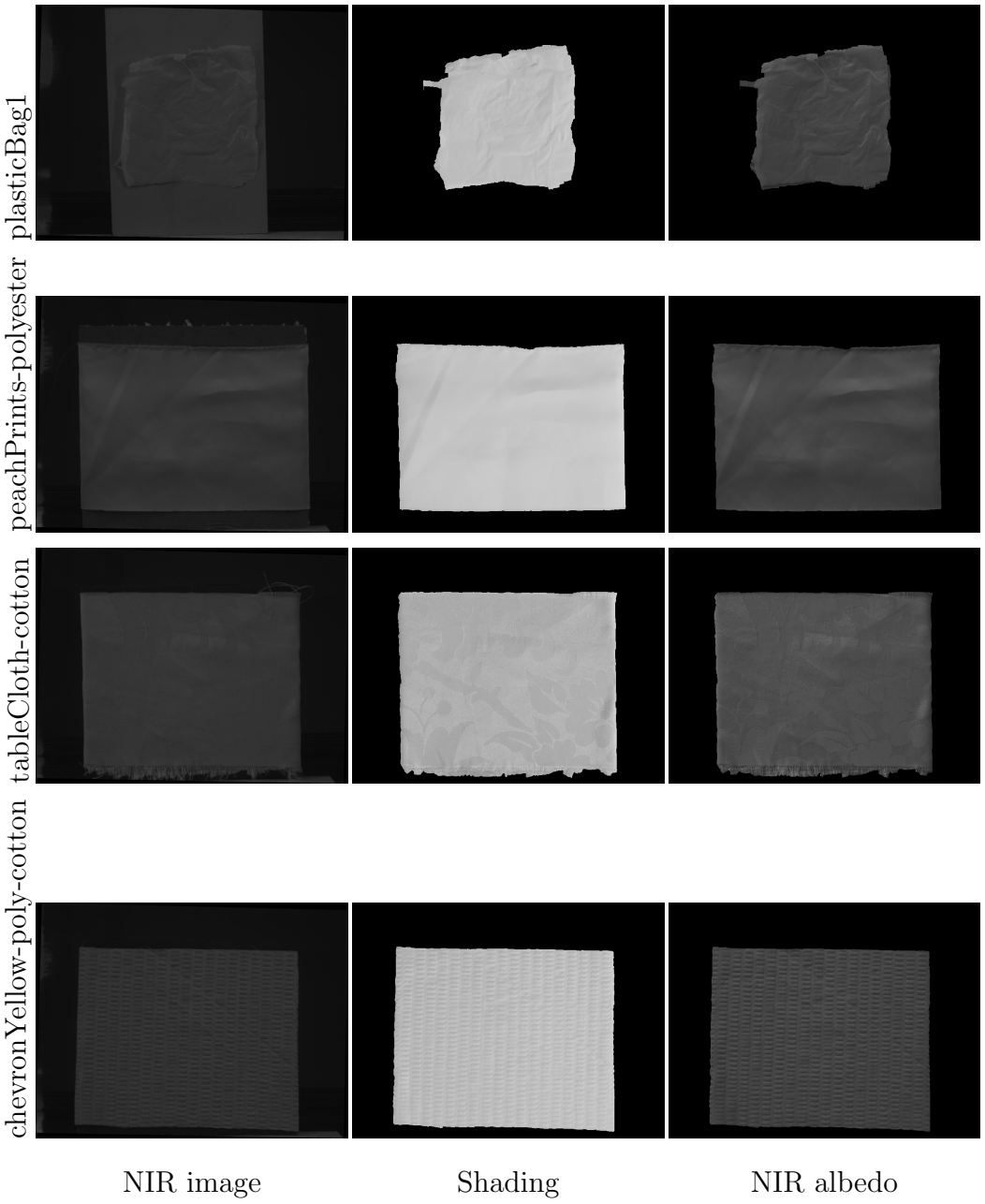


Figure 4.4: Examples of NIR images and the built shading and albedo layers.

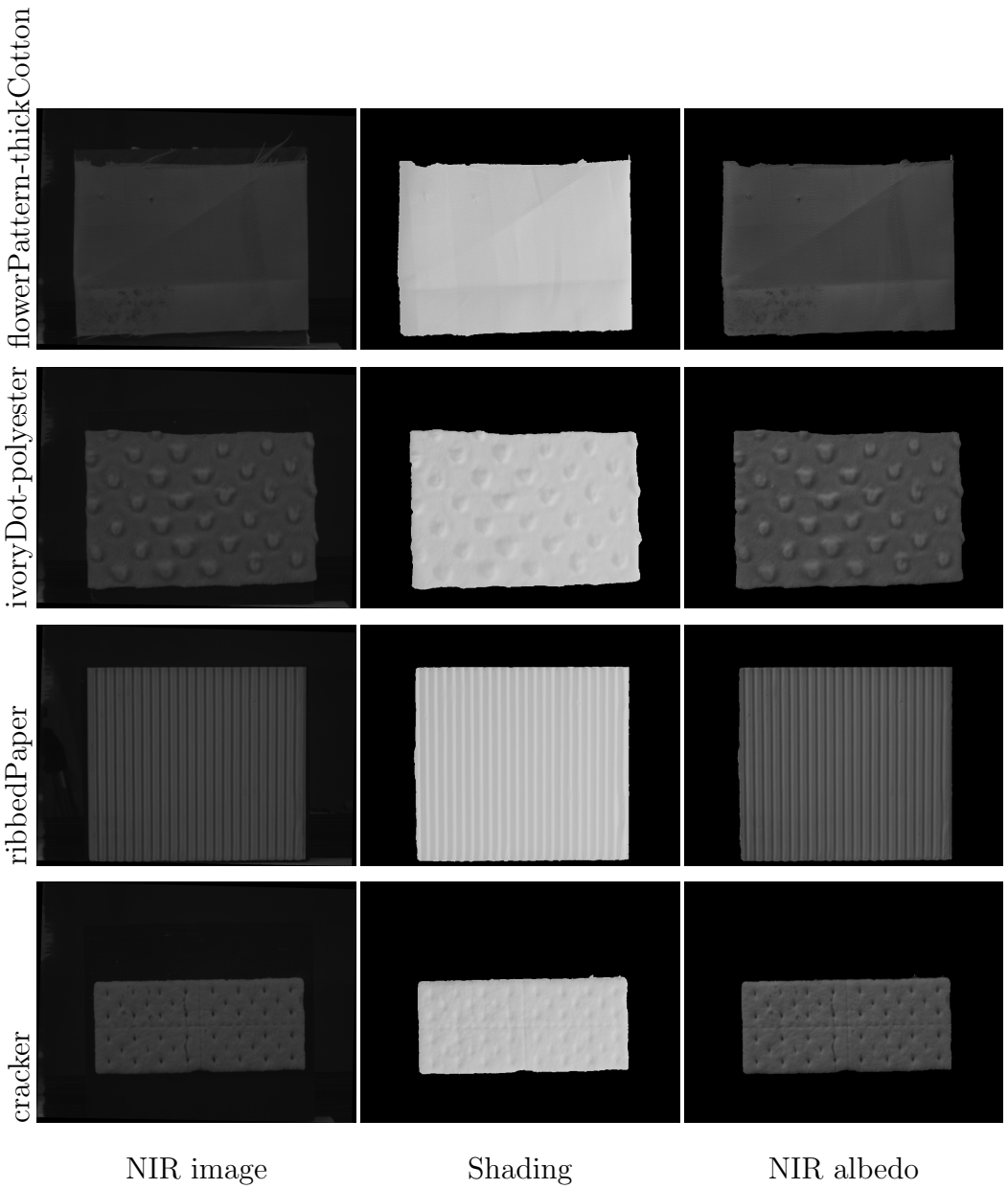


Figure 4.5: Examples of NIR images and the built shading and albedo layers.

Figure 4.6: Selected RGB-NIR image pairs that satisfy the property that surfaces of captured objects have diffused reflectance. From left to right and top to bottom, the materials are (same name used in NISAR dataset): artificialLeather-polyester-suede, artificialShell, blackNwhite-jacquardPolyester, blue-quilterFlannel, chevronYellow-poly-cotton, cracker, floralJacquard-polyester, flowerPattern-thickCotton, ivoryDot-polyester, jacquardPoly-spandex, jute, lamb-swool, leaf-twoColor, olefin_polyester-cotton, olefin-cotton-polyester, orange, peachPrints-polyester, pinkChiffon, plasticBag1, pureColorDyedBCFNylon-white, straw, straw2, tableCloth-cotton, teelStich-thinCotton, woodTexture-quilterFlannel, woolScarf and yellow-seersuckerCotton.

