# Color Spill Suppression in Chroma Keying

**by**

**Ya Luo**

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the M.A.Sc. degree in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

**Abstract**

Alpha matting is one of the key techniques in image processing and is used to extract accurate foreground from a still image or video sequences. Chroma keying is a special case of alpha matting with a solid background color.

Color spill is one of the difficulties in chroma keying, and it has not been effectively solved by current methods. Sometimes, an image contains both reflected regions and transparent regions. When the foreground in such images is chroma keyed, reflection on the foreground is often falsely treated as transparency and causes unreal foreground extraction and composition. This problem is called **color spill**. Color spill suppression aims to extract the opaque foreground with the correct transparency descriptor (i.e. $\alpha$ value) and remove the reflected background color on it. When the background color presented on the foreground is simultaneously caused by reflection and transparency, color spill suppression becomes extremely challenging. It is because that the reflection removal and the actual transparency estimation is a dilemma. Our proposed method for color spill suppression is to separate reflected regions from transparent regions, and process reflected regions as foreground while keeping transparency unchanged at the same time.

In this thesis, we propose a novel method for color spill suppression for chroma keying. The quality of the estimated alpha matte could be significantly improved. In our approach, we suppress color spill by using the polarization and the optical flow algorithm based on disparity estimation. Specifically, we make the assumption that reflection changes more than transparency when the scene is captured by a binocular camera with a polaroid filter. Based on this assumption, we took stereo images with polarization filter, registered stereo images by optical flow and conducted the variance

analysis on histograms of input images to separate transparency and reflection. Our experiments show that the opaque foreground with background color spill can be reliably extracted while the real transparency can be kept.

**Acknowledgements**

First and foremost, I would like to express my heartfelt thanks to my supervisor, Professor Jiying Zhao, for bringing the problem of color spill suppression in chroma keying to me, and also thank for his patience, guidance, encouragement and feedbacks throughout my entire research work.

Moreover, I would like to express my gratitude to my friends and lab colleagues for their help, advice and companion.

Last but not least, I would like to thank my parents for their love and encouragement. They are always my greatest support in my study and life.

This thesis is dedicated to my family.

# Table of contents

# List of tables

# List of figures

# Nomenclature

## Abbreviations

| | |
|---|---|
| RGB | Red, Green, Blue (color model space) |
| HSV | Hue, Saturation, Lightness (color model space) |
| MAP | Maximum A Posteriori |
| KNN | K Nearest Neighbor |
| CIE | International Commission on Illumination |
| LPP | Local Preserving Projects |
| ICA | Independent Components Analysis |
| SVD | Singular Value Decomposition |
| AE | Adobe After Effects CS6 |

# Chapter 1

# Introduction

## 1.1 Chroma keying and alpha matting

With the fast development of science and technology, various methods have been proposed for information collecting and processing. Image, one of widely used media in our life, has always been compelling for its large amount of information, fast speed, and long-distance transmission. Image segmentation is an important and basic topic in image processing. In the field of image segmentation, the matting problem is a special case that needs our attention.

In image editing and film industry, matting is commonly used to accurately extract foreground from a still image or video sequences. Generally, foreground can be considered as the extracted object, while the rest of the frame or image can be considered as the background. Afterward, the background can be replaced by another one to composite a new image with the extracted foreground. In 1984, Poter and Duff [1] proposed the landmark paper to model the way that foreground and background

color mix. Generally, it can be expressed as the equation shown as follows:

$$I_{(x,y)} = \alpha_{(x,y)} F_{(x,y)} + (1 - \alpha_{(x,y)}) B_{(x,y)} \,, \tag{1.1}$$

where $x$ and $y$ indicate the coordinate components of the pixel; $F$ is the color of foreground; $B$ is the color of background; $I$ is the color of the observed pixel. The corresponding value $\alpha$ is the blending factor varying from 0 to 1. If $\alpha = 0$, the pixel is completely background and if $\alpha = 1$, the pixel is completely foreground. In Equ. (1.1), there are three unknown variables $F$, $B$ and $\alpha$. It is impossible to solve the equation just with one known value ($I$) directly. Various matting methods are proposed to find an optimal $\alpha$ with constraints and cost functions. With appropriate matting model, the foreground objects can be accurately extracted from the image. An example of alpha matting is shown in Figure 1.1. In the alpha matte (b) of this figure, the black indicates the background and the white indicates the foreground.

(a) Original image (b) Alpha matte

**Figure 1.1:** An example of alpha matting.

To solve the matting equation, there are several constraints. One of the constraints is called trimap. Trimap is a rough segmentation of a given image. As shown in Figure

1.2 (a), the given image is divided into foreground, background and unknown areas, where black region is referred to as background, white region is referred to as absolute foreground and gray region is referred to as unknown region.

Another commonly used constraint is called strokes. Parts of the given image are labeled as the foreground/background using the strokes, and the rest of the image is regarded as the unknown area. The example of strokes is shown in Figure 1.2 (b), where black lines indicate the background, white lines indicate the foreground and the rest part of the image is the unknown region.

(a) Trimap (b) Strokes

**Figure 1.2:** Side information used in alpha matting.

In matting methods, trimap and strokes are always drawn by users. With a good trimap or strokes, the result of alpha matting can be improved. Otherwise, the matting quality could be significantly degraded by using bad trimap/strokes. When we are trying to extract the foreground from complicated background in a still image, the matting problem is also referred to as alpha matting, which often involves manual operations. When we are trying to efficiently extract the foreground from a video sequence, automation is an essential concern.

Chroma keying is a special case of alpha matting, which aims to fast and efficiently extract foreground by using solid background color, such as green or blue. And the

process of extracting foreground object is called 'keying'. This technique is widely used in the weather forecast and the film production. For example, the actor appears to be standing in the outer space while he or she is actually filmed in the studio in front of a big green or blue screen as the background.

Since alpha matting has no requirement for background, it is often used in still image like photos, while chroma keying with a solid background is often used in videos for its efficiency and automation like film shooting. A professional studio environment for chroma keying is shown in Figure 1.3.

**Figure 1.3:** A professional studio environment for chroma keying [2].

However, there are still limitations in chroma keying systems. For example in Figure 1.4, the background light reflects on the woman's face and coat, and these reflections will be falsely treated as transparency in regular chroma keying systems. If it is composed of a new background like a natural background, we can see the background through her face and coat since the regions are processed as transparency, which is not consistent with common sense. Similarly, as shown in Figure 1.5, the green color on the windows is due to transparency while the green tint on the back

chair is due to reflection. When conducting chroma keying, reflections on the window and the chair are treated as transparency, which is unreasonable. This problem is called color spill. Color spill in chroma keying has not been effectively solved by current methods. And there is little literature about color spill in chroma keying. Our proposed method aims to suppress color spill, which can also be expressed as the separation of reflection and transparency. An effective color spill suppression method is supposed to improve the quality of alpha matte with higher accuracy.

(a) Original image (b) Alpha matte (c) Color spill in Chroma keying with a new background

**Figure 1.4:** Color spill in Woman image when conducting chroma keying.

(a) Original image (b) Alpha matte (c) Color spill in Chroma keying with extracted foreground

**Figure 1.5:** Color spill in Car image when conducting chroma keying.

## 1.2 Polarization

In order to suppress color spill, we propose to use the polaroid filters on images.

### 1.2.1 Introduction to polarization

Polarization refers to the fact that the direction of oscillation is perpendicular to the direction of the propagation of waves. It is an important cue to distinguish the transverse waves from the longitudinal waves. Longitudinal waves are waves in which the displacement of the medium is in the same direction as, or the opposite direction to the direction of wave transferred. Transverse waves are moving waves, and they are made up of oscillations which are occurring perpendicularly to the propagation [3]. Light is a kind of transverse electromagnetic wave. The polarization of light was first introduced by Malus in 1980.

If the oscillation of light is confined to one direction, the light is referred to as linearly polarized light. Meanwhile, natural light such as sunshine is unpolarized light, since the light consists of many wave trains with random directions of oscillation. Partially polarized light consists of all the possible directions of oscillation and its orthogonal components have the maximum and minimum amplitudes respectively. The intensity of partially polarized light can be weakened by passing through a polaroid, while the linearly polarized light can be eliminated. Also, the natural light can be expressed as a group of linearly polarized light and partially polarized light.

The unpolarized light can be turned into linearly polarized light by using a sheet of material called polaroid. As shown in Figure 1.6, the light from a luminous source is unpolarized, and it may be a group of polarized light from all directions. Then

the light is passing through a sheet of polaroid filter, and only a component which is oscillated in a specified direction is allowed to pass through the polaroid. If the linearly polarized light is passing through another sheet of polaroid filter, all the light may be eliminated. In our homemade environment for chroma keying, the light source is a kind of unpolarized light.

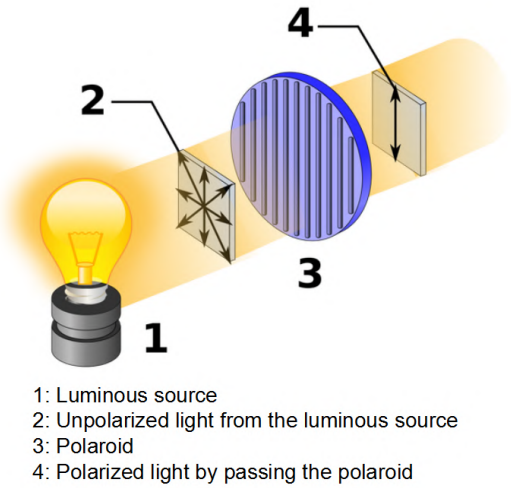



**Figure 1.6:** Transform from unpolarized light to polarized light [3].

### 1.2.2 Malus's law

Malus's law was proposed by Malus in 1808. Malus' law indicates the intensity of linearly polarized light passing through a sheet of polaroid, and is defined as follows [4]:

$$I = I_o(\cos\theta)^2\,, \tag{1.2}$$

where $I_o$ is the intensity of linearly incident polarized light; $\theta$ is supposed to be the

angle between the oscillated direction of incident light and the polarized direction of the polaroid. When $\theta = 0°$ or $\theta = 180°$, the intensity of the received light is maximum. When $\theta = 90°$, the light is eliminated by the polaroid.

### 1.2.3 Physical properties of reflection and Brewster's angle

When an unpolarized light is passing through an incident plane, the reflected light and refracted light are partially polarized in general. The reflected and refracted light will be linearly polarized only at a specific incident angle $\theta_b$, which is also called Brewster's angle [5]. And the reflected light is perpendicular to the refracted light at the Brewster's angle. Then incident light is supposed to pass through a flat surface, such as the glass. According to Fresnel equations [6], we can get the reflection coefficients $r_\parallel$ and $r_\perp$, which refer to the direction of polarization lying in the incident plane and perpendicular to the incident plane respectively. And the equations are defined as follows:

$$r_\perp = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2}\,, \tag{1.3}$$

$$r_\parallel = \frac{n_1 \cos\theta_2 - n_2 \cos\theta_1}{n_1 \cos\theta_2 + n_2 \cos\theta_1}\,, \tag{1.4}$$

where $n_1$ is the refractive index for air; $n_2$ is the refractive index for glass; $\theta_1$ is the incident angle and $\theta_2$ is the angle of refraction. Here $\theta_2$ is calculated according to the

Snell's law of refraction [7], and the equation is defined as follows:

$$\frac{\sin\theta_2}{\sin\theta_1} = \frac{n_1}{n_2}\,. \tag{1.5}$$

And the reflected light can be expressed as a combination of two orthogonal components $R_\parallel$ and $R_\perp$. They are defined by the equations shown as follows:

$$R_\parallel = r_\parallel^2\,, \tag{1.6}$$

$$R_\perp = r_\perp^2\,. \tag{1.7}$$

Figure 1.7 is an example of Brewster's law. The incident light is unpolarized with the Brewster's angle $\theta_p$. The reflected light and refracted light are partially polarized. And the reflected light can be expressed as the combination of two components: parallel to the incident plane $R_\parallel$ and perpendicular to the incident plane $R_\perp$. Based on the definition of the Brewster's angle, the reflected light is linearly polarized which leads to $R_\parallel = 0$. According to Equ. (1.4) and Equ. (1.6), we can get the definition of Brewster's angle expressed as follows:

$$\tan\theta_p = \frac{n_2}{n_1}\,. \tag{1.8}$$

Meanwhile, the reflected light and refracted light are perpendicular to each other when the incident angle is Brewster's angle $\theta_p$, which is around $56^\circ$ for glass and air. It can be clearly observed in Figure 1.8 that the intensity of reflection varies with

the angle of incidence. When most of the light passes directly without reflection and refraction through materials like glass, it is called transparency. The transparency is a part of the original light source, which is unpolarized and has little change through polarization. The property that reflection varies more than transparency through polarization is used in our proposed method.

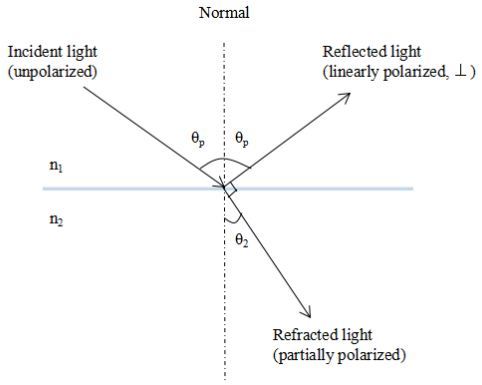



**Figure 1.7:** An example of Brewster's law.

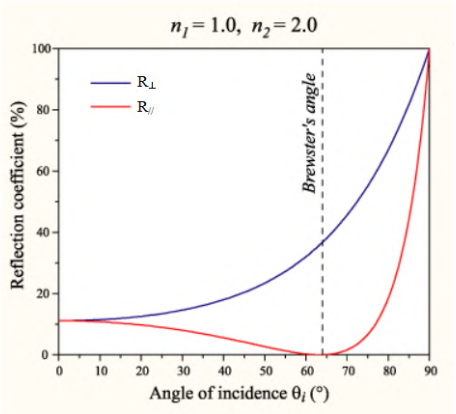



**Figure 1.8:** The property of Brewster's angle [8].

### 1.2.4 Specular reflection and diffuse reflection

Reflection from a surface can be divided into specular reflection and diffuse reflection. Specular reflection refers to reflection followed the law of reflection that incident angle equals the reflected angle. While diffuse reflection refers to the group of reflected lights with random reflected angle, which is happened inside or on the surface of an object. As shown in Figure 1.9, (a) is specular reflection and (b) is diffuse reflection.

Specular reflection has the property of polarization, and it can be distinguished with transparency by polarization. While diffuse reflection is unpolarized, stereo images can be used to help the separation in this thesis.