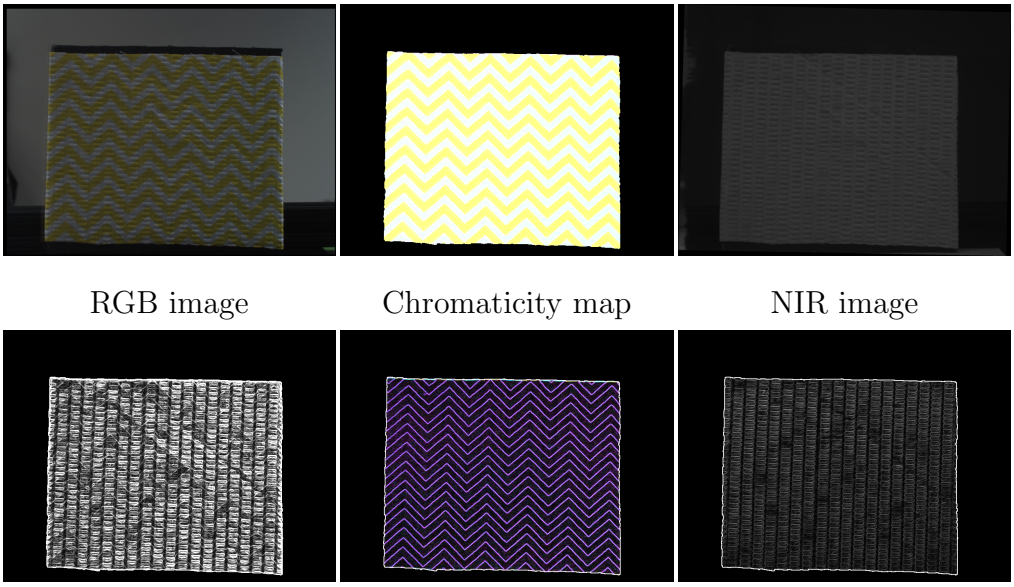


Figure 4.7: Example of the image gradient, the albedo map, the albedo gradient, the chromaticity map and the chromaticity gradient. The chromaticity gradient is employed to filter the image gradient, and then to recover the albedo gradient.

The corresponding pseudo-albedo edge is actually:

$$\nabla PA_k = \nabla(\frac{R_k}{R_{nir}}) = \frac{\nabla R_k R_{nir} - R_k \nabla R_{nir}}{R_{nir}^2}. \tag{4.7}$$

For most of the images in the NISAR dataset we were working on, the NIR image normally had a uniform albedo. Thus Equation 4.7 results in $\nabla PA_k = \frac{\nabla R_k R_{nir}}{R_{nir}^2}$. However, the texture of NIR could sometimes become complicated, such as that presented in Figure 4.9.

The chromaticity edge that was masked by thresholding the chromaticity gradient was first defined:

$$TC(x,y) = \begin{cases} 1, & ||\nabla CH|| > quantile(\nabla CH, 0.85), \\ 0, & \text{otherwise.} \end{cases} \tag{4.8}$$

where $quantile(\nabla CH, 0.85)$ is the 85% quantile number of ∇CH. The TC has non-zero magnitudes (value 1) at pixels where the chromaticity edge and the albedo edge are highly correlated. However, TC has zero magnitudes in regions where there should be no albedo edges, which indicates that no chromaticity edges exist.

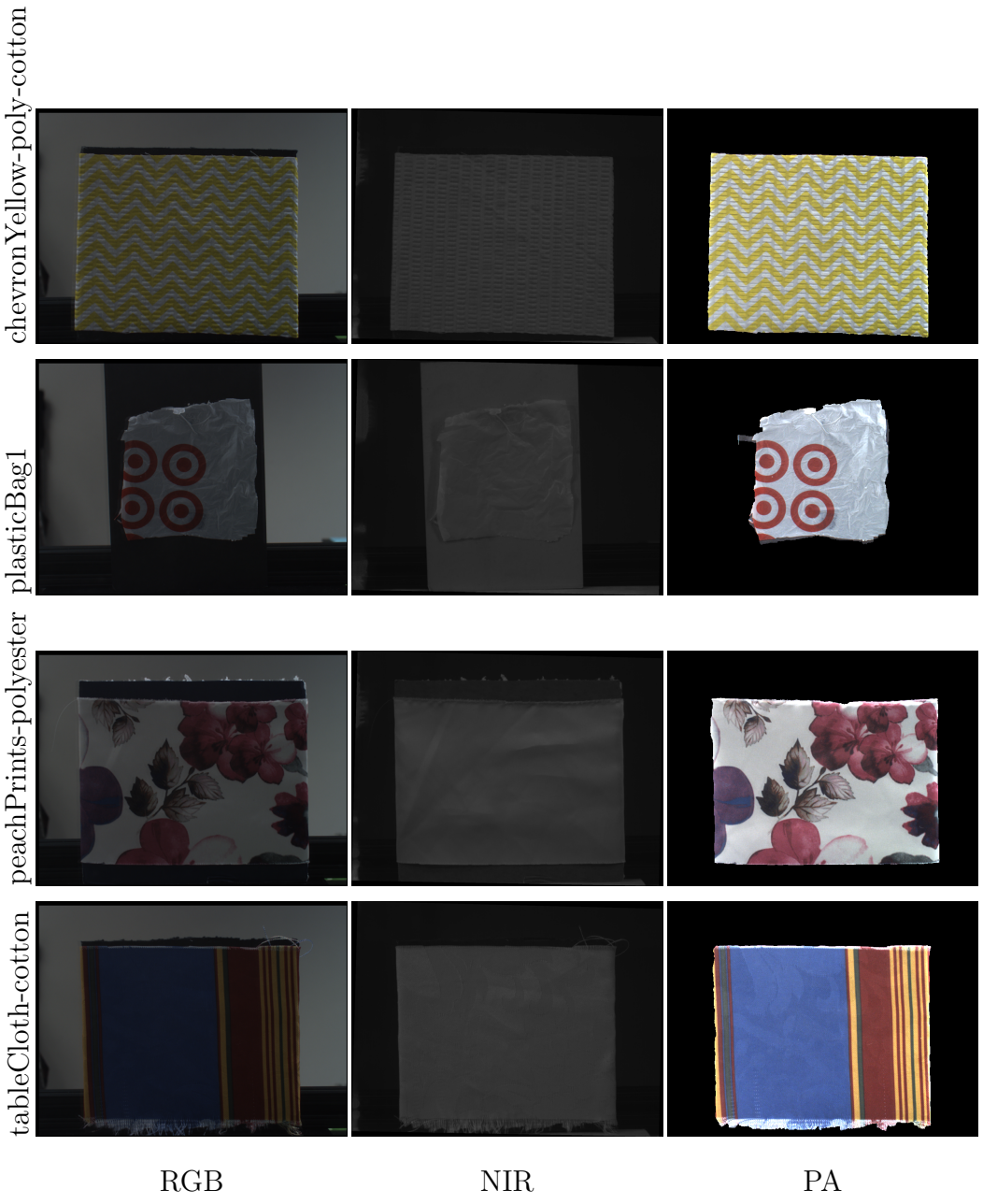
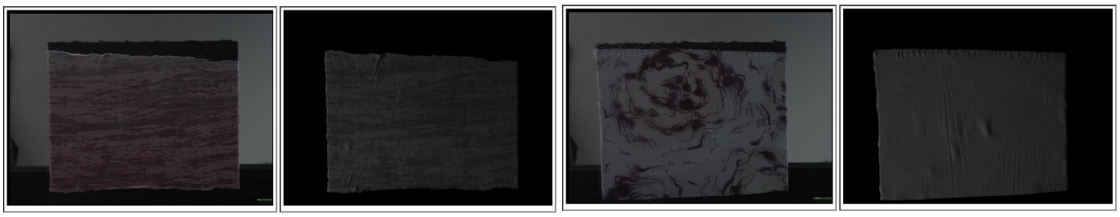
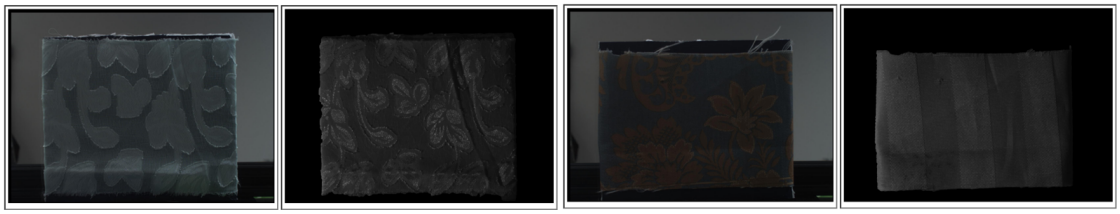


Figure 4.8: Examples of RGB and NIR image pairs with their pseudo-albedo (PA).



Similar texture appearance | Rich texture only in RGB



Rich texture only in NIR | Different textural appearance

Figure 4.9: Complicated textures in RGB and NIR images.

Then, a filtered pseudo-albedo edge map is defined by incoperating information on the chromaticity edge as :

$$TC \circ \nabla PA_k = TC \circ \nabla(\frac{I_k}{I_{nir}}) \tag{4.9}$$

$$= TC \circ \frac{\nabla R_k R_{nir} - R_k \nabla R_{nir}}{R_{nir}^2} \tag{4.10}$$

$$= TC \circ \frac{\nabla R_k}{R_{nir}} - TC \circ \frac{R_k \nabla R_{nir}}{R_{nir}^2} \tag{4.11}$$

where $\circ$ is pixel-wise multiplication. According to the definition and properties of TC, the term of $TC \circ \frac{R_k \nabla R_{nir}}{R_{nir}^2}$ in Equation 4.9 is actually zero in regions that contain no albedo-edges in the RGB image. Therefore, by filtering a pseudo-albedo edge using a chromaticity edge, a filtered pseudo-albedo edge that has approximately the same location is obtained as the albedo edge (see Figure 4.12). The filtered pseudo-albedo edge has approximately the same value as that of the albedo edge because of the constant ratio of the NIR albedo and because of the definition of the TC map.