### 1.2.5 Application of polarization

Besides the application of separating reflection on glass, polarization can be used in many fields in the daily life. For example, polaroids is used on a special pair of glass for 3D movies. The oscillated directions of polaroids on glasses are perpendicular to each other. By passing through the polaroids on glasses, each eye will have a specific view with a great three-dimensional sense for the audience. And photographers always use polarization to adjust the brightness of the photo.

## 1.3 Stereo images

Stereo images are used to help separate reflection and transparency when the reflection is not specular reflection. In our proposed method, we use a binocular camera to take a pair of stereo images from the left and the right views simultaneously. The

example of a pair of stereo images is shown in Figure 1.10. Stereo images simulate the way that human perceives objects from both eyes. Stereo matching is often used to generate three-dimensional depth from two-dimensional images. In our proposed method, stereo matching based on optical flow algorithm is used for color spill suppression.

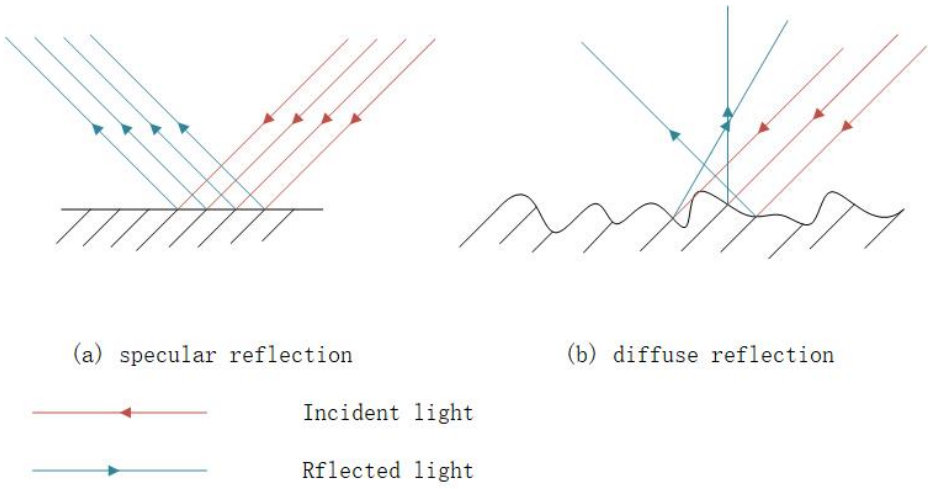



**Figure 1.9:** Specular and diffuse reflection.

(a) an image from the left eye (b) an image from the right eye

**Figure 1.10:** An example of a pair of stereo images.

## 1.4 Contributions

This thesis propose a novel method for color spill suppression in chroma keying, which can significantly improve the quality of the estimated alpha matte. The main contributions are summarized as follows:

1. We propose to use polarization to discriminate the specular reflection and the transparency.

2. We propose to use a stereo imaging system [9] to discriminate the diffuse reflection and transparency.

3. A color spill suppression system is established to improve the quality of the original alpha matte.

## 1.5 Thesis structure

In this thesis, we focus on color spill suppression to improve the quality of alpha matte. And it consists of five chapters.

In Chapter 1, the basic concepts of alpha matting, chroma keying, polarization, and stereo images are introduced. The main contributions of the thesis are summarized.

In Chapter 2, we review some representative algorithms on blue screen matting and alpha matting. Since the color spill suppression is highly relevant to transparency and reflection separation, we review the classical methods of reflection separation on glass in image analysis.

In Chapter 3, we present our proposed method in detail.

In Chapter 4, we show the experimental results by using our method and conduct visual comparison with other state-of-the-art color spill suppression methods to demonstrate the higher accuracy and better quality of our generated alpha matte.

In Chapter 5, the conclusion is drawn and some future work is discussed.

# Chapter 2

# Literature review

Our research work is to propose a novel method to solve color spill problems in chroma keying. In this chapter, we are going to review the matting methods for both blue screen and natural images. Then we talk about some classical methods for separating reflections and transparency in image.

## 2.1 Blue screen matting

Blue screen matting is a special case of alpha matting, which aims to solve the matting problem with a solid backing color, such as blue. The term "solid" means the chrominance of the background color is constant and the lightness may change within a narrow range. Therefore, the blue screen matting method is easier to extract the foreground from a solid background than the normal alpha matting. In the film industry, the blue screen matting is often called as **chroma keying**. The term "keying" means the process of extracting the background screen color and composing the fore-

ground with a new background. And we do not differentiate the blue/green screen matting and chroma keying in our thesis.

### 2.1.1 Difference matting

Some films and videos have been recorded by using blue screen matting for many years. However, the patents of these methods are often protected [10][11][12][13]. In this section, we are going to introduce a representative technique which is called the difference matting. This matting method was proposed by Petro Vlahos in 1964 [10]. The realization was based an equipment called Ultimatte Corporation which is similar to the video studio we use today. The difference matting method mainly focuses on the difference between the blue channel and the green channel on the foreground objects. And the modeled object color is shown in Equ. (2.1):

$$C_0 = C_f + (1 - \alpha_0)C_b \,, \tag{2.1}$$

where $C_0$ represents the observed color; $C_f$ and $C_b$ represent the foreground color and background color respectively. However, the coefficient $\alpha$ is not the completely same as the matting equation we used tody.

Based on this equation, Vlahos made the assumption in paper [10] that the blue channel of the foreground object is related to the green channel by Equ. (2.2):

$$B_0 \leq aG_0 \,, \tag{2.2}$$

where $B_0$ indicates the value in the blue channel of the foreground object and $G_0$ is the green channel. The range of constant $a$ is within [0.5, 1.5] [13].

According to Vlahos's earliest patent [11], the value of $\alpha$ can be estimated by using Equ. (2.3):

$$\alpha_0 = max\left(min\left(1 - a_1\left(min\left(B_0, B_k\right) - a_2 G_0\right), 1\right), 0\right) , \tag{2.3}$$

where $B_0$ and $G_0$ are the values of blue and green channel; $a_i$ ($i$=1,2) is tunable parameters within the range of [0.5, 1.5]. And we find out the minimum value of the background color in the blue channel as $B_k$. With the estimated value of $\alpha$, the foreground color can be obtained as shown in Equ. (2.4):

$$\begin{aligned} R_f &= R_f \\ G_f &= G_f \\ B_f &= min\left(B_f, a_2 G_f\right) \end{aligned} , \tag{2.4}$$

where $R_f$, $G_f$,and $B_f$ refer to the red, green and blue components of the foreground color respectively. To improve the result, Vlahos proposed the method [13] to obtain the estimation of $\alpha$ by using Equ. (2.5):

$$\alpha_0 = 1 - a_1\left(B_0 - a_2\left(a_5 max\left(a_3 R_0, a_4 G_0\right) + \left(1 - a_5\right) min\left(a_3 R_0, a_4 G_0\right)\right)\right) , \tag{2.5}$$

where $a_i$ ($i$=1,2,3,4,5,...) are the tunable parameters; $R_0$, $G_0$, $B_0$ are the values of red, green and blue channel of the observed pixel. The estimated value of $\alpha$ can be obtained by tuning the parameters $a$. And with $\alpha$, the foreground objects can be extracted from the film by using difference matting method. This approach was the beginning and could be considered as the basic ideal of some blue screen matting techniques today. However, the blue screen matting can only be used on a constant

background color. To extract foreground objects from natural images, we have to use natural image matting methods which are introduced in the next section.

## 2.2 Natural image matting

Matting refers to the problem of extracting accurate foreground from a still image or video sequences. Generally, foreground can be considered as the extracted object, while the rest of the frame or image can be considered as the background. The extracted foreground can be composted with another background to generate a new image or frame. Matting is one of the key techniques in image processing and editing, and is widely used in real life, such as the film industry. As we introduced in Section 1.1, matting problems can be expressed by Equ. (1.1) with color vectors $F(x, y)$, $B(x, y)$ and a blending factor $\alpha(x, y)$. And this problem can also be presented as Equ. (2.6) with the matrix form:

$$\begin{bmatrix} I_R \\ I_G \\ I_B \end{bmatrix} = \alpha \begin{bmatrix} F_R \\ F_G \\ F_B \end{bmatrix} + (1 - \alpha) \begin{bmatrix} B_R \\ B_G \\ B_B \end{bmatrix}, \tag{2.6}$$

where $R$, $G$ and $B$ indicate the three channels in RGB color space. As it is shown in Equ. (2.6), by given a single input image, we have to solve the equation with 7 unknown variables from 3 known values $I_R$, $I_G$ and $I_B$. Therefore, there is no unique solution to this problem. To obtain a good estimate of the blending factor $\alpha$, most matting approaches use constraints to the problem. The constraints can be categorized as user guidance, such as a manual trimap on the natural image and prior

assumptions on image statistics.

In this thesis, we conduct the literature review based on natural image matting. In the following sections, several methods with different prior assumptions on image statistics to the matting problem are reviewed.

Depending on the applications, the matting problems can also be classified by other means, such as video matting [14, 15], shadow matting [16, 17], and environment matting [18, 19].

### 2.2.1 Sampling-based matting

Since there are 7 unknown variables in Equ. (2.6), we can do some prior works to estimate the foreground color and background color for each pixel. Then the corresponding value $\alpha$ can be obtained by using the given pixel color, estimated foreground color, and estimated background color. In the sampling-based method, samples are chosen in neighbors, and the strong correlation between nearby image pixels are used to estimate the background and foreground color. Moreover, the key problem of the sampling-based method is how to choose reliable samples and estimate the foreground/background color. Generally, nearby pixels always have similar colors and correlation in a single image. And this local correlation can be widely used in image processing, such as segmentation and super-resolution. In this section, we are going to review three representatives for early sampling-based methods. And since these methods are used on natural image processing, a roughly trimap is manually drafted by users to indicate the foreground, background, and unknown regions.

In 2000, an early sampling-based method for matting problem was proposed by Ruzon and Tomasi [20]. The Ruzon method is based on the assumption that unknown

region is a narrow band along the foreground boundary and it can be represented by a chain of pixels. With the given pixel in the unknown region, foreground/background color and corresponding value $\alpha$ for this pixel can be estimated by using the Ruzon method [20]. In this approach, the samples are collected by using a rectangular window, as shown in Figure 2.1.

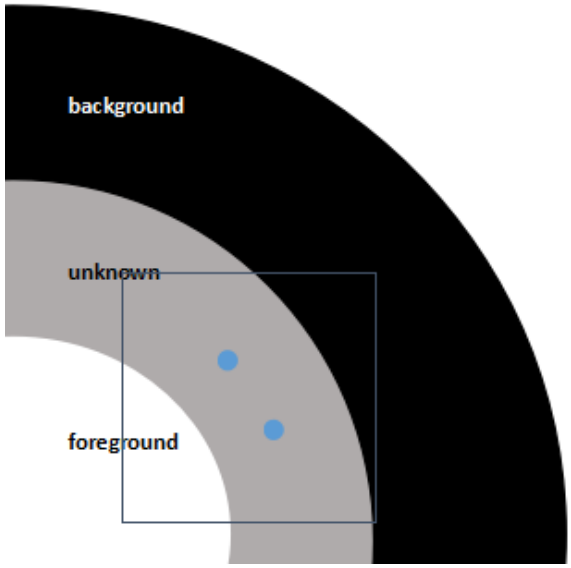



**Figure 2.1:** Sampling strategy of Ruzon's method [20].

In Figure 2.1, the black region and white region indicate background and foreground respectively, while the gray region indicates the unknown region within which the blue dots indicate the given pixels. The window covers pixels in the foreground, the background ,and the unknown region. The color of pixels in each window are modeled by using Gaussian distribution functions in CIE-Lab color space [21]. The distribution of observed color can be expressed as the linear combination of the foreground and background Gaussian distributions with the blending factor $\alpha$. When the intermediate distribution for which the observed color has maximum probability, the

optimal $\alpha$ value is estimated. Then the foreground color and background color of the unknown pixel can be estimated by the mean value of the foreground and background Gaussian pairs.

Bayesian matting method was proposed by Chuang in 2001 [22]. This approach was based on Ruzon's matting method to improve the effect. In the Bayesian matting method, a continuously sliding window is used to collect samples, as shown in Figure 2.2.

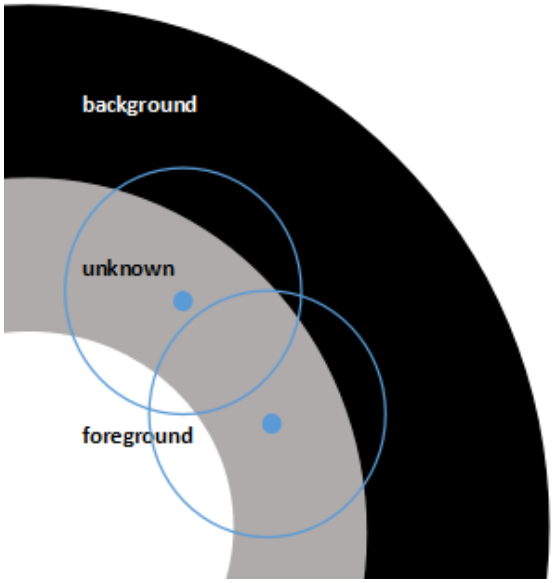



**Figure 2.2:** Sampling strategy of Bayesian matting method [22].

Pixels in the window are also modeled by using Gaussian distribution functions, which is similar to the Ruzon's method. Then a well-defined Bayesian framework is used to solve the problem by using maximum a posteriori (MAP) technique. Mathematically, the unknown foreground value, background value, and corresponding $\alpha$ value for the given pixel can be estimated by using the MAP technique [22] as shown

in Equ. (2.7):

$$\begin{aligned}
&\underset{F,B,\alpha}{\operatorname{argmax}} P\left(F,B,\alpha|C\right) \\
&= \underset{F,B,\alpha}{\arg\max} P(C|F,B,\alpha)\, P\left(F\right) P\left(B\right) P\left(\alpha\right)/P\left(C\right) \quad , \\
&= \underset{F,B,\alpha}{\arg\max}\, L\left(C|F,B,\alpha\right) + L\left(F\right) + L\left(B\right) + L\left(\alpha\right)
\end{aligned} \tag{2.7}$$

where $C$ is the given pixel color; $F$, $B$, and $\alpha$ indicate the foreground, background and blending factor respectively; $L\left(\bullet\right)$ indicates the log likelihood of $P\left(\bullet\right)$, which means $L\left(\bullet\right) = \log P\left(\bullet\right)$; $L\left(C|F,B,\alpha\right)$ in Equ. (2.7) is defined in Equ. (2.8):

$$L\left(C|F,B,\alpha\right) = -\|C - \alpha F - (1-\alpha)B\|^2 \big/ \sigma_C^2 \,, \tag{2.8}$$

where $\sigma$ indicates the local color variance. And $L\left(F\right)$, $L\left(B\right)$ are expressed in Equ. (2.9) and Equ. (2.10):

$$L\left(F\right) = -\left(F - \bar{F}\right)^T \sum\nolimits_F^{-1} \left(F - \bar{F}\right) \Big/ 2 \,, \tag{2.9}$$

$$L\left(B\right) = -\left(B - \bar{B}\right)^T \sum\nolimits_B^{-1} \left(B - \bar{B}\right) \Big/ 2 \,, \tag{2.10}$$

where $\bar{F}$ and $\bar{B}$ indicate the mean value of the foreground and background respectively, as shown in Equ. (2.11) and Equ. (2.12):

$$\bar{F} = \frac{1}{W} \sum_{i \in N} w_i F_i \,, \tag{2.11}$$

$$\bar{B} = \frac{1}{W} \sum_{i \in N} w_i B_i \,. \tag{2.12}$$

Since the window is constant, $P(C)$ in Equ. (2.7) can be dropped. And we can obtain the $\alpha$ value which is defined in Equ. (2.13) by maximizing the posterior probability:

$$\alpha = \frac{(C - B) \cdot (F - B)}{\|F - B\|^2} \,. \tag{2.13}$$

As shown in Figure 2.3, the image is divided into three regions by trimap, and the blue dot is the given pixel in the unknown region. The foreground color, background color, and corresponding value for this pixel could be estimated by using the Knockout method. Knockout method [23, 24] choose samples in neighbors. And in this approach, the foreground color for the given pixel can be estimated by using a weighted sum of nearby known pixel colors in the foreground region. And the weight factors used in this method are proportional to the distance between known pixels and chosen samples. Similarly, the background color can also be estimated in this way. Then the corresponding factor $\alpha$ in each channel can be estimated by Equ. (2.14):

$$\alpha_i = \frac{C_i - B_i}{F_i - B_i} \,. \tag{2.14}$$

In conclusion, sampling-based methods are used to estimate the corresponding

factor by collecting nearby foreground and background pixels as samples. However, the methods can only work well when the image color is smooth and the trimap is well-defined, since most of the samples are collected along the boundaries of trimap.