4.3.4 The MAP Model

$\log(I)$, $\log(R)$, and $\log(L)$ are denoted as $\hat{I}$, $\hat{R}$, and $\hat{L}$ and are written as $\hat{I} = \hat{R} + \hat{L}$ in the *log* space. Without loss of generality, I, R, and L are still written for $\hat{I}$, $\hat{R}$, and $\hat{L}$, though they will be dealt with in *log* space from now on. Since an edge in

normal space is still an edge in *log* space, the above explanation will still be valid for the following discussion.

The proposed probabilistic model presents the intrinsic decomposition problem as a MAP formulation, $p(R_k, L|I_k, I_{nir})$, where k represents R, G, and B channels. By using Bayes' theorem:

$$p(R_k, L|I_k, I_{nir}) \propto p(I_k, I_{nir}|R_k, L)p(R_k)p(L), \tag{4.12}$$

where $p(R_k, L|I_k, I_{nir})$ represents the likelihood and $p(R_k)$ and $p(L)$ denotes the priors on the albedo and the shading layers, these terms will now be discussed separately. Since the operation proceeds independently for R, G, and B channels, the subindex k have been omitted for simplicity and the following discussion is valid for all three channels.

Likelihood $p(I|R, L)$. The likelihood of an observed image given that albedo and shading are based on the imaging model. According to the previous discussion, the pseudo-albedo (which is obtained from RGB and NIR images) is close to the real albedo layer except for a global constant ratio. It is assumed that this global constant ratio would be 1 in this chapter. By assuming a Gaussian distribution, the definition of the likelihood is

$$p(I, I_{nir}|R, L) \triangleq \prod_p N(PA_p - R_p|0, \sigma_0), \tag{4.13}$$

Albedo prior $p(R)$. Here, the albedo prior, $p(R)$, is defined to satisfy two objectives. The prior should serve as a regularization term that reduces the ill-posedness of the decomposition problem. Two components are introduced for $p(R)$: the global prior, $p_g(R)$, and the local prior, $p_l(R)$, *i.e.*,

$$p(R) \triangleq p_g(R)p_l(R). \tag{4.14}$$

Global albedo prior $p_g(R)$. Research indicates that for simple scenes composed of simple objects or a single object, the albedo layer satisfies the piece-wise constant assumption [WO04]. The albedo layer is piece-wise constant, and the albedo gradient is reasonably sparse but has a significantly large value [Bla85] compared to the shading layer, which is smooth everywhere and thus has a relatively small gradient. Therefore, the idea of ℓ_0-smoothing [XLXJ11] was adopted that was

originally introduced to create a piece-wise smooth artistic effect for image processing. Such an assumption provides the ℓ_0 prior for the statistics of an albedo image. The final definition of the global prior, $p_g(R)$, is written as:

$$p_g(R) \triangleq \prod_p \frac{1}{Z_p} e^{-||\nabla R_p||_0}, \tag{4.15}$$

Local albedo prior $p_l(R)$. A plausible representation of albedo is available in the proposed scenario. Thus, a pseudo-albedo image is used in this prior to constrain the gradients of the albedo image, which is significantly effective in suppressing artifacts from the shading edge. The prior is motivated by the fact that the proposed pseudo-albedo image provides very strong clues for the intrinsic albedo layer. Therefore, apart from using a general distribution of natural images, we propose applying the following regularization term: the input image gradient is constrained so that it is similar to the pseudo-albedo image gradient. The errors are defined to follow a Gaussian distribution with a 0z mean and standard deviation σ_1:

$$p_l(R) \triangleq \prod_p N(\nabla R_p - TC \cdot \nabla PA_p|0, \sigma_1), \tag{4.16}$$

The value of σ_1 is gradually increased over the course of optimization, as will be described more fully in Subsection 4.4, since this prior becomes less important as the albedo layer estimate becomes more accurate. The TC (Equation 4.8) is a chromaticity mask with the same size as the RGB and NIR images. TC has non-zero magnitudes (value 1) at pixels where the chromaticity edge and the albedo edge are highly correlated. TC has zero magnitudes, on the other hand, in regions where there should be no albedo edges, which is indicated by the lack of chromaticity edges.

Prior $p(L)$. Variations in the shading layer are relatively smaller than variations in albedo. It has been assumed that variations in the shading layer would approach to zero. Therefore:

$$\begin{aligned} p(L) &\triangleq \prod_p N(\nabla L_p|0, \sigma_2) N(\Delta L_p|0, \sigma_2) \\ &= \prod_p N(\nabla I_p - \nabla R_p|0, \sigma_2) N(\Delta I_p - \Delta R_p|0, \sigma_2), \end{aligned} \tag{4.17}$$

where σ_2 is the standard deviation. Based on Equations 4.6, 4.7 and the assump-

tion that R_{nir} is of value 1, $p(L)$ can be redefined as:

$$p(L) \triangleq \prod_{p} N(\nabla PA_p - \nabla R_p|0, \sigma_2) N(\Delta PA_p - \Delta R_p|0, \sigma_2), \tag{4.18}$$

In fact, according to the definition of PA, ∇R and ΔR actually should have value closer to ∇PA and ΔPA than to ∇I and ΔI, definition of 4.18 should be more meaningful than 4.17.

4.3.5 Optimization Model

The pipeline of the proposed method is outlined in Figure 4.10.

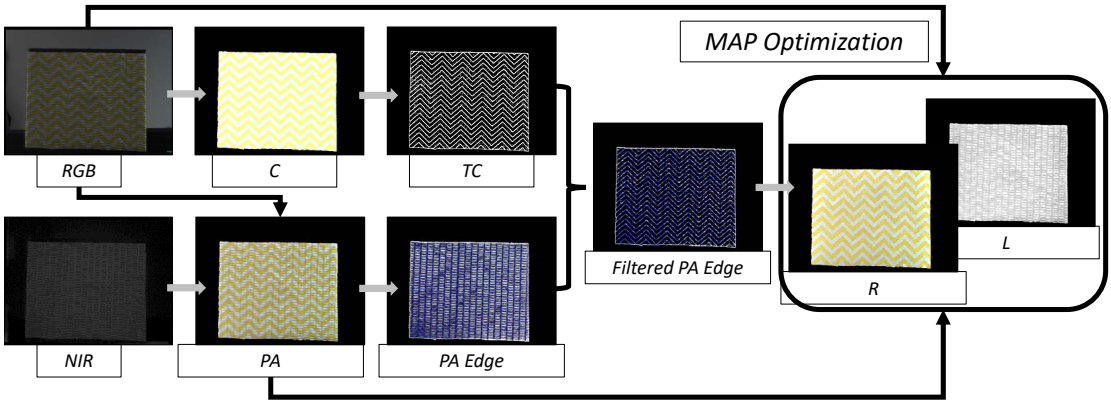


Figure 4.10: This figure outlines the pipeline of the proposed approach: 1) calculating the pseudo-albedo PA from input RGB and NIR image by using 4.6, 2) calculating chromaticity mask TC by using 4.8 3) calculating the filtered PA edge from PA and TC by using 4.9, and 4) combining PA, filtered PA edge, and input RGB image in the proposed MAP model to obtain decomposed albedo layer R and shading layer L.

Minimizing $p(I, I_{nir}|R, L)p(R)p(L)$ involves minimizing its *logarithm*, thus combining Equations 4.13 4.15 4.16 and 4.18 and replacing the coefficient related to standard deviation with λ_0, λ_1 and λ_2, the optimization model under the MAP framework 4.12 was built as follow: (Equation 4.19):

$$\begin{gathered} \min_{R,L} \left\{ ||\nabla R||_0 + \lambda_0 ||PA - R||_2 + \lambda_1 ||\nabla L||_2 + \lambda_2 ||\Delta L||_2 \right\} \\ \text{s.t.} \quad I_p \leq R_p \leq 0, \ max(L) = 0, \end{gathered} \tag{4.19}$$

The PA is assigned the pseudo-albedo defined in Equation 4.6, whereas the initialization of L is assigned to the NIR image according to the definition of PA

(Figure 4.8). Here $max(L) = 0$ is the largest value among all pixels. It comes from the white patch assumption for white-balancing and under the assumption in this work that the illumination of the scene is of white color.

The original images in the implementation in this work were normalized to $[1/256, 1]$. The R should fall in the range of $[I, 0]$ (note that the discussion here is within *log* space). $||\cdot||_2$ means the ℓ_2-norm, subscript p means the value at a pixel, and the *sum* operator runs over the whole image. In the proposed work, λ_0 is set to be 1 thus is omitted in following.

Further, TC (Equation 4.9) has been included in the proposed optimization problem as a guide to filter the effects of gradients that are not from the albedo edge:

$$\min_R \Big\{ ||\nabla R||_0 + ||PA - R||_2 + \lambda_1 ||TC \circ \nabla PA - \nabla R||_2 + \lambda_2 ||TC \circ \Delta PA - \Delta R||_2 \Big\}$$
$$\text{s.t.} \quad I_p \leq R_p \leq 0, \; max(L) = 0, \tag{4.20}$$

Noting that $L = I - R$, Equation 4.20 is written into pixel-wise form as:

$$\min_R \Big\{ C(\nabla R) + ||PA - R||_2 + (\lambda_1 (TC \circ \nabla PA - \nabla R)^2 + \lambda_2 ||TC \circ \Delta PA - \Delta R|| \Big\}$$
$$\text{s.t.} \quad I_i \leq R_i \leq 0. \tag{4.21}$$

where $C(I) = \#\{p || \partial_x I_p| + |\partial_y I_p| \neq 0\}$ indicates the number of non-zero elements and $\circ$ is the pixel-wise product. Here the constraint $max(L) = 0$ was removed as we white-balanced the input image using the white-patch assumption.