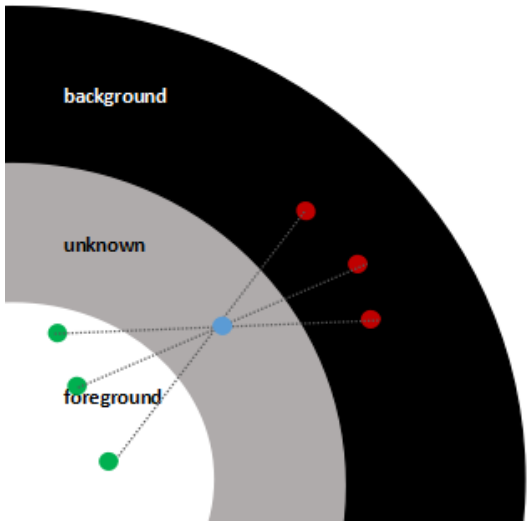



**Figure 2.3:** Sampling strategy of Knockout matting method [23, 24].

### 2.2.2 Propagation-based matting

Propagation based natural image matting methods are proposed [25–27] to avoid the limitation of color sampling-based methods. Propagation-based methods use different affinities between neighboring pixels to estimate the corresponding factor $\alpha$. In this section, we will review some classic propagation-based methods.

The Poisson matting method makes the assumption that changes in the foreground and background are locally smooth [28]. The alpha matte gradient is proportional to the gradient of the original image. And the matte gradient can be approximated in Equ. (2.15):

$$\nabla \alpha_{(x,y)} \simeq \frac{1}{F_{(x,y)} - B_{(x,y)}} \nabla I_{(x,y)} \,, \tag{2.15}$$

where $(x, y)$ indicates the coordinate of the image; $\nabla$ is the gradient operator; $I$ indicates the image color for the pixel $(x, y)$; $F$ and $B$ are the foreground color and background color, which are the nearest neighbor of the unknown pixel. And Equ. (2.15) can be further expressed in Equ. (2.16):

$$\Delta \alpha_{(x,y)} \simeq div \left( \frac{\nabla I}{F_{(x,y)} - B_{(x,y)}} \right) \,, \tag{2.16}$$

where $\Delta$ indicates the Laplacian operator; $div$ indicates the divergence operator. This equation can be solved by using the Gauss-Seidel iteration with the over-relaxation method. When doing Poisson matting on textured images, since the pixels are collected in a small neighborhood and the local smoothness can not be guaranteed based on the assumption, the result may be inaccurate. Therefore, the improvement based on the Poisson matting was proposed [29]. It uses a set of local filters and operations to refine and improve the quality of the results, such as high-pass filter, diffusion filter and boosting brush.

To avoid the limitation in the Poisson matting approach when the local foreground and background color are not smooth, the closed-form matting method was proposed [30]. It makes the underlying assumption that the pixel color in local foreground and background can be considered as the linear combination of two constant colors which is also referred to as the line color model space. In the closed-form matting, the value

of $\alpha$ can be estimated by Equ. (2.17) based on the assumption:

$$\alpha_i = \sum_c a^c I_i^c + b^c, \forall i \in w\,, \tag{2.17}$$

where $i$ indicates the coordinate of the image; $c$ refers to the three channels in RGB color space; $a$ and $b$ are two constants in the small local window $w$. Moreover, in the closed-form matting method, a local cost function is introduced to find the optimal constant $a$ and $b$ in the window. The cost function is defined in Equ. (2.18) to minimize the energy:

$$J(\alpha, a, b) = \sum_{j \in I} \left( \sum_{i \in w_j} \left(\alpha_i - a_j^c I_i^c - b_j^c\right)^2 + \varepsilon \left(a_j^c\right)^2 \right), \tag{2.18}$$

where $w_j$ is the local window around the pixel $j$; $\varepsilon\left(a_j^c\right)^2$ is a regularization term on $a$ which is used for numerical stability and minimization; the two constants $a$ and $b$ can be eliminated by using the derivations in [30]. And Equ. (2.18) can be expressed as a quadratic cost function with $\alpha$ value as in Equ. (2.19):

$$J(\alpha) = \alpha^T L \alpha\,, \tag{2.19}$$

where $L$ is an $N$ by $N$ matrix; $\alpha$ in this equation refers to an $N$ by 1 vector, and $N$ is the number of pixels in the unknown region. In the matrix $L$ which is also called

matting Laplacian, the $(i, j)$th entry can be formulated by Equ. (2.20):

$$L(i,j) = \sum_{k|(i,j)\in w_k} \left[ \delta_{ij} - \frac{1}{|w_k|} \left( 1 + (I_i - \mu_k) \left( \sum{}_k + \frac{\varepsilon}{|w_k|} I_3 \right)^{-1} (I_j - \mu_k) \right) \right], \quad (2.20)$$

where $\delta_{ij}$ is defined in Equ. (2.21) as the Kronecker delta; $|w_k|$ indicates the number of pixels in the local window $w_k$; $u_k$ indicates the mean value of the pixel colors in the window $w_k$; $\sum_k$ is a 3 by 3 covariance matrix which is also refers to as the pixel colors in the window, and $I_3$ is a 3 by 3 identity matrix.

$$\delta_{\mathrm{ij}} = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases} \quad (2.21)$$

Finally, the optimal $\alpha$ can be estimated with the Laplacian matrix defined in Equ. (2.22):

$$\hat{\alpha} = \arg\min \alpha^T L \alpha, \quad s.t. \quad \alpha = 1 \quad or \quad 0, \forall i \in \Omega\,, \quad (2.22)$$

where $\Omega$ indicates the pixels in the foreground and background. The closed-form matting can perform better compared to the Poisson matting by introducing the cost function.

Another classic propagation method based on the affinity is called random-walk matting [25]. It uses the probability that a random walker starts from the location for

a given unknown pixel $I$ to the foreground before it reaches the background region to estimate the corresponding factor $\alpha$. And a weighted graph for each pixel as a node is introduced in the random-walk method. The weighted graph is shown in Figure 2.4.

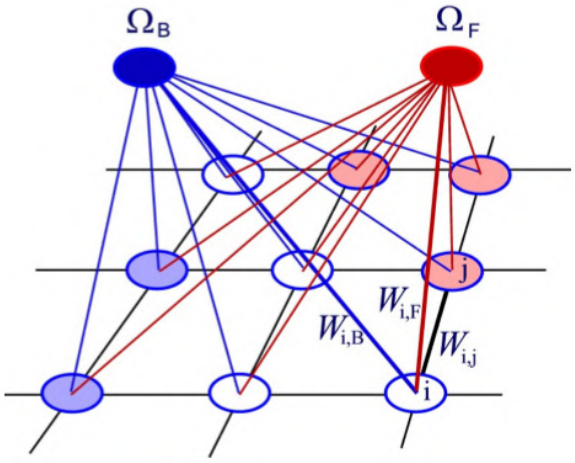



**Figure 2.4:** A weighted graph [31].

In Figure 2.4, $\Omega_F$ and $\Omega_B$ refer to the virtual foreground and background pixel nodes respectively. And the white nodes refer to the unknown pixels in the unknown region. The other blue and red nodes refer to the given pixels in the unknown region. A data weight $W_{iF}$ and $W_{iB}$ can be defined between each pixel and the foreground/background node. The edge weight $W_{ij}$ is defined between every two nearby pixels.

The energy function in the random-walk matting method is constrained by using the Locality Preserving Projections (LPP) technique of He and Niyogi [32]. And the function can be expressed in Equ. (2.23):

$$ZLZ^T x = \lambda ZDZ^T x\,, \tag{2.23}$$

where the matrix $Z$ is defined as 3 by N with a vector $I_i$, which indicates the RGB color at the pixel, as a column; $D$ is the diagonal matrix with the elementary $D_{ii} = \sum w_{ij}$, and the Laplacian matrix $L$ is defined in Equ. (2.24):

$$L_{ij} = \begin{cases} w_{ii} & \text{if } i = j \\ -w_{ij} & \text{if } i \text{ and } j \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases}, \tag{2.24}$$

where $w_{ii} = \sum_j w_{ij}$, $w_{ij}$ indicates the edge weight between pixel $i$ and pixel $j$. The dimension of Laplacian matrix $L_{ij}$ is defined as $N$ by $N$, where $N$ is the number of pixels in the image. By denoting the Laplacian matrix to Equ. (2.23), the edge weight $w_{ij}$ can be obtained in Equ. (2.25):

$$w_{ij} = \frac{exp(-\left\|I_i - I_j\right\|^2 / \sigma^2)}{\sum_k w_{ik}}, k \in \text{neighbors of } i\,, \tag{2.25}$$

where $\sigma$ is a tunable parameter that can be set automatically or by users. And $I$ is the transformed color [33] of the pixel.

The corresponding factor $\alpha$ can be estimated in the same way as closed-form

matting according to the weight $w_{ij}$ between pixels. In conclusion, the random-walk method estimates the corresponding factor $\alpha$ based on the probability from a given unknown pixel to a foreground pixel with a weight graph.

Moreover, KNN [34, 35] algorithm is also a commonly used method in propagation matting. It uses the $K$ nearest neighbor (KNN) to construct affinities based on the nonlocal principle. KNN matting makes the assumption that the corresponding factors $\alpha$ of pixels with a similar color and texture in the unknown region are also similar. In the KNN approach, a feature vector $X(i)$ at a given pixel $i$ can be expressed by Equ. (2.26):

$$X\left(i\right)=\left(\cos\left(H_i\right),\sin\left(H_i\right),S_i,V_i,x_i,y_i\right)\ , \tag{2.26}$$

where $H$, $S$, $V$ indicate the coordinate in HSV color space respectively as hue, saturation and lightness; $x$ and $y$ indicate the spatial coordinate respectively. We can obtain the affinity between two pixels $i$ and $j$ by using the feature vector for each pixel as in Equ. (2.27):

$$K\left(i,j\right)=1-\frac{\left\|X\left(i\right)-X\left(j\right)\right\|}{C}\,, \tag{2.27}$$

where $C$ is a parameter referred to as the least upper bound of $\left\|X\left(i\right)-X\left(j\right)\right\|$, and it can guarantee the affinity factor $K\left(i,j\right)$ varies from 0 to 1. With the affinity, we can estimate the $\alpha$ by the same way as closed-form matting.

The KNN matting does not rely on the local color-line model and it can perform better on the images with complex textures, while closed-form matting has better results when the change of image color is smooth.

### 2.2.3 Combination methods for natural image matting

The sampling-based methods have better performance in images with simple colors and can intuitively estimated the foreground and background colors by collecting samples in the neighborhood, while the propagation-based methods have better performance when the images have complex colors and texture. To improve the quality of results, recent researches proposed novel approaches [31, 36] by combining the sampling-based method and propagation-based method.

The robust matting method [31] is one of the representative combination methods which is also based on closed-form matting method. And it focuses on analyzing the confidence of foreground and background samples for unknown pixels. The method can obtain high-quality alpha matte by minimizing the energy with high confidence samples. The energy function is defined in Equ. (2.28):

$$E = \sum_{z \in \psi} \left[ \hat{f}_z (\alpha_z - \hat{\alpha}_z)^2 + \left(1 - \hat{f}_z\right) \left(\alpha_z - \delta\left(\hat{\alpha}_z > 0.5\right)\right)^2 \right] + \lambda \cdot J\left(\alpha, a, b\right) , \quad (2.28)$$

where $\hat{\alpha}_z$ indicates the estimation of the corresponding factor $\alpha$ and $\hat{f}_z$ is the confidence values [31], which is defined in sampling-based method to choose good foreground samples; $\hat{f}_z \left(\alpha_z - \hat{\alpha}_z\right)^2 + \left(1 - \hat{f}_z\right) \left(\alpha_z - \delta\left(\hat{\alpha}_z > 0.5\right)\right)^2$ indicates the data weight, while $J\left(\alpha, a, b\right)$ is based on the edge weight; $J\left(\alpha, a, b\right)$ is defined as the cost function

in closed-form matting as in Equ. (2.18); $\delta$ is the Boolean function [31] returning 0 or 1; $\lambda$ is the adjustable parameter set by users, which is always set to be 0.1. The energy function can also be considered as the combination of the data weight and edge weight. An optimal confidence for background and foreground can be calculated in Equ. (2.29):

$$f(F,B) = \exp\left(\frac{-R(F,B)^2 \cdot \omega(F) \cdot \omega(B)}{\sigma^2}\right), \tag{2.29}$$

where $\sigma$ is an adjustable parameter; $R(F,B)$ refers to the color linearity in Equ. (2.30); $\omega(F)$ and $\omega(B)$ refer to the color similarities in the foreground and background respectively as in Equ. (2.31):

$$R(F,B) = \frac{\|C - (\alpha F + (1-\alpha)B)\|}{\|F - B\|}. \tag{2.30}$$

$$\begin{aligned} \omega(F) &= \exp\left(-\frac{\|F-C\|^2}{D_F^2}\right) \\ \omega(B) &= \exp\left(-\frac{\|B-C\|^2}{D_B^2}\right) \end{aligned}, \tag{2.31}$$

where $D_F/D_B$ indicates the minimum distance between the sample of foreground /background and unknown pixel $C$.

With the estimated pair of foreground and background colors, the value of $\alpha$ can be estimated. The robust matting approach combines the advantages of sampling-

based methods by using the confidence value and the advantages of propagation-based methods by using the energy function as the constraint with the data weight and edge weight. Therefore, the quality and accuracy of the mattes can be improved by using the robust matting approach.

## 2.3 Separating reflections from glass

Polarization is always used to remove the reflection on glass. As we often see in real life, virtual image of an object is reflected on glass and the virtual image will overlap with the image of another object behind the glass, as shown in Figure 2.5.

**Figure 2.5:** An example of reflection on glass [5].

Therefore, the methods by using polaroid are proposed to separate the reflection and transparency. We will review some basic approaches based on polaroid and non-polarization-based approach to eliminate the reflected light.

### 2.3.1 Image-based method

This method is the earliest one to separate real and virtual objects from their overlapping images based on polaroid [37]. In this approach, several images are obtained by rotating a polarizing filter. Based on the physical property we introduced in Chapter 1, the reflected light can be almost eliminated when the incident angle is Brewster's angle. Therefore, in this approach, if we obtain enough images by rotating the polaroid filter, there will be an image with less reflection. By rotating the polaroid filter, we could obtain images at each polaroid filter angle $\theta$. The real image of an object A can be separated from the overlapped images by selecting a minimum image intensity of the given images $I(x, y; \theta)$ for each pixel, which can be expressed by Equ. (2.32):

$$\hat{I}_A(x, y) = \min_{\theta \in \Theta} I(x, y; \theta) \ , \tag{2.32}$$

where $\hat{I}_A(x, y)$ indicates the real image of the object A. Meanwhile, the virtual image of an object B can be separated by using the equation shown in Equ. (2.33) and Equ. (2.34):

$$I_{MAX} = \max_{\theta \in \Theta} I(x, y; \theta) \tag{2.33}$$

$$\begin{aligned} I_{MAX} &= I_A(x, y) + I_B(x, y) \\ \hat{I}_B(x, y) &= I_{MAX}(x, y) - \hat{I}_A(x, y) \end{aligned} \ . \tag{2.34}$$

However, this approach is based on the assumption that the reflected surface is planar and smooth, and the reflection should be specular. Also, the result is not

accurate. Reflection can not be fully removed by simply using this method and weak reflection may still remain in the image. Therefore, Schechner et al. proposed approaches [38, 39] to reduce the remaining reflection. However, the methods are based on the prior knowledge of incident angle to minimize reflection which is not realistic. In 1999, H.Farid and E.Adelson [40] proposed the approach by using independent component analysis (ICA) without the prior knowledge.

### 2.3.2 ICA Method

Another typical method is widely used to separate reflections from images. This approach combines polarization and independent components analysis (ICA) to remove reflections from a planar surface [41]. This method is to remove specular reflections from a planar surface, like taking photos through a planar glass. Two images are captured at different polarization angles. The images can be considered as a linear combination of a reflection layer and a transparency layer. The intensity of light at a single point in the image $y_1$ can be expressed by Equ. (2.35) [41]:

$$y_1 = aP + bR\,, \tag{2.35}$$

where $P$ and $R$ indicate the light intensity of the transparency layer and the reflection layer. And $a$, $b$ indicate the multiplicative constants. The constants $a$ and $b$ will change along with the different polarization angle. Similarly, another image $y_2$ can be expressed in Equ. (2.36):

$$y_2 = cP + dR\,, \tag{2.36}$$

where $c$, $d$ are different multiplicative constants.

To separate the transparency layer $P$ and reflection layer $R$, we have to solve the two equations with six unknowns. Therefore, the independent components analysis (ICA) algorithm [42, 43] was introduced to solve the problem. This algorithm is based on the assumption that the transparent image $P$ and the reflected image $R$ are independent, which means there is no correlation between real view and objects.

In this approach, two obtained images $y_1$ and $y_2$ are used for separation. Equ. (2.35) and Equ. (2.36) can be expressed in matrix form shown in Equ. (2.37):

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} , \tag{2.37}$$
$$Y = MX$$

where $x_1$ and $x_2$ are the row vector form of transparency layer $P$ and reflection layer $R$. And $M$ is the matrix embodies the linear mixing. $Y$ is overlapped images and $X$ is the original images. The purpose of the method is to recover the original images $x_1$ and $x_2$ from the overlapped images $y_1$ and $y_2$. Moreover, $M$ is assumed to be full rank. Therefore, the original images can be estimated with Equ. (2.38):

$$\hat{X} = M^{-1}Y , \tag{2.38}$$

where $\hat{X}$ is estimated original images. The matrix $M$ can be expressed according to the singular value decomposition (SVD) as in Equ. (2.39):

$$M = R_1 S R_2 , \tag{2.39}$$

where $R1$ and $R2$ indicate the rotation matrices, and $S$ is the scaling matrix. As the right column shown in Figure 2.6, the ICA algorithm is to transform a parallelogram into a square in geometry. The estimated original image X can be obtained by rotation and scaling.

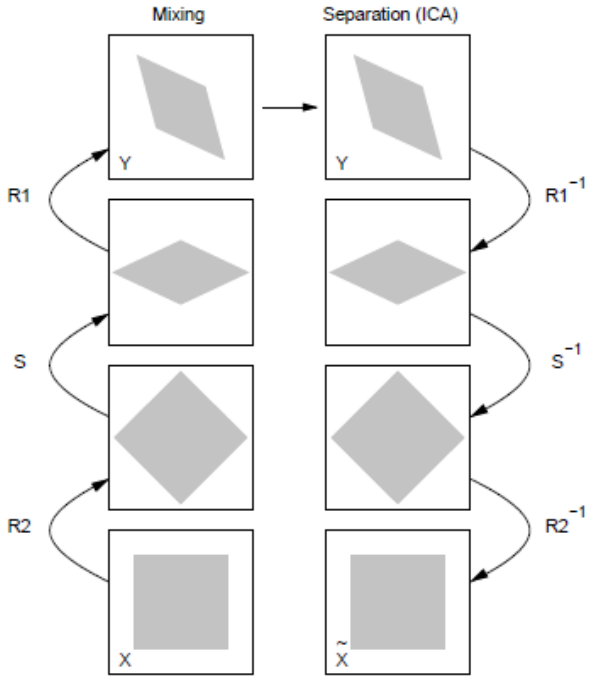



**Figure 2.6:** The strategy of the ICA algorithm in geometry [41].

For example, as shown in Figure 2.7, we can separate the mixed images by using the ICA algorithm. In Figure 2.7, images on the top are the mixed images which are combined with a transparent layer and reflected layer. The two overlapped images on the top are mixed with different coefficient in Equ. (2.35) and Equ. (2.36). By using the ICA algorithm, we can estimate the two original images as transparent layer and reflected layer, which is shown at the bottom of Figure 2.7.

**Figure 2.7:** An example of the ICA method to separate reflection.

According to this ICA approach, another ICA method for multiple polarized images based on sparsity image gradients was proposed by Bronstein to improve the accuracy and efficiency [44].

### 2.3.3 Misalignment

Another useful method to separate reflection from multiple images captured behind glass is called misalignment [45]. In this approach, the notion of repetition of space-time behavior characteristics is used to solve the problem. It makes the assumption

that the dynamics in one of the layers is repetitive while the other one is arbitrary. As shown in Figure 2.8, there are four frames captured from a movie. At the bottom of the figure, there are four corresponding frames based on the repetitive dynamics. We will use the two sets of frames with repetitions to separate the reflection.

**Figure 2.8:** An example of frames with repetitions [45].

First, the temporal distances between the occurrences of repetitive dynamics in the video sequences are detected. Then the global space-time alignment approach [46] proposed by Ukrainitz and Irani is used to detect and align repetitions in the multiple frames. As shown in Figure 2.9, in the video sequence, we can find two corresponding repetitions segment $A$ and segment $B$, and the segment $B$ can be aligned with $A$ by using spatio-temporal parametric transformation which can maximize a global similarity. The similarity measure is locally computed within small space-time windows by using covariance and variance of intensities. And the global space-time transformation can be recovered by using the alignment algorithm with the local sim-

ilarity, which is similar to optical flow algorithm. And the global similarity measure $M$ in paper [46] is formulated in Equ. (2.40):

$$M\left(\vec{p}\right) = \sum_{(x,y,t)\in A} C\left(w_A\left(x,y,t\right), w_B\left(x+\mu_1, y+\mu_2, t+\mu_3\right)\right), \tag{2.40}$$

where $C$ is the local similarity within each space-time window based on the local covariance and variance, and $M$ is the global similarity between two sequences; $\vec{p}$ is the unknown parametric transformation; $w_A$ and $w_B$ indicate the small window of sequences A and B respectively. According to the equation, the alignment problem is expected to find the spatio-temporal transformation $\vec{p}$ which can maximizes $M\left(\vec{p}\right)$. And in this alignment algorithm [46], the Newton method and Gaussian pyramid for both sequences $A$ and $B$ are used to find the transformation $\vec{p}$.

The misalignment of the repetitive frames of sequence $A$ and $B$ by using the alignment algorithm is shown in Figure 2.10. Moreover, as in Figure 2.11, a local space-time refinement can be used to reduce the residual misalignment in this approach. The refinement is to find a better local match within the other corresponded repetitions in the sequences. And the recovered transformation of frames in sequence $B$ can be obtained.

With the recovered transformation frames, we can separate the two layers. The first layer of repetitive dynamics can be extracted by using the median operator pixel $(x, y, t)$ in the domain of the spatio-temporal derivatives. Since the arbitrary dynamics are not aligned, it can be removed with the median. The second layer of arbitrary behaviors can be obtained by using the algorithm [47] defined in Equ. (2.41):

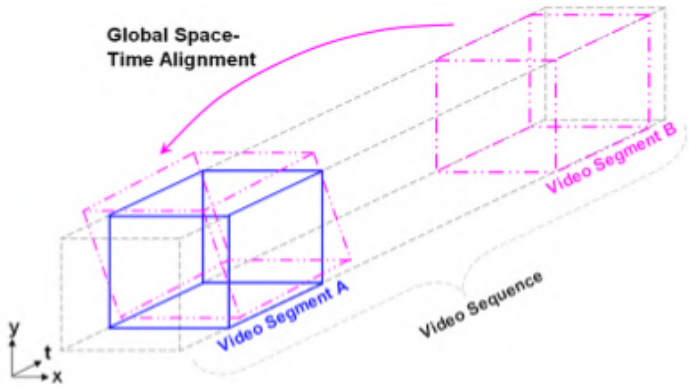



**Figure 2.9:** Global space time alignment [45].

**Figure 2.10:** Misalignment after applying alignment algorithm [45].

**Figure 2.11:** No residual misalignment by using local refinement [45].

$$L_2 = S - \alpha L_1 \,, \tag{2.41}$$

where $S$ refers to the input sequence; $L_1$ is the first extracted layer from the recovered frames. The second layer $L_2$ can be estimated by seeking the $\alpha$ which can minimize the correlation between the two layers $L_1$ and $L_2$. The final result of two separated layers by using the algorithm is shown in Figure 2.12. By using the misalignment algorithm based on global to local space-time alignment in a sequence, the transparent layers can be separated.

**Figure 2.12:** Layer separation results [45].

# Chapter 3

# Proposed method for color spill suppression

The proposed color spill suppression method system contains four parts: image preprocessing, region segmentation, reflection/transparency separation, and outliers processing. The overall procedures are presented in Figure 3.1. Generally, as shown in Figure 3.1, we input two images. One image is considered as the target image, on which we will conduct color spill suppression. The other image is considered as the reference image, which is used to compare the difference with the target image for suppression. Table 3.1 describes the images we use in different situations in our proposed method. For example, if the reflection on objects is specular, we use polarization to suppress color spill. We input an image at a random polarization angle $I_{\theta 1}$ as the target image, and input a reference image $I_{\theta 2}$ which has the same view but at a different polarization angle for comparison. If there is diffuse reflection on objects, we will use stereo matching with optical flow to conduct color spill suppression.