The following subsection explains how the optimization problem (Equation 4.21) was solved by using an iteration strategy called the half-quadratic separation scheme [WYYZ08].

4.3.6 Solver

The proposed objective function was non-convex due to the $C(I)$ component. The half-quadratic separation scheme [WYYZ08] was used to solve this problem. For this, auxiliary variables h_p^x and h_p^y were introduced to make the problem solvable

and we yielded the objective function:

$$\min_{R,(h^x,h^y)} \Big\{ C(h^x,h^y) + \sum_p ((PA_p - R_p)^2 + \beta((\partial_x R_p - h_p^x)^2 + (\partial_y R_p - h_p^y)^2) \\ +\lambda_1((TC \circ \partial_x PA_p - \partial_x R_p)^2 + (TC \circ \partial_y PA_p - \partial_y R_p)^2) + \lambda_2 ||TC \circ \Delta PA_p - \Delta R_p||) \Big\} \tag{4.22}$$

where $C(h^x,h^y) = \#\{p||h_p^x| + |h_p^y| \neq 0\}$ is the number of pixels with non-zero (h^x,h^y) value, β is a weight that is increased during the optimization (starting from around 10 and multiplied by η each time in our implementation, where η is used to accelerate the number of iterations and set to be between one and two). Equation 4.22 for fixed β can be minimized by alternating between computing R and (h^x,h^y), as is explained in the following.

Computing (h^x,h^y) Keeping R fixed, similarly to that in Xu *et al.* [XLXJ11], minimize the following problem:

$$\min_{(h^x,h^y)} \Big\{ C(h^x,h^y) + \sum_p (\beta((\partial_x R_p - h_p^x)^2 + (\partial_y R_p - h_p^y)^2) \Big\} \tag{4.23}$$

The closed-form solution for Equation 4.23 (proof can be found in Xu *et al.* [XLXJ11]) is:

$$(h^x,h^y) = \begin{cases} (\partial_x R_p, \partial_y R_p), & (\partial_x R_p)^2 + (\partial_x R_p)^2 > \frac{1}{\beta} \\ (0,0), & \text{otherwise.} \end{cases} \tag{4.24}$$

Computing R Fixing (h^x,h^y), the subproblem of Equation 4.22 w.r.t. R was solved. the information provided by the chromaticity map was simultaneously used as a filter by considering the relation between the chromaticity map and the albedo layer to avoid the influence of the gradients caused by the shading as much as possible. Therefore, the chromaticity gradient map was included as a guide by using TC in Equation 4.8. The subproblem of estimating R then corresponds to minimizing the problem:

$$\begin{aligned} E_R &= \Big\{ \textstyle\sum_p (\beta((\partial_x R_p - h_p^x)^2 + (\partial_y R_p - h_p^y)^2) \\ &+ \lambda_1((TC_p \partial_x PA_p - \partial_x R_p)^2 + (TC_p \partial_y PA_p - \partial_y R_p))^2 + \lambda_2 ||TC_p \Delta PA_p - \Delta R_p||) \Big\} \end{aligned} \tag{4.25}$$

Denote the Fourier transform operator as $\mathcal{F}$ and its inverse as $\mathcal{F}^{-1}$. Apply the

Fourier transform to all functions within the square terms and obtain

$$\begin{aligned} E_{\mathcal{F}(R)} &= \Big\{ \textstyle\sum_p (\beta((\mathcal{F}(\partial_x)\mathcal{F}(R_p) - \mathcal{F}(h_p^x))^2 + (\mathcal{F}(\partial_y)\mathcal{F}(R_p) - \mathcal{F}(h_p^y))^2) \\ &+ \lambda_1((\mathcal{F}(TC \circ \partial_x PA_p) - \mathcal{F}(\partial_x)\mathcal{F}(R_p))^2 + (\mathcal{F}(TC \circ \partial_y)\mathcal{F}(PA_p) - \mathcal{F}(\partial_y)\mathcal{F}(R_p)))^2 \\ &+ \lambda_2 ||\mathcal{F}(TC \circ \Delta PA_p) - \mathcal{F}(\Delta)\mathcal{F}(R_p)||) \Big\}. \end{aligned} \tag{4.26}$$

Similar to the deduction of Equation 2.16, Optimal R can be found by:

$$\begin{aligned} R &= \mathcal{F}^{-1}(A/B), \\ A &= \mathcal{F}(PA) + \beta(\overline{\mathcal{F}(\partial_x)}\mathcal{F}(h^x) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(h^y)) \\ &\quad +\lambda_1(\overline{\mathcal{F}(\partial_x)}\mathcal{F}(TC \circ \partial_x PA) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(TC \circ \partial_y PA)) \\ &\quad +\lambda_2\overline{\mathcal{F}(\Delta)}\mathcal{F}(TC \circ \Delta)\mathcal{F}(PA), \\ B &= \mathcal{F}(1) + (\beta + \lambda_1)(\overline{\mathcal{F}(\partial_x)}\mathcal{F}(\partial_x) + \overline{\mathcal{F}(\partial_y)}\mathcal{F}(\partial_y)) \\ &\quad +\lambda_2\overline{\mathcal{F}(\Delta)}\mathcal{F}(TC \circ \Delta), \end{aligned} \tag{4.27}$$

where $\mathcal{F}$ is the Fourier transform, $\circ$ is the pixel-wise product, and / is the pixel-wise division. Here, $\overline{\mathcal{F}(\cdot)}$ is the complex conjugate, $\mathcal{F}(1)$ is the Fourier transform of the Dirac delta function, and TC is defined as in Equation 4.26 where β is the same as in Equation 4.22. Parameter τ added to the denominator is a small number that was necessary to avoid the division by zero in the algorithm. The multiplication and division are both performed pixel-wise.

The whole process is outlined in Algorithm 3.

Algorithm 3 RGB-NIR Intrinsic Decomposition

Input:

input image I; smoothness weight λ_1, λ_2; initial β_0; iteration number $i_{\max}$; increasing rate η;

Initialization: $R \leftarrow PA$; $\beta \leftarrow \beta_0$; $i \leftarrow 0$.

repeat

with R fixed, computing (h^x, h^y) using Equation 4.24;

with (h^x, h^y), compute R using Equation 4.27;

$\beta = \eta * \beta, i++$;

until $i >= i_{\max}$

$L = I - R$;

Output: R and L;

4.4 Experiments

Experiments were carried to test the optimization speed of the proposed algorithm. In addition, comparisons with the state-of-the-art methods were shown to demonstrate the effectiveness of the proposed method. Evaluations were done on the estimated ground-truth data built from NISAR data [CNK16].

The experiments were done on a PC with an Intel i7-4820K CPU (3.7 GHz) and with 16GB of random-access memory (RAM). The implementation was done by using Matlab without any graphics processing unit (GPU) acceleration. The proposed algorithm was applied on the NISAR dataset of 100 materials [CNK16]. The size of all the images was reduced by 0.5 in the experiment here and resized them to 814×615.

Some results produced by the proposed method are listed in Figure 4.12. We can see that the proposed method is effective for materials with rich textures. It can be seen that the obtained albedo layer and shading layer look reasonable for the chevronYellow-poly-cotton, tableCloth-cotton, and plasticBag1 materials and are suitable for the priors that the two layers should have. However, it can be seen that the proposed method does not handle the peachPrints-polyester quite well.

The NIR image played an important role in the proposed optimization. To view its effects more intuitively, the results without an NIR image was generated by replacing the PA with the input RGB image, I_k $(k = r, g, b)$. It can be seen in Figure 4.11 that by using intensity together with the gradient and Laplacian information on PA (therefore, those of the RGB & NIR image pair), much better decomposition results could be obtained.

4.4.1 Comparisons

The performance of the proposed method was compared with several representative methods of intrinsic image estimation and reported the runtime per image as well as the means squared error (MSE) [GJAF09], local mean squared error (LMSE) [GJAF09], structural similarity index (SSIM) [WBSS04], which are summarized in Table 4.2. The methods in Shen and Yeo [SY11] and Gehler *et al.* [GRK+11] were built on complicated inference and their frameworks used more

constraints and learning-based techniques or spatial-spectral combination techniques for intrinsic decomposition. These methods are considered to be state-of-the-art regarding performance on the Massachusetts Institute of Technology (MIT) dataset. Work in Yu and Sato [2], which was also an edge-based method that employed a similar framework, was also compared. The source code of Gehler *et al.* [GRK+11] was taken from their website. The source code of Shen and Yeo [SY11] and Yu and Sato [2] were implemented with Matlab according to the descriptions reported in their corresponding research papers.

Table 4.1: Quantitative comparison with previous methods on 77 images with estimated shading GT. (m is minutes, and s is seconds.)