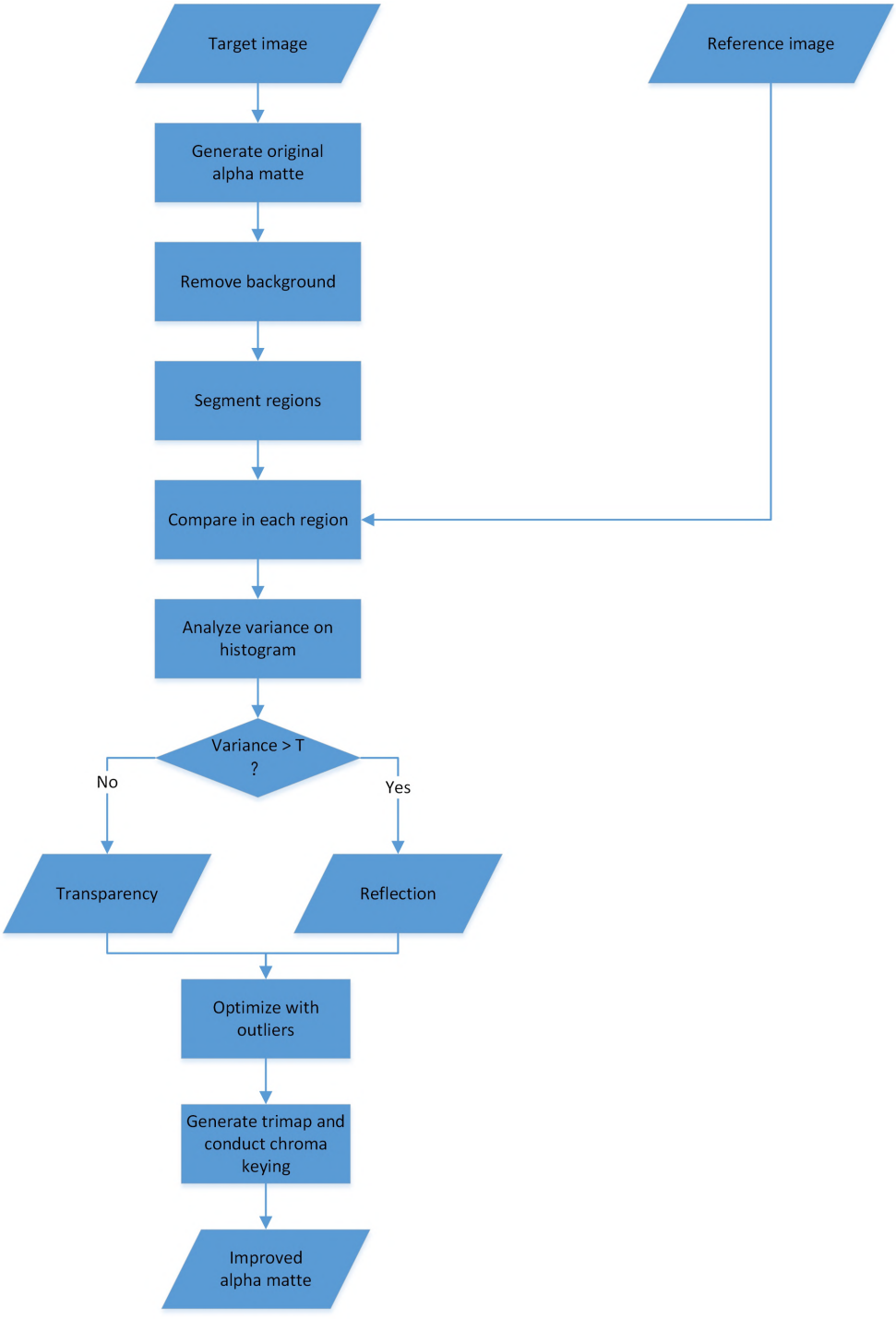



**Figure 3.1:** The flow chart of the proposed method.

**Table 3.1:** Input images for our proposed method.

| Image | Definition | Use | Application |
|---|---|---|---|
| $I_{\theta 1}$ | Image at a random polarization angle $\theta 1$ | Target image | Specular reflection |
| $I_{\theta 2}$ | Image at a random polarization angle $\theta 2$ | Reference image | Specular reflection |
| $I_L$ | Image from the left eye | Target image | Diffuse reflection |
| $I_R$ | Image from the right eye | Matching image for pre-processing | Diffuse reflection |
| $W$ | Warped image obtained from $I_L$ and $I_R$ | Reference image | Diffuse reflection |
| $I_{L\theta 1}$ | Image from the left eye at a random polarization angle $\theta 1$ | Target image | Both specular and diffuse reflection |
| $I_{R\theta 1}$ | Image from the right eye at a random polarization angle $\theta 1$ | | Both specular and diffuse reflection |
| $I_{L\theta 2}$ | Image from the left eye at a random polarization angle $\theta 2$ | | Both specular and diffuse reflection |
| $I_{R\theta 2}$ | Image from the right eye at a random polarization angle $\theta 2$ | Matching image for pre-processing | Both specular and diffuse reflection |
| $W_\theta$ | Warped image obtained from $I_{L\theta 1}$ and $I_{R\theta 2}$ | Reference image | Both specular and diffuse reflection |

We capture a pair of stereo images from the left eye and the right eye for pre-processing. The procedure of pre-processing is shown in Figure 3.2. The image from the left eye $I_L$ is chosen as the target image and the image from the right eye $I_R$ is the matching image. A warped image can be obtained by reconstructing the matching image with optical flow algorithm. The warped image $W$ is the reference image for comparison. Moreover, if the situation is complicated with both diffuse and specular reflection, we combine polarization and stereo matching to conduct suppression. Similarly, we captured two pairs of stereo images at different polarization angles such as $I_{L\theta 1}$, $I_{R\theta 1}$, $I_{L\theta 2}$ and $I_{R\theta 2}$. By pre-processing with $I_{L\theta 1}$ and $I_{R\theta 2}$, a warped image $W_\theta$ is obtained. We choose the image from the left eye at a random polarization

angle $I_{L\theta 1}$ as the target image and the warped image $W_\theta$ as the reference image in our proposed method. To reduce shadows from background, we remove background while keeping the foreground completely by generating a mask. Then we segment the unknown region into several segments. In our method, we compare two input images in each region and set a threshold (T) by analyzing variance distribution to separate transparency and reflection on the target image. The separated reflection and transparency can help us generate an improved and more accurate alpha matte. In the alpha matte, reflective regions are considered as foreground while transparency remains unchanged. The detailed description will be presented in the following sections.

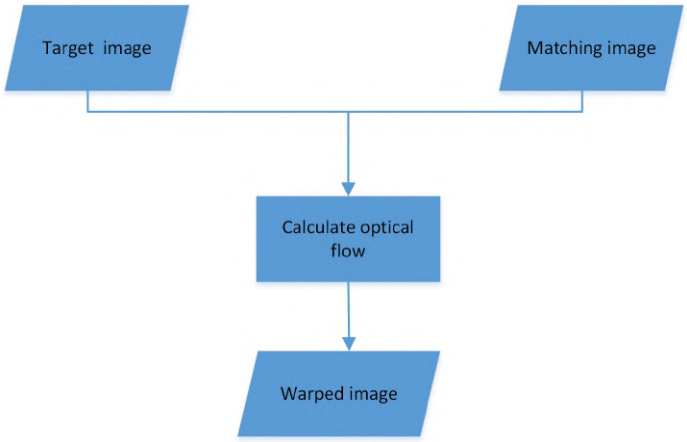



**Figure 3.2:** The flow chart for pre-processing.

## 3.1 Polarization

The reflection on the specular surface is partially polarized and its intensity can be manipulated by using a polaroid filter. This phenomenon is widely used to reduce the

glare effect on the glass. Figure 3.3 (a) presents a photo that is taken in front of the glass, and the reflection on the glass severely degrades the imaging quality. Figure 3.3 (b) shows another photo taken with the polaroid filter, in which the reflection on the glass is significantly reduced. A sketch of the imaging of Figure 3.3 is also presented in Figure 3.4.

**Figure 3.3:** Photographs taken through window.

In our proposed method, we make use of the property of polarization to verify our assumption that the intensity of reflected light will change more than transparency after polarization. However, the color spill suppression conducted on images which are captured in our MCRlab environment is different from the problem of separating reflection on glass. We can not use the current methods and algorithms based on polarization to separate the background reflection in chroma keying. A simple schematic diagram of our studio environment is shown in Figure 3.5. As we can see in this figure, we have a reflective cylinder and a transparent cuboid in front of the green screen. We can capture an image with transparency on the cuboid and reflection from background color on the top of the cylinder. The background color is the

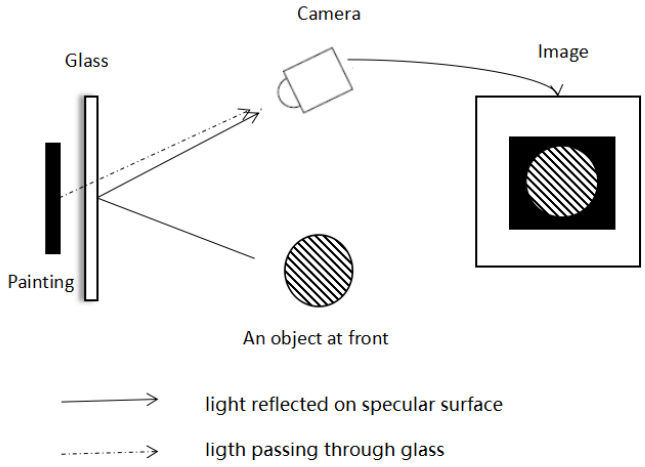



**Figure 3.4:** An example of photograph taken through glass. The image is the combination of the painting and reflected front objects.

diffuse reflection from a rough green screen which consists of unpolarized lights and some partially polarized lights, while specular reflection consists of partially polarized lights. Since transparency is the light passing directly through the object, it is part of the background color. After polarization, only a small part of transparency varied, which is less than the specular reflection.

The used polaroid filter is shown in Figure 3.6. The intensity of light will change while the polaroid is rotated. The polarized images with different polarization angles can be obtained and two of them are randomly chosen as the input images.

Meanwhile, the reflected light is complicated, which is mixed with specular reflection and diffuse reflection. In the problem of reflection separation on glass, since the incident plane is plain glass, the reflected lights are almost specular reflection which can be polarized. But in our problem, the incident plane is the surface of objects with different shapes and materials. If the light reflects on a rough surface, the re-

flection will be diffused which is similar to the backing color and transparency. We cannot separate transparency and diffuse reflection only by polarization. Hence we use stereo images for capturing the color spill phenomenon from multiple perspectives to conduct the suppression.

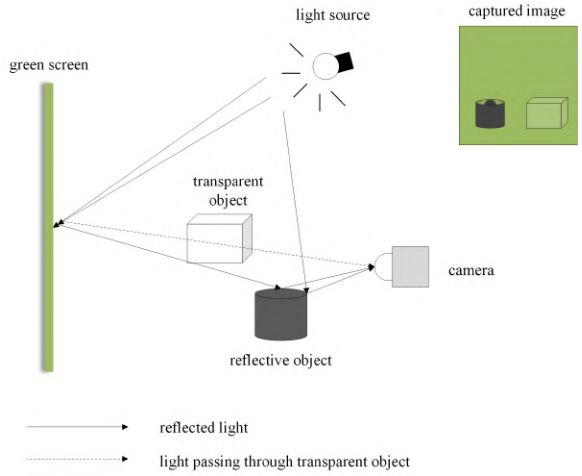



**Figure 3.5:** An example of a photograph of transparent and reflective objects in chroma keying.

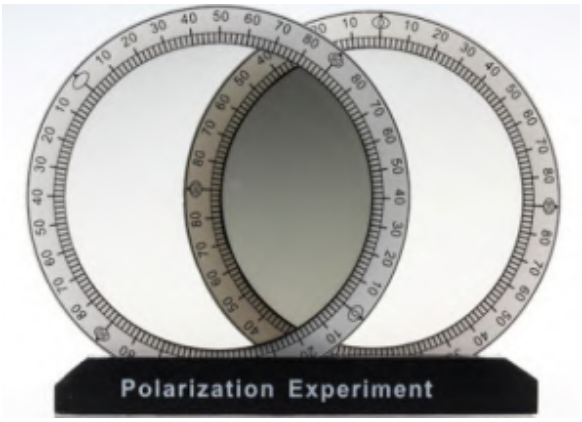



**Figure 3.6:** Polaroid filters used in experiments.

## 3.2 Stereo matching using optical flow

If the green color is diffused reflection which is reflected on a rough surface of objects, it is always weak and faint. When we change the angle of the reflective surface of objects from the left eye to the right eye, the position and intensity of diffuse reflection may change, while the transparency will have small changes in color and intensity. Stereo matching using optical flow for color spill suppression in our method is based on the assumption that the transparency has little color variation viewing from arbitrary directions, while the diffuse reflections change more in both color and intensity [48, 49].

Stereo images simulate the way that human perceives objects from both eyes. Stereo matching is always used to generate three-dimensional depth from two-dimensional images by finding out the corresponding pixels and their relationship. In our proposed method, we use optical flow to find corresponding pixels and calculate their difference between two stereo images when the light is diffusely reflected. The reflection and transparency can be separated by comparing the difference, which is based on the assumption that the specular reflections significantly change in both color and intensity, and diffuse reflections have a little color variation but more than transparency from arbitrary directions [50].

Since the stereo images are taken from two viewing points, there is a slight displacement between these two images. Normally the optical flow is an efficient way to estimate the motion vectors of the corresponding points in the adjacent video frames [51]. In the stereo matching method, we use the optical flow vectors that result from the displacement between the pair of stereo images to determine the correspondence between points [52].

In the conventional concept, the disparity of the corresponding points in frames is due to the movement of the points. The displacement of the viewing points can also introduce the disparity between corresponding points. In this case, one image from the stereo pair can be considered as being moving at time t, the other image can be considered as being moving at time $t + \Delta t$. The pixel intensity relation can be described by the following equation:

$$I((x, y), t) = I((x, y) + (\Delta x, \Delta y), t + \Delta t)\,, \tag{3.1}$$

where $I((x, y), t)$ refers to the image intensity of the two-dimensional point $(x, y)$ at time $t$. The instantaneous direction of the two-dimensional motion path at the point $(x, y)$ can be described with the vector $(u, v)$. And components of the vector are $u(x, y) = dx/dt$ and $v(x, y) = dy/dt$. At the time $\Delta t$, the displacement can also be indicated as $(\Delta x, \Delta y)$, where $\Delta x = u \cdot \Delta t$ and $\Delta y = u \cdot v\Delta t$.

The approach of finding out the vector $(u, v)$ was first introduced by Horn and Schunck in 1981 [53]. Horn and Schunck proposed the basic algorithm and introduced the constraint equation of optical flow by combining the two-dimensional velocity vectors for movement which are also called the "motion field" [53, 54]. In our proposed method, we adopt the approach from Liu [9]. The core of the algorithm is based on [55]. To obtain the displacement vector $(u, v)$, a constraint function, which was proposed in paper [55], is introduced to minimize the energy. The energy function is indicated in the following equation:

$$E(u,v) = E_{data}(u,v) + \lambda E_{smooth}(u,v)\,, \tag{3.2}$$

where the total energy $E(u,v)$ is constituted by a data term $E_{data}(u,v)$ which is defined in Equ. (3.4), and a smoothness term $E_{smooth}(u,v)$ which is defined in Equ. (3.4); $\lambda$ is the regularization parameter that is greater than 0.

$$E_{data}(u,v) = \int_{\Omega} \Psi\left(\left(|I(\mathbf{x}+\mathbf{w}) - I(\mathbf{x})|^2\right) + \gamma\left(\nabla I(\mathbf{x}+\mathbf{w})\right) - \nabla I\,|I(\mathbf{x})|^2\right) d\mathbf{x}\,, \tag{3.3}$$

where the function $\Psi(a^2) = \sqrt{a^2+\epsilon^2}$ is for minimization, and $\epsilon$ is a small positive constant chosen to be 0.001; $\gamma$ is a tunable parameter; $x$ is defined as $(x,y,t)^T$ and $w$ is defined as $(u,v,1)^T$; $\nabla$ is the gradient operator.

$$E_{smooth}(u,v) = \int_{\Omega} \Psi\left(|\nabla_3 u|^2 + |\nabla_3 v|^2\right) d\mathbf{x}\,, \tag{3.4}$$

where $\nabla_3$ is the spatio-temporal gradient which is defined as $(\partial_x, \partial_y, \partial_t)^T$. By minimizing the energy function, the displacement vector $(u,v)$ can be obtained. Then we know the motion of the object, and can "warp" the image with the optical flow vector.

Here image warping is a kind of digital image processing, and it is to manipulate the image with significant distortions on shapes. Warping is always used in image correction and for some creative purpose such as morphing [56]. An example of image

warping based on optical flow is shown in Figure 3.7:

(a) (b) (c)

**Figure 3.7:** An example for image warping.

In Figure 3.7, (a) is the original image, (b) is the matching image with some displacement, and (c) is the warped image obtained from the matching image by using the optical flow. By comparing the warped image (c) and the original image (a), we can find that there are only some reflected parts on the car surface having variations and movements, like the reflections in the red circle. We can use the difference between the original image and the warped image to separate reflection and transparency.

However, when the reflection is specular, it is always strong and has few changes from left eye to right eye in stereo images. Hence stereo matching may not work on specular reflection and we have to combine polarization and stereo matching for more complicated situations.

## 3.3 Region segmentation

As shown in Equ. (1.1), the $\alpha$ value is a blending vector varying from 0 to 1. If $\alpha = 0$, the pixel is completely background and if $\alpha = 1$, the pixel is completely foreground.

In our proposed method, the image is going to be segmented to different regions to do the separation. In other words, we extract the pixels where $0 < \alpha < 1$, no matter whether it is transparency or reflection.

Before segmentation, we have to generate an original alpha matte of the target image by using the AE (Adobe After Effects CS6).

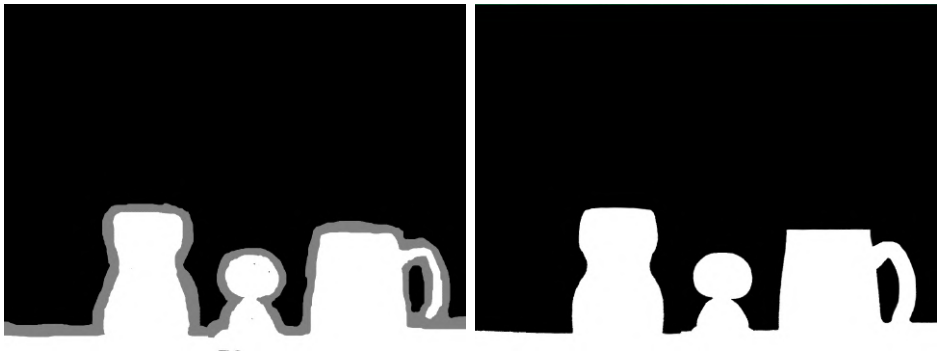
Since the lighting condition of chroma keying in real life is poor, it will introduce severe interference, such as shadows on background. If we directly generate the alpha matte of the image with poor lighting condition, the result could be noisy (see the example in Figure 3.8). As we can see, a large part of the background is considered as unknown region and will be treated as transparency, which will also influence the analysis of segmentation and separation. To avoid the influence from the background, we can generate a mask with a user-defined trimap to remove the background. The user-defined trimap and the mask are shown in Figure 3.9.


(a) (b)

**Figure 3.8:** Original image and its alpha matte taken from home made environment.

By combining the mask with the original alpha map, we can remove the shadow on the background. For example, Figure 3.10 shows the alpha map after using the

mask. However, we cannot distinguish if the pixel is reflection or transparency in the unknown region. In our proposed method, the next step is to segment and gather the pixels when $0 < \alpha < 1$ in the unknown region.



(a) The trimap of the original image. (b) The generated mask of the original image.

**Figure 3.9:** User-defined trimap and its mask.

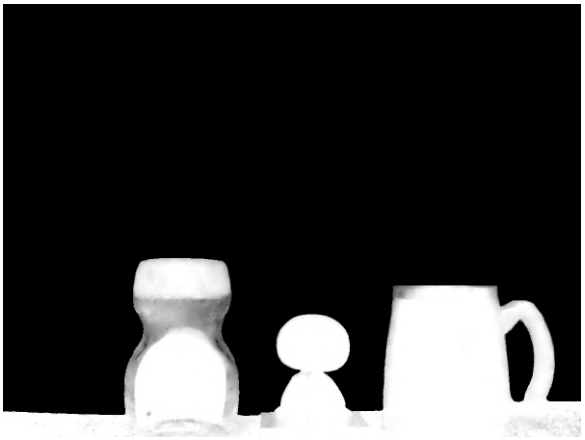


**Figure 3.10:** An alpha map by using the mask to remove the shadows in background.

The analysis is based on the assumption that reflection has large variation through a polaroid filter. Since the variation is displayed more obviously on a whole region than individual pixels, the analysis is conducted on each region instead of the individual pixel. We connect adjacent pixels with $0 < \alpha < 1$ as one region. Since noise is

always introduced to the image, one reflection region may be divided into several sub-regions. In the sub-regions, the difference between reflected light and transparency is not obvious to meet our assumption. Hence we introduce the "dilation" into our proposed method to connect all sub-regions in the same part of reflection or transparency to one region.

Dilation is one of the basic operations in mathematical morphology and it usually uses a structuring element for probing and expanding the shapes contained in the input image [57]. The dilation operation is used on a binary image, where pixels with $0 < \alpha < 1$ are replaced by 1 and others are replaced by 0. The binary dilation is a kind of shift-invariant operator. In mathematical morphology, a binary image can be considered as the subset of a Euclidean space $R^d$, where $d$ indicates the dimension and equals $0, 1, 2, \cdots$. In a Euclidean space $E$, $A$ is a binary image and $B$ is a structuring element which is also considered as a subset of $R^d$. And dilation of $A$ by $B$ can be indicated by using the following equation:

$$A \oplus B = \cup_{b \in B} A_b \,, \tag{3.5}$$

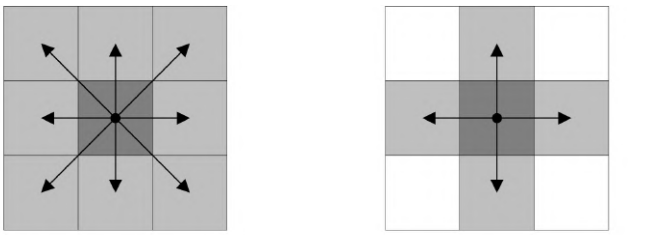
where $A_b$ is the translation of $A$ by $B$. Meanwhile, dilation has the characteristic and can be expressed by the following equation:

$$A \oplus B = A \oplus B = \cup_{\alpha \in A} B_a \,. \tag{3.6}$$

In our proposed method, we use a disk in the dilation operation to enlarge the region. If the disk is regarded as the subset $B$ and the original sub-region is regarded

as $A$. The dilation operation can be considered as the locus of the points covered by $B$ when the center of $B$ moves inside $A$. For example, in Figure 3.11, the operator is the disk which can also be regarded as the subset $B$. And the original subset of pixels with $0 < \alpha < 1$ is the small square which is also $A$. When doing the dilation of $A$ by $B$, the large square with round corners will be obtained. By doing the dilation operation, the regions are divided and then the analysis proceed in each region respectively. As shown in Figure 3.12, the subset $B$ can be chosen as 4 pixels connected or 8 pixels connected. In our method, we choose the subset with 8 pixels connected as shown in (a). If their edges or corners touch, the adjoining pixels are considered as part of the same region.


**Figure 3.11:** An example for dilation [57].



(a) Subset B with 8 pixels connected (b) Subset B with 4 pixels connected

**Figure 3.12:** Dilation subset size chosen in our method.

The image where $0 < \alpha < 1$ will be divided into several regions to do the separation. For example, as shown in Figure 3.13, the unknown region in the alpha matte is divided into four parts (a), (b), (c) and (d) by using dilation. To separate reflections and transparency, we will calculate and compare in each divided part.

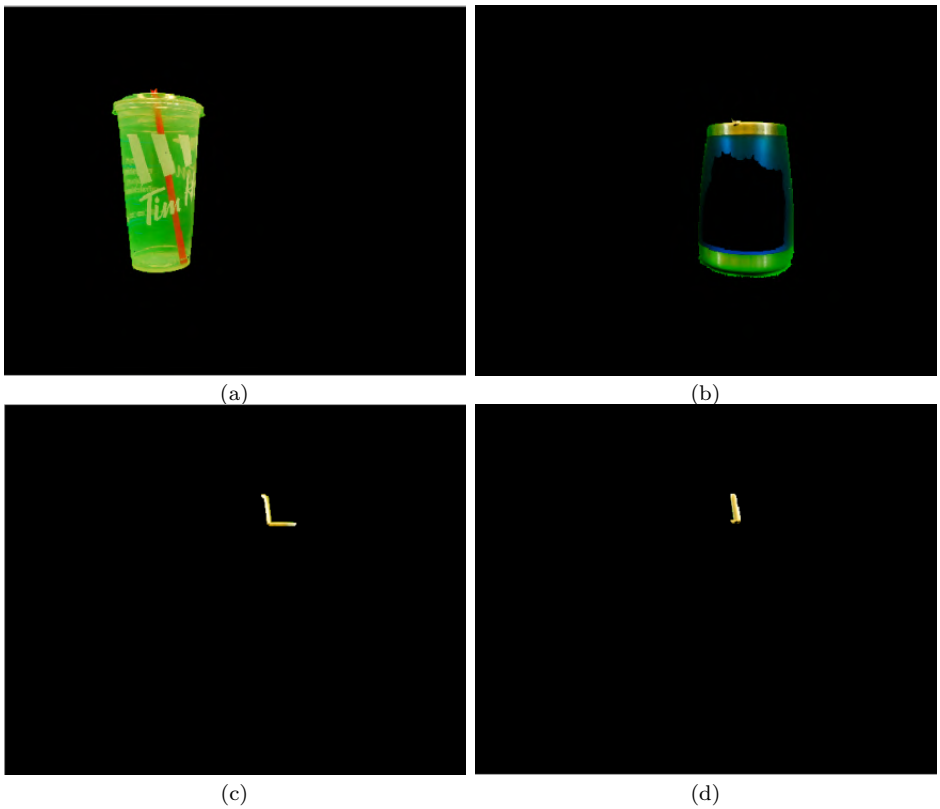



**Figure 3.13:** Segmented regions.

However, there are still some discrete pixels with $0 < \alpha < 1$ are left as shown in Figure 3.14. We call these pixels outliers. And we will deal with the outliers individually.

**Figure 3.14:** Outliers.

## 3.4 Reflection and transparency separation

The separation of reflection and transparency is an essential task for our proposed method. To separate reflection from transparency, we are going to analyze the differences between two images in each segmented region in HSV color space.

HSV (Hue, Saturation, Value) color model is one of the most common cylindrical-coordinate representations of points in an RGB color model [50]. This color space model was first introduced by Smith in 1978, and it is also called Hexcone Model. HSV color model is more intuitive than RGB color model which is based on Cartesian (cube) representation. In HSV model, $H$ means hue and it indicates the basic attribute of the color which also refers to the name of the color, such as red, yellow, etc. $S$ mans saturation, and it refers to the purity of color. $V$ is brightness or lightness. Since reflection is related to the saturation and brightness of objects, HSV color model can help us to analyze more intuitively than RGB model.

The transformation of the color space from RGB to HSV is defined in following

equations [50]:

$$h = \begin{cases} 0^\circ & \text{if } max = min \\ 60^\circ * ((g-b)/(max-min)) + 0^\circ & \text{if } max = r, g \geq b \\ 60^\circ * ((g-b)/(max-min)) + 360^\circ & \text{if } max = r, g \leq b \\ 60^\circ * ((b-r)/(max-min)) + 120^\circ & \text{if } max = g \\ 60^\circ * ((r-g)/(max-min)) + 120^\circ & \text{if } max = b \end{cases}, \tag{3.7}$$

$$s = \begin{cases} 0, & \text{if } max = 0 \\ (max-min)/max = 1 - min/max & \text{otherwise} \end{cases}, \tag{3.8}$$

$$v = max\,, \tag{3.9}$$

where $r$, $g$ and $b$ are coordinates of a color in RGB model, and the values are varying from 0 to 1; $max$ is the maximum value among $r$, $g$ and $b$, while $min$ is the minimum value among $r$, $g$ and $b$; $h$, $s$ and $v$ are coordinates of a color in HSV color space; where $h \in [0^\circ, 360^\circ]$, and it is the metric hue angle; $h$ can also be normalized to $[0, 1)$. $s \in [0, 1)$ and $v \in [0, 1)$ stand for saturation and brightness.

Histogram and variance are used to separate reflection from transparency in our proposed methods. Histogram and variance are often used in image processing for analysis. For example, they are used in specularity highlight removal based on light

field imaging [50], which is similar to our problem to be solved. Here we use variance in HSV space to indicate the difference between two images. The equation is shown as follows:

$$D_k = \sqrt{(|H_{1k}^2 - H_{2k}^2|) + (|S_{1k}^2 - S_{2k}^2|) + (|V_{1k}^2 - V_{2k}^2|)}\,, \tag{3.10}$$

where $D_k$ is the calculated variance in each region respectively. $k$ indicates the index of the region and it equals $0, 1, 2, \cdots$; where $H_{1k}$, $S_{1k}$ and $V_{1k}$ are components of hue, saturation and brightness of the first image in the region labeled as $k$; where $H_{2k}$, $S_{2k}$ and $V_{2k}$ are the hue, saturation and brightness of the second image in the region labeled as $k$.

In each segmented region, we observe the variance by using the histogram. With the underlying assumption that reflection has large variation on polarization and viewing from different directions, we can observe that in the histogram, more pixels in reflected regions are distributed in sections with large variation, while more pixels in transparent regions are centralized within the sections of small variation.

To separate reflection from transparency, we set a threshold $h_{thres}$ to detect and classify based on the histogram. The threshold $h_{thres}$ is adjusted within $[0, 1]$ with different images. If the variance of the region exceeds a given threshold, it indicates the region is reflection, since in this region, more pixels have large variation. Otherwise, the region is transparency.

## 3.5 Outliers

When analyzing the variance distribution histogram, we find that the regularity of distribution in some small regions is not obvious and correct based on our assumption since the pixel numbers in these regions are too few. To optimize the result, we deal with discrete pixels or pixels in small regions separately, and call them outliers. A threshold is set to separate outliers and regions. In our method, if the pixel numbers in a region is less than 50, we regard these pixels in the region as outliers. After separating regions into reflection and transparency, we are going to deal with the outliers. The specific procedure of outliers classification is shown in Figure 3.15. The distance between the outlier and the nearest reflection/transparency pixel $d_r/d_t$ is calculated and compared. If $d_r < d_t$, the outlier is considered as the reflection. Otherwise, the outlier is considered as transparency. By using outliers, the suppression will be more accurate.