Method	**Runtime**	**SMSE**	**SLMSE**	**SSSIM**
Shen & Yeo [SY11]	$\sim$ 40 m	0.2695	0.0500	0.1367
Gehler et al. [GRK+11]	$\sim$ 10 m	0.0321	0.0029	0.8013
Yu & Sato [2]	$\sim$ 20 s	0.1303	0.0061	0.7174
Proposed (unique parameters)	$\sim$ 15 s	0.0120	0.0042	0.7582
Proposed (fine tuned parameters)	$\sim$ 15 s	**0.0101**	**0.0024**	**0.8308**

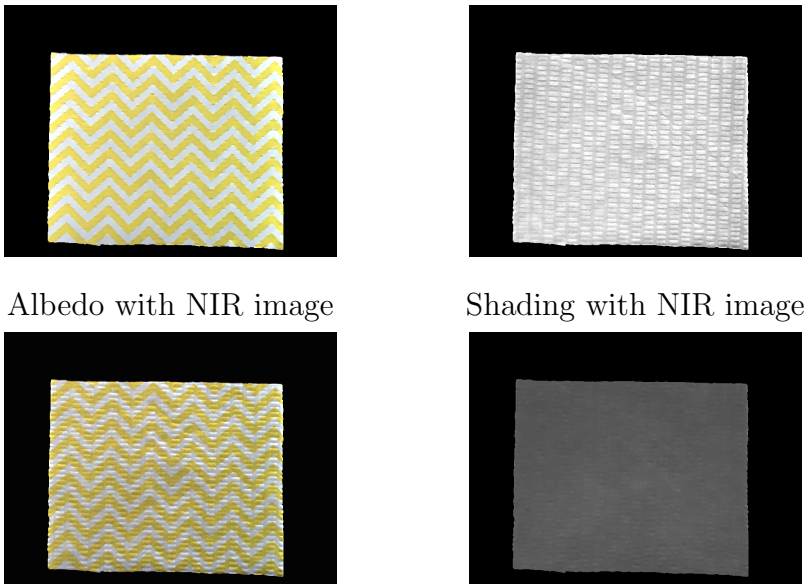
Table 4.4.1 compares the results from 77 images with the estimated shading GT and Table 4.4.1 lists the results from 27 selected images with the estimated albedo and shading GT. The results of the proposed method were listed by setting the proposed optimization with unique parameters for all images and by fine tuning all the parameters for each image to obtain better comparisons. From Tables 4.4.1 and 4.4.1, It can be seen that the proposed optimization was more efficient than the others. The proposed method could achieve performance close to methods employing complexity models (e.g., [GRK+11]), and ran much faster when using unique parameters for all images. The highest performance could even be achieved by fine-tuning all the parameters.

Of the three measurements MSE, LMSE, and SSIM that was used, It can be seen that the proposed method always achieved the best performance with MSE, which represents the global accuracy of comparison with GT. As can be seen from Figure 4.13, the results by Gehler *et al.* [GRK+11] on material chevronYellow-poly-cotton were actually not as accurate compared to the others. There are

Table 4.2: Quantitative comparison with previous methods on 27 selected images with estimated albedo & shading GT. (m is minutes, s is seconds.)

Method	**Runtime**	**RMSE**	**SMSE**	**RLMSE**	**SLMSE**	**RSSIM**	**SSSIM**
Shen & Yeo [SY11]	$\sim$ 40 m	0.0374	0.0847	0.0017	0.0028	0.8904	0.6711
Gehler et al. [GRK+11]	$\sim$ 10 m	0.0146	0.0337	**0.0045**	0.0028	**0.9453**	0.7993
Yu & Sato [2]	$\sim$ 20 s	0.0192	0.2186	0.0060	0.0056	0.5914	0.7109
Proposed(unique parameters)	$\sim$ 15 s	0.0118	0.0134	0.0101	0.0059	0.9034	0.7682
Proposed(fine tuned parameters)	$\sim$ 15 s	**0.0103**	**0.0115**	0.0051	**0.0026**	0.9434	**0.8182**

also a few similar examples of the results generated by Gehler *et al.* [GRK+11]. However, the LMSE and SSIM by Gehler *et al.* [GRK+11] are still higher than those of the others. One possible reason is that LMSE and SSIM represent the local accuracy of comparison more with the GT. They contain little information on the comparison of global accuracy. Another reason that might cause this phenomenon is that the obtained estimated GT, especially the estimated albedo GT, might not be accurate enough or serve as a real GT. We intend to explore the latter limitation in the future work to build a more accurate GT.



Albedo with NIR image | Shading with NIR image

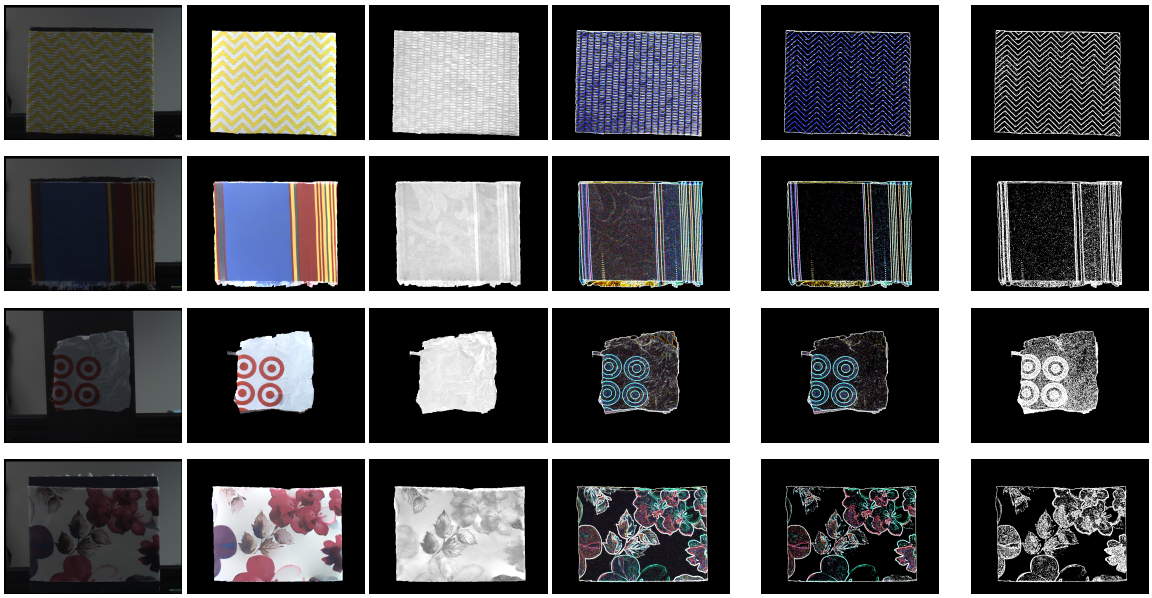
Albedo without NIR image | Shading without NIR image | chrom map

Figure 4.11: Results with or without NIR image.

4.4.2 Structure-Texture Separation via RGB-NIR Image Pairs

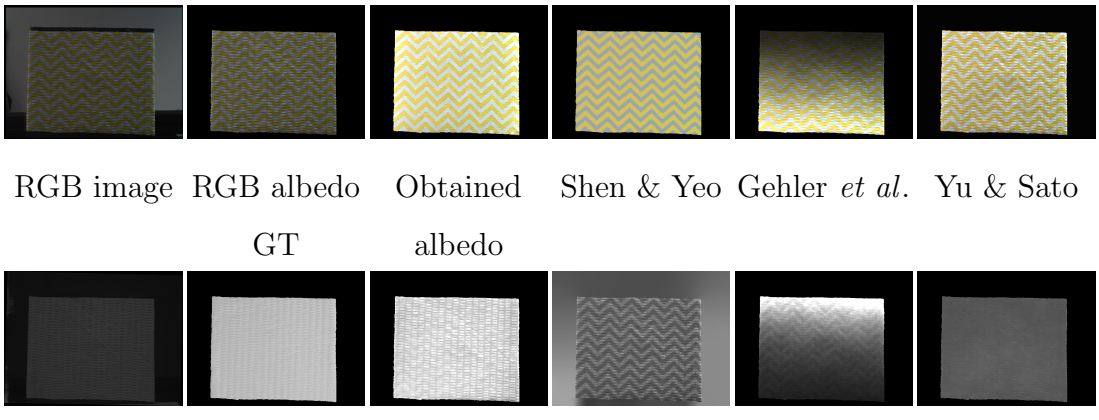
As it was discussed in subsection 4.3, the NIR image has a textureless appearance and its gradient is roughly the gradient of the shading layer under the assumption that NIR has a uniform albedo, where it has been assumed that the NIR albedo has a constant value of 1.

The mask generated from feature asymmetry (FA) was replaced in the WLS with the masks generated from the filtered pseudo-albedo gradient (Equation 4.8). According to the discussion above, it has been shown that the location of the



Input image | Obtained albedo | Obtained shading | PA edge | Filtered PA edge | Chrom map

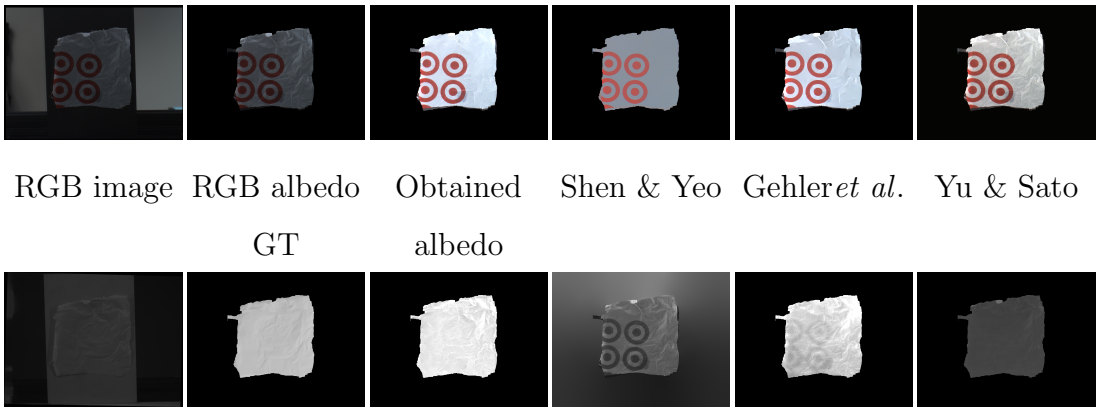
Figure 4.12: Results obtained from the proposed method. Top to bottom materials are: chevronYellow-poly-cotton, tableCloth-cotton, plasticBag1, and peachPrints-polyester.



RGB image | RGB albedo GT | Obtained albedo | Shen & Yeo | Gehler *et al.* | Yu & Sato

NIR image | Shading GT | Obtained shading | Shen & Yeo | Gehler*et al.* | Yu & Sato

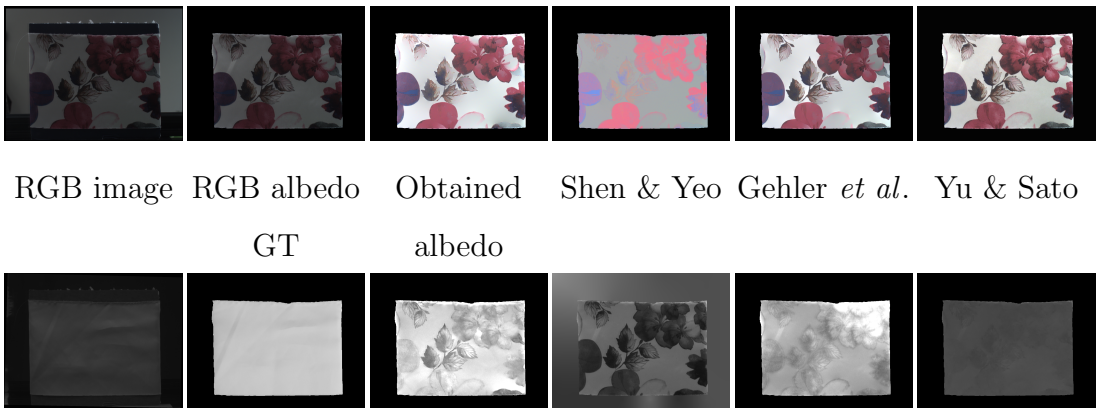
Figure 4.13: Comparisons of material chevronYellow-poly-cotton.



RGB image | RGB albedo GT | Obtained albedo | Shen & Yeo | Gehler *et al.* | Yu & Sato

NIR image | Shading GT | Obtained shading | Shen & Yeo | Gehler *et al.* | Yu & Sato

Figure 4.14: Comparisons of material plasticbag1.

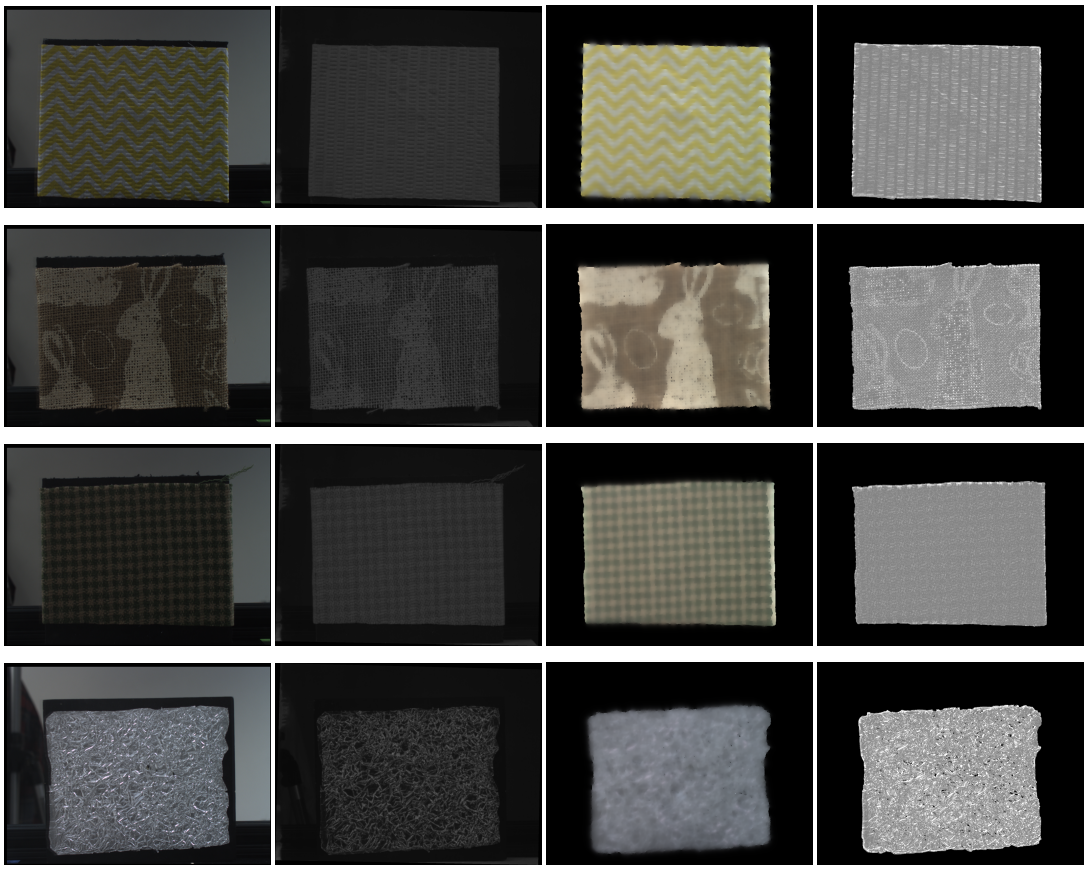


RGB image | RGB albedo GT | Obtained albedo | Shen & Yeo | Gehler *et al.* | Yu & Sato

NIR image | Shading GT | Obtained albedo | Shen & Yeo | Gehler *et al.* | Yu & Sato

Figure 4.15: Comparisons of material peachPrints-polyester.

albedo edge of an NIR image corresponds to the location of the intrinsic structure edge of the captured scene. Some results are shown in Figure 4.16. It can be seen that with the strong clue of the NIR image for the shading edge, a textureless structure layer could be obtained efficiently.



RGB image NIR image Structure layer Texture layer

Figure 4.16: Structure-texture separation using filtered pseudo-albedo gradient 4.8.

Chapter 5

Conclusion

This chapter provides a summary of the work presented in the previous three chapters in this dissertation. While each previous chapter had a self-contained summary and discussion, this chapter served to re-iterate these summaries and discussions. Here, potential directions for future work will also be described.

5.1 Summary

- Chapter 2 discussed the chromaticity map that is considered to be an important clue for the albedo layer in intrinsic image decomposition. We demonstrated the usefulness of the chromaticity gradient in the proposed work, which is an even stronger clue that indicates the albedo edge. Thus chromaticity gradient is important for the reconstruction of the albedo layer. A novel method of edge-based intrinsic image decomposition was proposed employing chromaticity and the Retinex assumption under the ℓ_0-norm based model. A fast solver using FFT and inverse FFT was designed. The proposed method ran much faster than the state-of-the-art methods with comparable performance on the MIT dataset [GJAF09] and the dataset provided by Bousseau [BPD09].

  One limitation of the chromaticity gradient is that there is ambiguity between strong variations of reflectance (*i.e.*. monochromatic variations between black, gray, and white) and variations in shadow, as can be seen from Figure 2.9. The proposed strategy could handle the limitations that have

been explained in subsection 2.3 to some extent. However, we will try to find extra clues in the future to distinguish variations in reflectance and variations due to shadow.

- Chapter 3 proposed a method for addressing the problem of structure preserving image smoothing by incorporating an FA-based edge map as weights into the WLS optimization framework. It was demonstrated that the proposed method offers at-least comparable qualitative results to those of the state-of-the-art methods but with significantly reduced computational cost. A major shortcoming of the proposed method is that there may occasionally be missing structure edge information in the FA edge map, which thus results in artifacts and color bleeding. Also, the proposed method is not sufficiently effective for low-contrast contours, which is the same as that for other methods. Future research will be in the direction of adopting multiscale strategies to solve these problems.

- Chapter 4 proposed the use of an NIR image together with an RGB image that was aimed at obtaining an efficient intrinsic image decomposition algorithm. To be more specific, the properties of an NIR image to certain materials was applied: the profile of intensity for the NIR image had less texture than that of an RGB image. The relationship between an albedo edge and chromaticity edge, between that shading edge and NIR image based was analyzed on the gradient operations. A pseudo-albedo map was defined that had almost the same properties as the albedo layer with the assumption made in this chapter. With these observations, a MAP framework was adopted and the intrinsic decomposition problem was modeled as a probability problem. Further, MAP optimization with Plancheral's theorem was solved by using a method of half-quadratic splitting. Significant acceleration was achieved by efficiently solving two split sub-quadratic problems. The proposed method outperformed state-of-the-art methods and achieved much lower computational complexity.

5.2 Future Directions

- **Fast Sparse Edge-based Intrinsic Image Decomposition Guided by Chromaticity Gradient**

  The proposed method could be accelerated by optimizing our code using a parallel computing technique and by other means for it be able to run in real time. It is possible to convert the proposed method into real-time to improve computational efficiency with a more efficient architecture and with parallel computing techniques such as GPU-acceleration. It is also possible to include temporal information and make the proposed method capable of real-time intrinsic video decomposition.