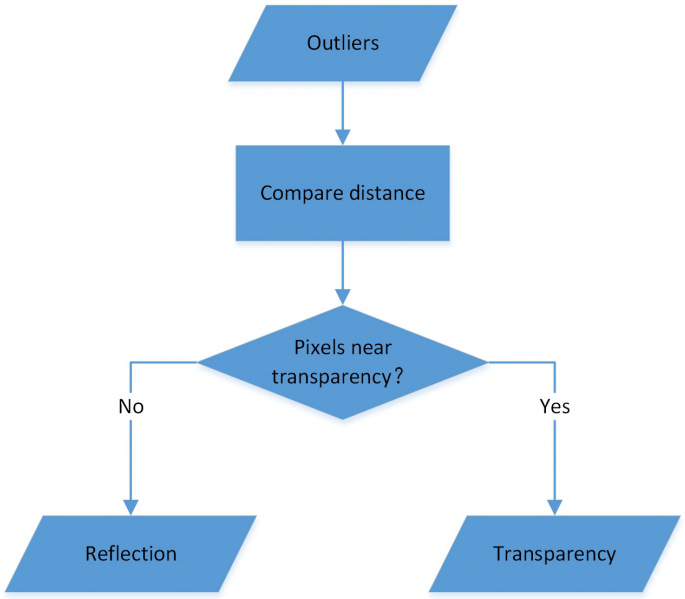



**Figure 3.15:** Separation of reflection and transparency for outliers.

## 3.6 Summary

Our proposed method is used to separate reflection from transparency on objects in chroma keying, which is also called color spill suppression. The reflection on objects is considered as the foreground and the value of $\alpha$ is set to 1. The value of $\alpha$ in transparent regions remains unchanged. With our proposed method, the quality of alpha matte can be improved.

# Chapter 4

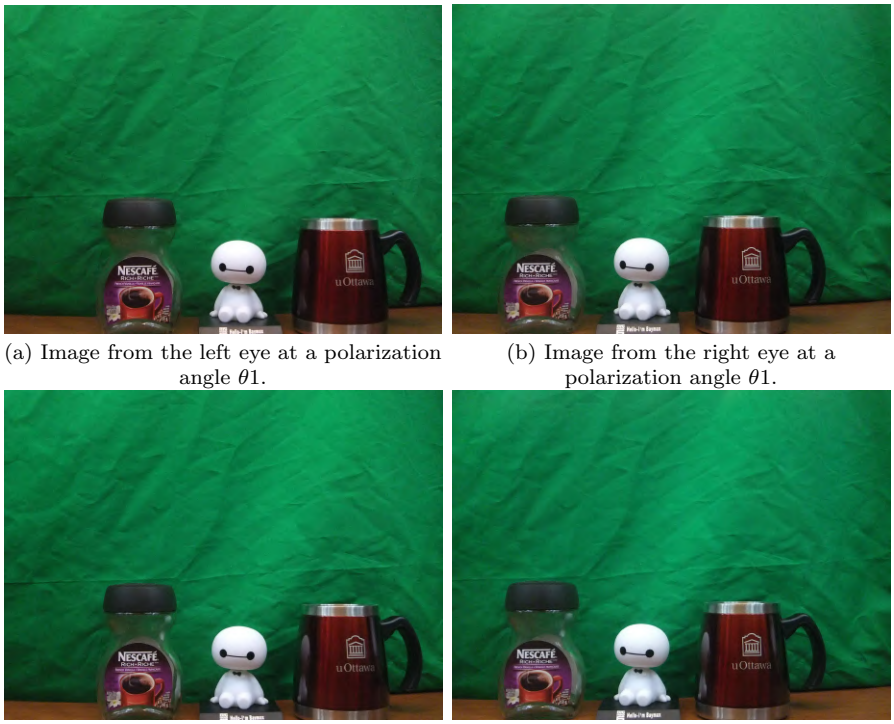
# Implementation and experimental results

In this chapter, we demonstrate the experimental results of our proposed method for color spill suppression. In our experiments, all the stereo images are taken by the binocular camera (Fujifilm 3D W3), and we consider the alpha matte conducted on target image by using the Adobe After Effects CS6 (AE) without color spill suppression as our original alpha matte. We will demonstrate the visual results of each step.

## 4.1 Pre-processing results

In our approach, we use stereo matching when there are diffuse reflections on objects. Stereo images are obtained by rotating the polaroid filter and using the binocular camera. For example, in a complicated situation, we capture two pairs of stereo

images at different random polarization angles as shown in Figure 4.1. In this figure, the reflection on the toy's head is diffuse, the reflection on the mug is specular and the bottle on the left is transparent.



(a) Image from the left eye at a polarization angle $\theta 1$.

(b) Image from the right eye at a polarization angle $\theta 1$.

(c) Image from the left eye at a different polarization angle $\theta 2$.

(d) Image from the right eye at a different polarization angle $\theta 2$.

**Figure 4.1:** Example of capturing stereo images at different polarization angels.

In Figure 4.1, (a) is taken from the left eye at a random polarization angle and (b) is taken from the right eye at the same polarization angle. Then we rotate the polaroid filter to change the polarization angle, another pair of stereo images from the left eye and right eye can be captured as shown in Figure 4.1(c) and (d). As demonstrated in the last chapter, we need to pre-process the stereo images. We choose two of these four images. One is an image taken from the left eye at a polarization angle which is considered as a target image, on which we will conduct segmentation and color spill

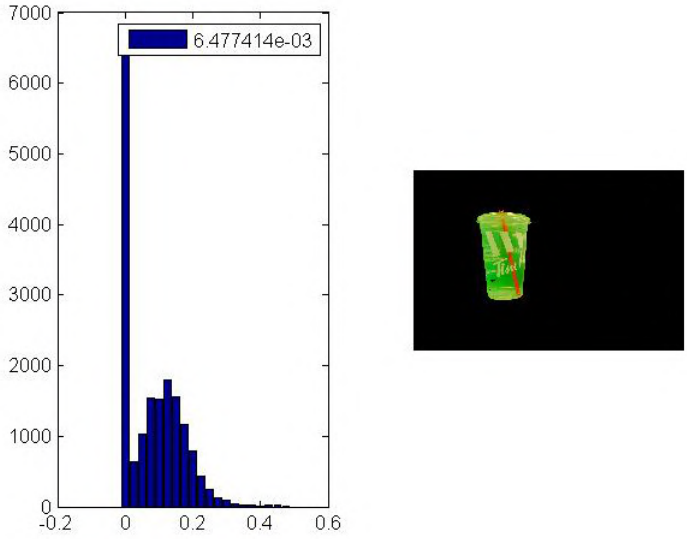
suppression. And the other image taken from the right eye at a different polarization angle is a matching image for disparity estimation. We can obtain a warped image by calculating optical flow to reconstruct the matching image with the disparity space. We use the warped image as the reference image to compare with the target image for suppression. An example of obtained warped image is shown in Figure 4.2.


**Figure 4.2:** Example of a warped image.

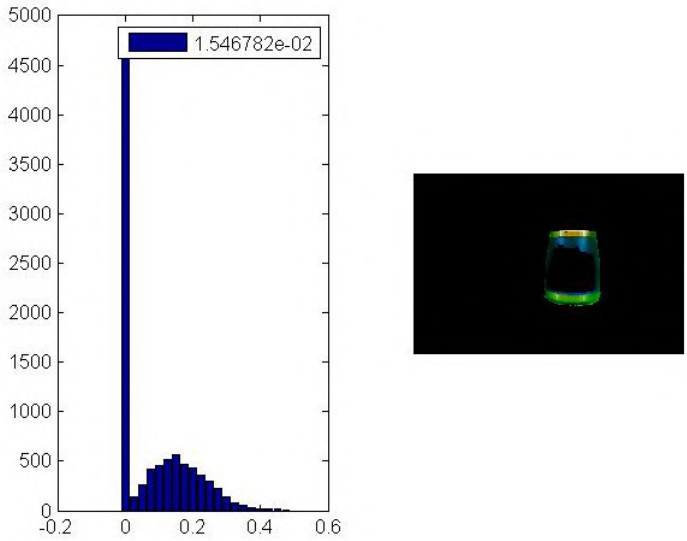
## 4.2 Histogram analysis for separation

As mentioned in Chapter 3, the target image is segmented into several regions and labeled. We conduct comparison and separation in each region, and use histograms to present the difference between the target image and the reference image. In the histogram, the Y coordinate indicates the number of pixels, and the X coordinate indicates the value of variance. The bar in the histogram means how many pixels are distributed in the indicated variance range. The histogram expresses the distribution of pixels' variance in a region. It is obviously shown in Figure 4.3 that the scales of

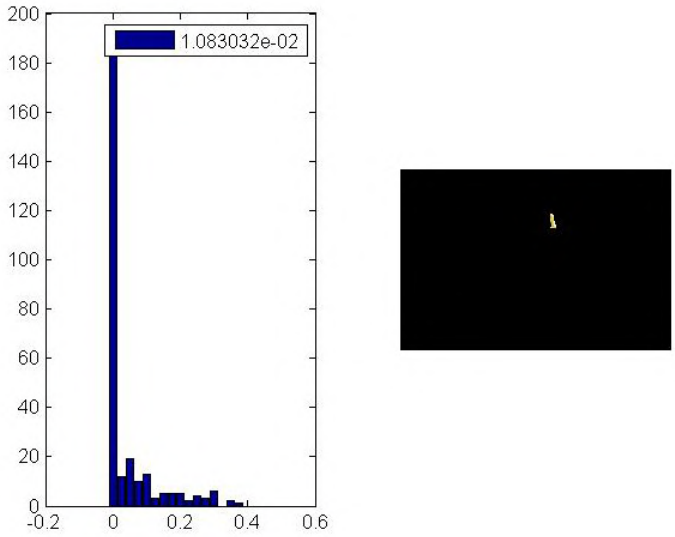
Y coordinate are different, since the size of each region is different and the number of pixels in each region is also different. If we used the same scale, the histograms of small regions like Figure 4.3 (c) and (d) would be compressed too small to be observed. For example, as shown in the right part of Figure 4.3, the unknown regions are segmented into four parts. In this figure, (a) is a transparent region on the cup, (b) is a reflected region on the bottle, (c) is a reflected region on the spoon and (d) is a reflected region on the straw. To the left of the figure is the histogram of variance distribution for each part. According to our assumption that the values in reflected regions change more than the values in transparent regions, we mainly focus on the large variance range such as last several bars in the histogram. By observation and analysis, the reflected regions indeed has more pixels in large variance range compared with transparent regions, which also demonstrates that the values in reflected regions change more than the values in transparent regions. Since the size of each region is different, we use the percentage instead of number of pixels to separate. The percentage is the number of pixels within the large variance range like last 10 bars over the total number of pixels. According to the percentage, we can set a threshold to conduct the suppression. If the percentage is larger than the threshold, it means that the region has more variation and it should be the reflection. Otherwise, the region should be the transparency. The threshold parameter is tunable according to different images.


6.477414e-03
7000
6000
5000
4000
3000
2000
1000
0
-0.2
0
0.2
0.4
0.6

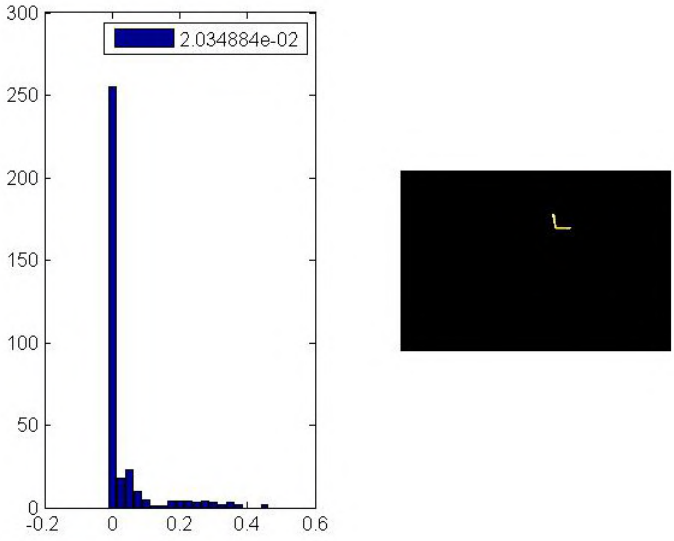
(a)


1.546782e-02
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-0.2
0
0.2
0.4
0.6

(b)




(c)

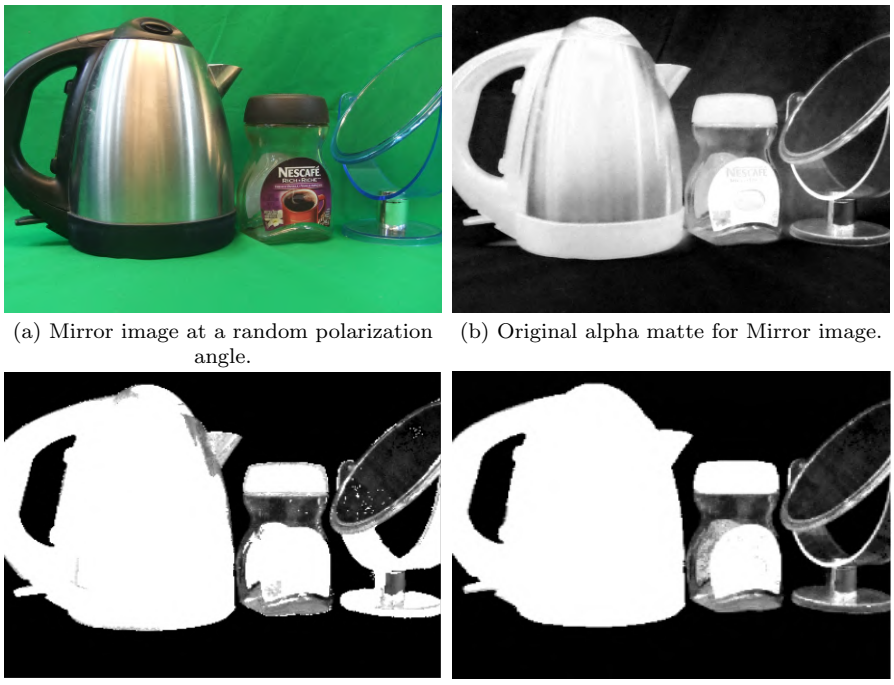



(d)

**Figure 4.3:** Histogram analysis in each segment of alpha matte. The scales of Y coordinate in the histograms are different.