  The proposed method is currently effective, especially for scenes with a single object or objects with simple textures. I would like to combine scene priors in the future, such as the lighting sources of scenes and the depth priors or categories of scenes into the proposed framework so that it can deal with more complicated scenes.

  Finally, the correlation between the albedo edge and chromaticity edge I analyzed here can be extended to be adopted in other fields of computer vision and image processing such as problems with photometric-stereo, materials analysis, color constancy, specularity removal, and shadow removal to enable these issues to be considered from another point of view.

- **Structure-Texture Separation via Feature Asymmetry-aware Weighted Least Square (FAWLS)**

  The computation of the FA map is currently 2D and the optimization WLS is 1D recursive iterations. It would also be interesting to transfer the computation of FA map into 1D recursive iterations to achieve higher computational efficiency.

  The location of FA could be shifted to a location of a few pixels (one three pixels) in contrast to its corresponding gradient edge, which limits its accuracy when being used as a structure edge indicator. Therefore, it would be useful and necessary work in the future to improve the location indicating properties of such phase-based features or to align such features with their

corresponding gradient edges.

- **Intrinsic Image Decomposition for Materials with Fine-Textures via RGB-NIR Images**

  Although the proposed method is currently limited to the analysis of scenes with a single object or objects with simple textures, it would be interesting to extend it to deal with more complicated scenes. The proposed framework would be able to deal with more complicated scenes by combining scene priors such as the lighting sources of scenes, depth priors, or the categories of scenes. The running time could be reduced by using parallel computing techniques or GPU-acceleration to make the proposed framework approach that of real-time processing.

  Although the assumption on the NIR image that it contains less texture than the RGB image can be observed in some materials such as the ones presented in Choe *et al.*'s dataset [CNK16], it has not widely been observed in general cases, *e.g.* in natural scenes. In fact, some research directions: *i.e.*, image denoising, image dehazing, image deblurring, and image enhancing for low light images have even assumed that the NIR image contains richer texture and higher contrast than the RGB image. Therefore, it would be interesting in future work to seek more intelligent clues to the NIR image for intrinsic image decomposition. One possible direction would be to look at the local correlation of the NIR image and shading layer in small regions, in which case the used assumption might still be valid.

  Another assumption used in this work was that the NIR albedo is almost uniform everywhere. Although this assumption is valid for almost all the images in the dataset [CNK16], in real life this assumption may not always be valid, even for a single object. Therefore, extending the proposed method to deal with a more general assumption of the NIR albedo would be an interesting future direction of research.

  Under the textureless assumption of the NIR image, the defined pseudo-albedo (PA) map has almost shading-free properties, which would be very useful to solve many computer vision problems such as object detection,

object classification and photometric stereo. It would be interesting to see how PA could be adopted in these domains and what promising results might be produced.

Bibliography

[AGCO06] Jean-François Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.

[BA87] Peter J. Burt and Edward H. Adelson. Readings in computer vision: Issues, problems, principles, and paradigms. chapter The Laplacian Pyramid As a Compact Image Code, pages 671–679. Morgan Kaufmann Publishers Inc., 1987.

[BBML11] Ahror Belaid, Djamal Boukerroui, Yves Maingourd, and Jean-François Lerallut. Phase-based level set segmentation of ultrasound images. *IEEE Transactions on Information Technology in Biomedicine*, 15(1):138–147, 2011.

[BBS14] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, 33(4):159:1–159:12, 2014.

[BCM05] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 60–65, 2005.

[Bel16] Ahror Belaid. *Feature Asymmetry of the Conformal Monogenic Signal*, pages 12–20. Springer International Publishing, 2016.

[Bla85] Andrew Blake. Boundary conditions for lightness computation in mondrian world. *Computer Vision, Graphics, and Image Processing*, 32(3):314–327, 1985.

[BLMV10] A. Buades, T. M. Le, J. M. Morel, and L. A. Vese. Fast cartoon + texture image filters. *IEEE Transactions on Image Processing (TIP)*, 19(8):1978–1986, 2010.

[BM12] Jonathan T Barron and Jitendra Malik. Color constancy, intrinsic images, and shape estimation. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 57–70, 2012.

[BM13] Jonathan T. Barron and Jitendra Malik. Intrinsic scene properties from a single rgb-d image. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17–24, 2013.

[BM15] Jonathan T. Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 37(8):1670–1687, 2015.

[BNB04] Djamal Boukerroui, J. Alison Noble, and Michael Brady. On the choice of band-pass quadrature filters. *Journal of Mathematical Imaging and Vision*, 21(1-2):53–80, 2004.

[BPD09] A. Bousseau, S. Paris, and F. Durand. User-assisted intrinsic images. *ACM Transactions on Graphics (TOG)*, 28(5):130:1–130:10, 2009.

[Bra00] R.N. Bracewell. *The Fourier Transform and Its Applications*. Electrical engineering series. McGraw Hill, 2000.

[BS11] M. Brown and S. Susstrunk. Multi-spectral sift for scene category recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 177–184, 2011.

[BT78] H. G. Barrow and J. M. Tenenbaum. *Recovering Intrinsic Scene Characteristics from Images*. Academic Press, 1978.

[CCFI14] J Chang, R. Cabezas, and J. W. Fisher III. Bayesian nonparametric intrinsic image decomposition. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 704–719, 2014.

[CK13] Qifeng Chen and Vladlen Koltun. A simple model for intrinsic image decomposition with depth cues. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 241–248, 2013.

[CNK16] Gyeongmin Choe, Srinivas G. Narasimhan, and In So Kweon. Simultaneous estimation of near ir brdf and fine-scale surface geometry. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2452–2460, 2016.

[CPTK14] Gyeongmin Choe, Jaesik Park, Yu-Wing Tai, and In So Kweon. Exploiting shading cues in kinect ir images for geometry refinement. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3922–3929, 2014.

[DD02] Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics (TOG)*, 21(3):257–266, 2002.

[DRC+15] Sylvain Duchêne, Clément Riant, Gaurav Chaurasia, Jorge Lopez-Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. Multiview intrinsic images of outdoors scenes with an application to relighting. *ACM Transactions on Graphics (TOG)*, 34(5):164, 2015.

[Dre03] Mark S. Drew. Recovery of chromaticity image free from shadows via illumination invariance. In *Proceedings of IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 32–39, 2003.

[DWL98] M. S. Drew, Jie Wei, and Ze-Nian Li. Illumination-invariant color object recognition via compressed chromaticity histograms of

color-channel-normalized images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 533–540, 1998.

[ESA12] Gamal ElMasry, Da-Wen Sun, and Paul Allen. Near-infrared hyperspectral imaging for predicting colour, ph and tenderness of fresh beef. *Journal of Food Engineering*, 110(1):127 – 140, 2012.

[FDL04] Graham D. Finlayson, Mark S. Drew, and Cheng Lu. Intrinsic images by entropy minimization. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 582–595, 2004.

[FFL10] Zeev Farbman, Raanan Fattal, and Dani Lischinski. Diffusion maps for edge-aware image editing. *ACM Transactions on Graphics (TOG)*, 29(6):145:1–145:10, 2010.

[FFLS08] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics (TOG)*, 27(3):67:1–67:10, 2008.

[FHLD06] G.D. Finlayson, S.D. Hordley, Cheng Lu, and M.S. Drew. On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(1):59–68, Jan 2006.

[FS01] Michael Felsberg and Gerald Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, 2001.

[FS04] M. Felsberg and G. Sommer. The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and Vision*, 21(1):5–26, 2004.

[FS09] Clement Fredembach and Sabine Susstrunk. Illuminant estimation and detection using near-infrared. In *Proceedings of SPIE/IST Electronic Imaging*, page 10.1117/12.806431, 2009.

[GJAF09] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 2335–2342, 2009.

[GO11] Eduardo S. L. Gastal and Manuel M. Oliveira. Domain transform for edge-aware image and video processing. *ACM Transactions on Graphics (TOG)*, 30(4):69:1–69:12, 2011.

[GRK+11] P. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf. Recovering intrinsic images with a global sparsity prior on reflectance. In *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*, pages 765–773, 2011.

[HCG14] Sk. Mohammadul Haque, Avishek Chatterjee, and Venu Madhav Govindu. High quality photometric reconstruction from a depth camera. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2283–2290, 2014.

[Hor70] B. K.P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, 1970.

[Hor74] B. K.P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4):277–299, 1974.

[HRRR11] MC Hanumantharaju, M Ravishankar, DR Rameshbabu, and S Ramachandran. Color image enhancement using multiscale retinex with modified color restoration technique. In *Proceedings of International Conference on Emerging Applications of Information Technology (EAIT)*, pages 93–97, 2011.

[JCTL14] Junho Jeon, Sunghyun Cho, Xin Tong, and Seungyong Lee. Intrinsic image decomposition using structure-texture separation and surface normals. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 218–233, 2014.

[KEE13] Levent Karacan, Erkut Erdem, and Aykut Erdem. Structure-preserving image smoothing via region covariances. *ACM Transactions on Graphics (TOG)*, 32(6):176:1–176:11, 2013.

[KL12] Johannes Kopf and Dani Lischinski. Digital reconstruction of halftoned color comics. *ACM Transactions on Graphics (TOG)*, 31(6):140:1–140:10, 2012.

[Kov97] Peter Kovesi. Symmetry and asymmetry from local phase. In *Tenth Australian Joint Converence on Artificial Intelligence*, pages 2–4, 1997.

[KSSC14] C. Kerl, M. Souiai, J. Sturm, and D. Cremers. Towards illumination-invariant 3d reconstruction using tof rgb-d cameras. In *Proceedings of International Conference on 3D Vision (3DV)*, pages 39–46, 2014.

[LBD13] Pierre-Yves Laffont, Adrien Bousseau, and George Drettakis. Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19(2):210 – 224, 2013.

[LM71] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America A (JOSA A)*, 61(1):1–11, 1971.

[MCB+12] Jos V. Manjn, Pierrick Coup, Antonio Buades, D. Louis Collins, and Montserrat Robles. New methods for mri denoising based on sparseness and self-similarity. *Medical Image Analysis*, 16(1):18–27, 2012.

[MCL+14] Dongbo Min, Sunghwan Choi, Jiangbo Lu, Bumsub Ham, Kwanghoon Sohn, and Minh N. Do. Fast global image smoothing based on weighted least squares. *IEEE Transactions on Image Processing (TIP)*, 23(12):5638–5653, 12 2014.

[Mey01] Yves Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*. American Mathematical Society, 2001.

[MLKS04] Yasuyuki Matsushita, Stephen Lin, Sing Bing Kang, and Heung-Yeung Shum. Estimating intrinsic images from image sequences with biased illumination. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 274–286, 2004.

[MO87] M. C. Morrone and R. A. Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6(5):303–313, 1987.

[MPN98] Miguel Mulet-Parada and J. Alison Noble. 2d+t acoustic boundary detection in echocardiography. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 806–813, 1998.

[MRBO86] M. C. Morrone, J. Ross, D. C. Burr, and R. Owens. Mach bands are phase dependent. *Nature*, 324(6094):250–253, 1986.

[MSH14] Takuya Mikami, Daisuke Sugimura, and Takayuki Hamamoto. Capturing color and near-infrared images with different exposure times for image enhancement under extremely low-light scene. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 669–673, 2014.

[NMY15] Takuya Narihira, Michael Maire, and Stella X. Yu. Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 2992–2992, 2015.

[PM90] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 12(7):629–639, 1990.

[PP11] Giuseppe Papari and Nicolai Petkov. Edge and line oriented contour detection: State of the art. *Image Vision Computation*, 29(2-3):79–103, 2011.

[QY13] Li Xu Shaojie Zhuo Xiaopeng Zhang Liang Shen Jiaya Jia Qiong Yan, Xiaoyong Shen. Cross-field joint image restoration via scale map. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 1537–1544, 2013.

[RFS14] Dominic Rufenacht, Clement Fredembach, and Sabine Susstrunk. Automatic and accurate shadow detection using near-infrared information. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(8):1672–1678, 2014.

[ROF92] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

[SBD15] Evan Shelhamer, Jonathan T. Barron, and Trevor Darrell. Scene intrinsics and depth from a single image. In *Proceedings of IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 235–242, 2015.

[Sch] K.J. Schlager. Non-invasive near infrared measurement of blood analyte concentrations. US Patent 4,882,492, 1989.

[SFS09a] Neda Salamati, Clment Fredembach, and Sabine Ssstrunk. Material Classification Using Color and NIR Images. In *Proceedings of IS&T/SID 17th Color Imaging Conference (CIC)*, pages 11–13, 2009.

[SFS09b] Lex Schaul, Clément Fredembach, and Sabine Süsstrunk. Color image dehazing using the near-infrared. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 1609–1612, 2009.

[SJA08] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. *ACM Transactions on Graphics (TOG)*, 27(3):73:1–73:10, 2008.

[SLCS12] Neda Salamati, Diane Larlus, Gabriela Csurka, and Sabine Susstrunk. Semantic Image Segmentation Using Visible and Near-Infrared Channels. In *Lecture Notes in Computer Science*, volume 7584, pages 461–471. Springer Berlin Heidelberg, 2012.

[SMHL00] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Skin detection in video under changing illumination conditions. In *Proceedings of International Conference on Pattern Recognition (ICPR)*, volume 1, pages 839–842 vol.1, 2000.

[SPBV12] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell. Names and shades of color for intrinsic image estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 278–285, 2012.

[SSD09] Kartic Subr, Cyril Soler, and Frédo Durand. Edge-preserving multiscale image decomposition based on local extrema. *ACM Transactions on Graphics (TOG)*, 28(5):147:1–147:9, 2009.

[STL08] L. Shen, P. Tan, and S. Lin. Intrinsic image decomposition with non-local texture cues. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, 2008.

[SY11] L. Shen and C. Yeo. Intrinsic images decomposition using a local and global sparse representation of reflectance. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 697–704, 2011.

[SYJL11] J. Shen, X. Yang, Y. Jia, and X. Li. Intrinsic images using optimization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3481–3487, 2011.

[Tan08] Robby T Tan. Visibility in bad weather from a single image. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[TFA05] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(9):1459–1472, 2005.

[TM98] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 839–846, 1998.

[TNI04a] R. T. Tan, K. Nishino, and K. Ikeuchi. Separating reflection components based on chromaticity and noise analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 26(10):1373–1379, 2004.

[TNI04b] Robby T. Tan, Ko Nishino, and Katsushi Ikeuchi. Color constancy through inverse-intensity chromaticity space. *Journal of the Optical Society of America A (JOSA A)*, 21(3):321–334, 2004.

[VRT13] Vibhav Vineet, Carsten Rother, and Philip H. S. Torr. Higher order priors for joint intrinsic image, objects, and attributes estimation. In *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*, pages 557–565, 2013.

[WBSS04] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004.

[Wei01] Y. Weiss. Deriving intrinsic images from image sequences. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 68–75, 2001.

[Wit84] Andrew P. Witkin. Scale-space filtering: A new approach to multi-scale description. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 150–153, 1984.

[WO04] Michael Werman and Ido Omer. Color lines: Image specific color representation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 946–953, 2004.

[WYYZ08] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[XLnGhMq09] Jie Xiong, Han Li-na, Geng Guo-hua, and Zhou Ming-quan. Real color image enhanced by illumination-reflectance model and wavelet transform. In *Proceedings of International Forum on Information Technology and Applications (IFITA)*, pages 691–695, 2009.

[XLXJ11] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image smoothing via l0 gradient minimization. *ACM Transactions on Graphics (TOG)*, 30(6):174:1–174:12, 2011.

[XYXJ12] Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia. Structure extraction from texture via relative total variation. *ACM Transactions on Graphics (TOG)*, 31(6):139:1–139:10, 2012.

[YR10] P WT Yuen and M Richardson. An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition. *The Imaging Science Journal*, 58(5):241–253, 2010.

[YS15a] J. Yu and Y. Sato. Fast sparse edge-based intrinsic image decomposition guided by chromaticity gradients. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 3753–3757, 2015.

[YS15b] J. Yu and Y. Sato. Structure-preserving image smoothing via phase congruency-aware weighted least square. In *Proceedings of Pacific Graphics Short Papers*, page 10.2312/pg.20151274, 2015.

[Yu16] J. Yu. Rank-constrained pca for intrinsic images decomposition. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 3578–3582, 2016.

[ZHZ14] Y. Zang, H. Huang, and L. Zhang. Efficient structure-aware image smoothingby local extrema on space-filling curve. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 20(9):1253–1265, 2014.

[ZKE15] Tinghui Zhou, Philipp Krahenbuhl, and Alexei A. Efros. Learning data-driven reflectance priors for intrinsic image decomposition. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 3469–3477, 2015.

[ZSM08] Xiaopeng Zhang, Terence Sim, and Xiaoping Miao. Enhancing photographs with near infra-red images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

Publications

[1] J. Yu and Y. Sato "Structure-Preserving Image Smoothing via Phase Congruency-aware Weighted Least Square". In *Proceedings of Pacific Graphics Short Papers, 10.2312/pg.20151274, 2015.*

[2] J. Yu, Y. Sato "Fast sparse edge-based intrinsic image decomposition guided by chromaticity gradients". In *Proceedings of IEEE International Conference on Image Processing (ICIP), pages 3753-3757, 2015.*

[3] J. Yu, C. Lu, Y. Sato "Sparsity-based Color Quantization with Preserved Image Details". In *Proceedings of SIGGRAPH Asia 2014 Posters, pages 32:1–32:1, 2014.*

[4] J. Yu, Y. Sato "Painterly Style Stroke Map based on Chromaticity and Blue Noise Sampling for Non-Photorealistic Rendering". *Poster Presentation of Meeting on Image Recognition and Understanding (MIRU), 2015.*

[5] J. Yu, M. Constable, J. Wang, K. L. Chan, M. S. Brown. "Aesthetic Interactive Hue Manipulation for Natural Scene Images". In *Proceedings of Pacific-Rim Symposium on Image and Video Technology (PSIVT), pages 201-214, 2015.*

[6] J. Yu, PA Heng, W. Wang, J. Qin "A Novel Coarse-to-fine Level Set Framework for Ultrasound Image Segmentation". In *Proceedings of International Conference on Artificial Intelligence and Pattern Recognition (AIPR), pages 8-17, 2015.*

[7] M. Wei, J. Yu, W. M. Pang, J. Wang, J. Qin, L. Liu, P. A. Heng "Bi-normal filtering for mesh denoising". *IEEE Transactions on Visualization and Computer Graphics (TVCG) 21 (1), pages 43-55, 2015.*