## 4.3 Optimization with outliers

In our proposed method, we introduce outliers as mentioned in Chapter 3 to optimize the result. As shown in Figure 4.4 (a), there are some reflections on the water kettle at the left, the bottle at middle and the mirror are transparency. And Figure 4.4 (b) is the original alpha matte. In the alpha matte, the color gray indicates transparency, black indicates background, and white indicates foreground. In this original alpha matte, the reflection on the bottle is considered as transparency incorrectly. To solve this problem, our proposed method without optimization is used and the final result is shown in Figure 4.4 (c). As we can see, there are still some error pixels, such as the pixels on the mirror. When the new background is composed, the error is more obvious and the image quality is not as good as expected. Hence, we introduce the outliers as optimization. The error pixels can be eliminated by using optimization and the quality of alpha matte can be improved as shown in Figure 4.4 (d).



(a) Mirror image at a random polarization angle.

(b) Original alpha matte for Mirror image.

(c) Alpha matte after color spill suppression without optimization.

(d) Alpha matte after color spill suppression with optimization.

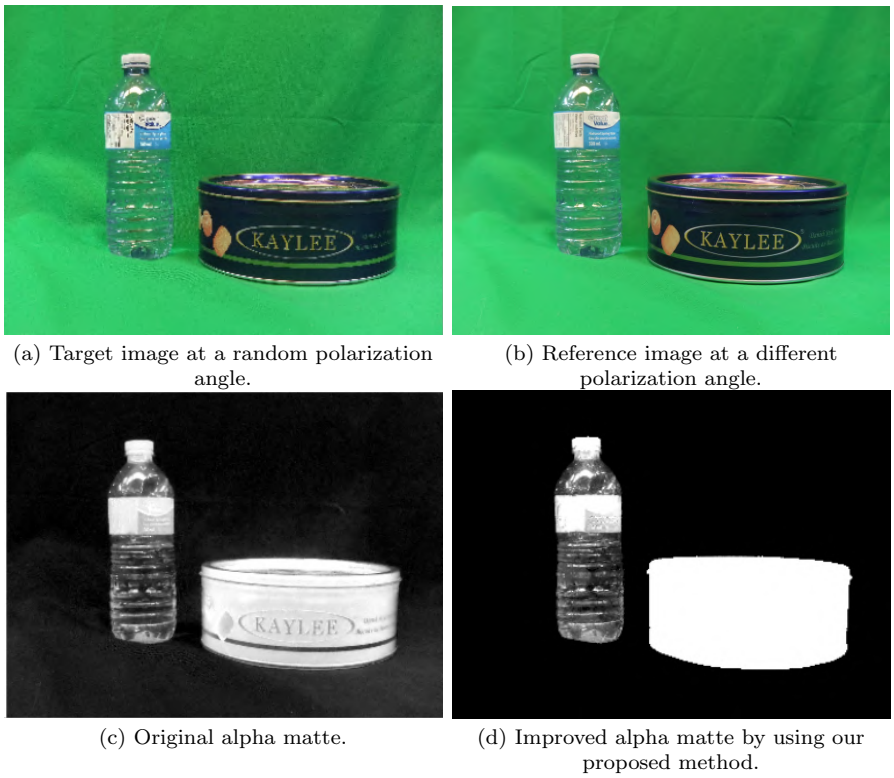
**Figure 4.4:** Comparison for optimization.

## 4.4 Experimental results

In this section, we test our approach on various images. All the test images are taken in our lab with the self-constructed experimental environment. The experimental results show the performance of our proposed method on color spill suppression and will be visually compared with other state-of-the-art matting methods on color spill suppression.

### 4.4.1 Performance of our proposed method

In the test image Cracker Box, since the reflection on the object is specular, we can use polarization to suppress color spill as shown in Figure 4.5.




(a) Target image at a random polarization angle.

(b) Reference image at a different polarization angle.

(c) Original alpha matte.

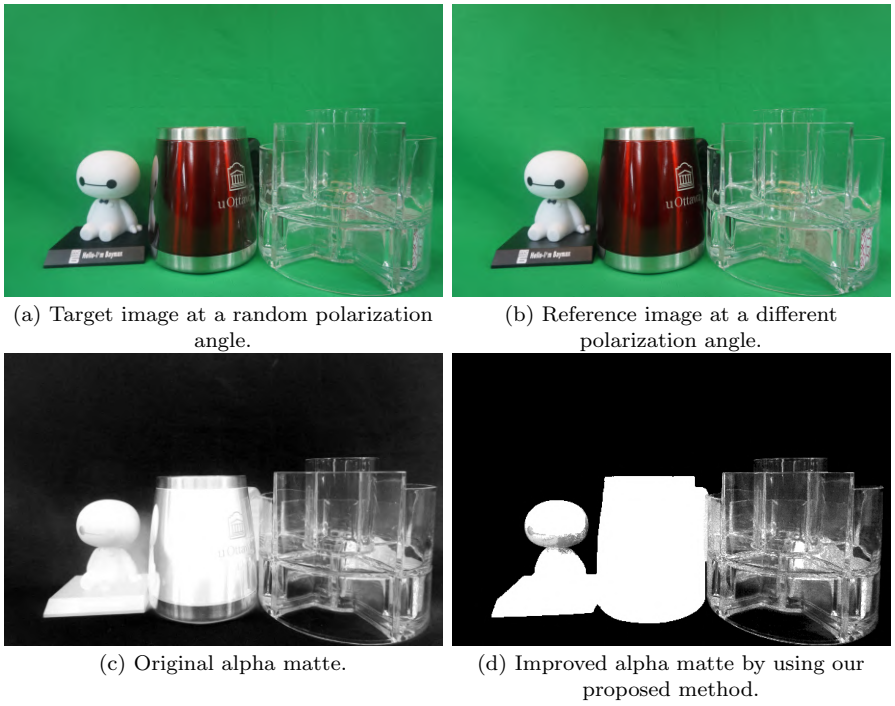
(d) Improved alpha matte by using our proposed method.

**Figure 4.5:** Cracker Box image and experimental result by using polarization.

In Figure 4.5, (a) is the target image at a random polarization angle, and (b) is the reference image at a different polarization angle. As we can see, the bottle is transparent and the cracker box is reflected to be green. The original alpha matte by using the AE is shown in Figure 4.5 (c). In the alpha matte, the gray regions indicate transparency, black regions indicate the background and white regions indicate the foreground. The reflection on the cracker box is considered as transparency by

mistake. To suppress this color spill, the unknown region in alpha matte is divided into several segments. We compare the variance on histogram between two input images in each segmented region. By comparison, a threshold can be set to separate reflection and transparency. Then an improved alpha matte by using our proposed method can be obtained as shown in Figure 4.5 (d). In the improved matte, reflection is separated and will be processed as foreground, while the transparency with the blending factor $\alpha$ remains the same.

However, as we mentioned, polarization's property is based on specular reflection. If there is diffuse reflection on objects, it can not be separated by only using polarization. We conducted color spill suppression on the Toy image by using polarization as shown in Figure 4.6.



(a) Target image at a random polarization angle.

(b) Reference image at a different polarization angle.

(c) Original alpha matte.

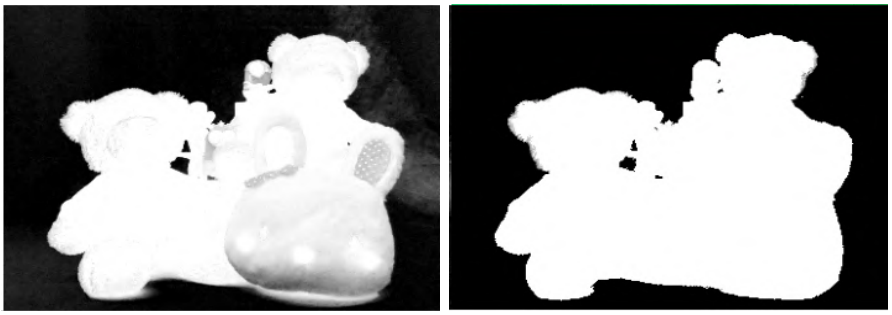
(d) Improved alpha matte by using our proposed method.

**Figure 4.6:** Toy image and experimental result by using polarization.

In Figure 4.6, (a) is the target image, and (b) is the reference image at a different polarization angle. In this figure, the reflection on the frosted toy is diffuse, the refection on the middle mug is specular and the container on the right is transparent. As shown in Figure 4.6 (c), all the reflections are considered as transparency in the original alpha matte, which is not correct. In this complicated situation, if we only use polarization, the improved alpha matte by using our method is shown in Figure 4.6 (d). The specular reflection in the middle can be processed as foreground, while the diffuse reflection on the left is still treated as transparency incorrectly.

Bear image is shown in Figure 4.7 (a). Since the surface of plush toys are rough, the reflection in the image are all diffuse reflection. Polarization can not separate reflection and transparency in this image. In this situation, we use stereo matching to suppress color spill. We capture a pair of stereo images and choose the image from the left eye as the target image to conduct color spill suppression on it. And we use the warped image obtained by stereo matching as the reference image for comparison in each segment, as shown in Figure 4.7 (b). The original alpha matte is shown in Figure 4.7 (c). Obviously, the diffuse reflection on the dolls are processed as transparency in the original alpha matte by mistake. The mistake can be corrected by using our proposed method and the improved alpha matte is shown in Figure 4.7 (d).

(a) Target image from the left eye. (b) Warped image.

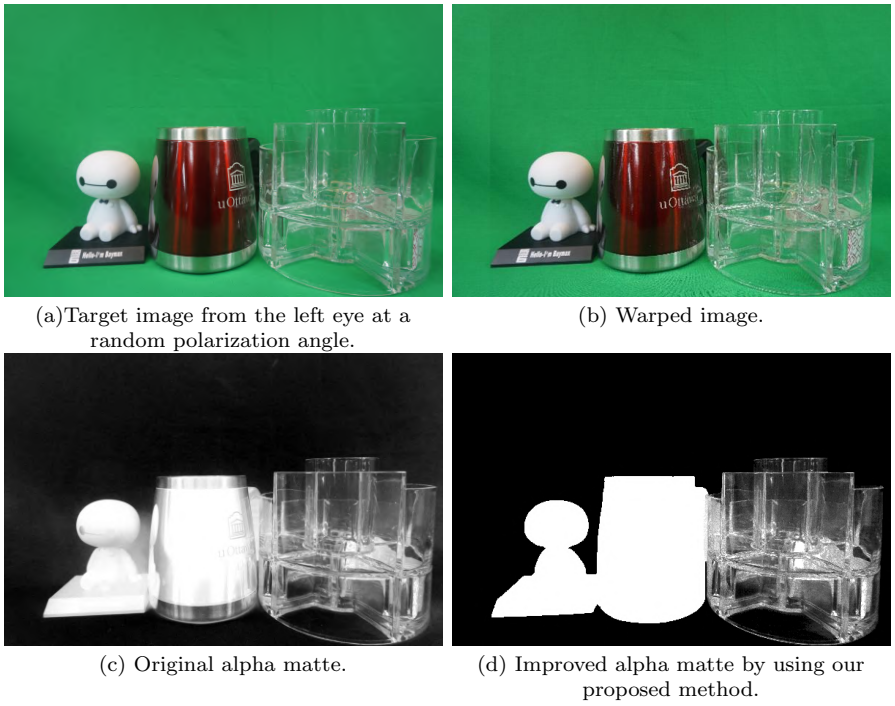


(b) Original alpha matte. (c) Improved alpha matte by using our proposed method.

**Figure 4.7:** Bear image and experimental result by using stereo matching.

Since the specular reflection is always strong, it might have few changes in specular reflection regions from left eye to right eye when we take stereo images. Hence, in this complicated situation with both specular reflection and diffuse reflection, we combine stereo matching and polarization to suppress color spill. For the input images, we could choose an image from the left eye at a random polarization angle as the target image and the warped image obtained by stereo matching as the reference image. For example, the diffuse reflection can not be suppressed by only using polarization in Toy image as we mentioned in Figure 4.6. We use our combined method to conduct the suppression on the toy image. As shown in Figure 4.8 (d), all the reflections can be separated from transparency in our improved alpha matte. The reflections on the toy and the mug are processed as foreground accurately while transparency remains

the same.



(a)Target image from the left eye at a random polarization angle.

(b) Warped image.

(c) Original alpha matte.

(d) Improved alpha matte by using our proposed method.

**Figure 4.8:** Toy image and experimental result by using polarization and stereo matching.

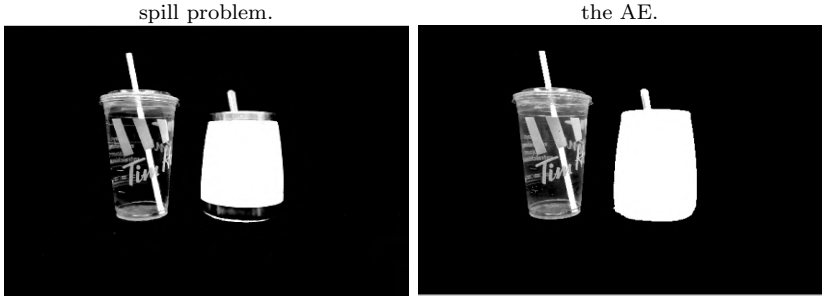
However, there are some potential problems in our proposed method, for example, the mirror reflection problem. Mirror reflection is a special case in chroma keying. Since the green reflection happened on the mirror is always the background color, the mirror reflection should be treated as transparency instead of foreground as other reflections in order to simulate the views that human perceives. Also, since the mirror reflection is strong and it is the unpolarzied light, it does not obey our assumptions. But our proposed method could still have better performance than others, and the visual comparison result will be shown in the next section.

### 4.4.2 Visual comparison

To show the improved quality by using our proposed method, we compare the visual quality of ours with the AE and Wang's method [2] which are the state-of-the-art software or algorithm. The original alpha mattes in the thesis are all from the AE without adjusting any parameters nor conducting color spill suppression. As shown in Figure 4.9, (a) is the original alpha matte for Mug image. In this image, the cup on the left is transparent and the mug on the right is wholly reflective. We use the AE to conduct the color spill suppression by adjusting parameters, and an improved alpha matte can be obtained as shown in Figure 4.9 (b). In AE's improved alpha matte, we can find that some reflection regions on the mug are corrected and processed as the foreground, which can improve the accuracy to some extent. However, there are still some reflections are treated as transparency by mistake. And the blending factor for transparency is also changed and some transparency is processed as background. The improved alpha matte with color spill suppression by using Wang's method [2] is shown in Figure 4.9 (c), which is similar to the result of the AE. Wang's method also uses tunable parameters to suppress color spill. It can suppress color spill to some extent while some regions with color spill still remain as mistakes. Our proposed method's result by using polarization is shown in Figure 4.9 (d). As we can see, it can separate the reflection from unknown regions and process it as foreground, while the blending factor in transparency can be kept. The quality and accuracy of our proposed method is better than the AE and Wang's method.

(a) The original alpha matte with color spill problem.

(b) The improved alpha matte by using the AE.

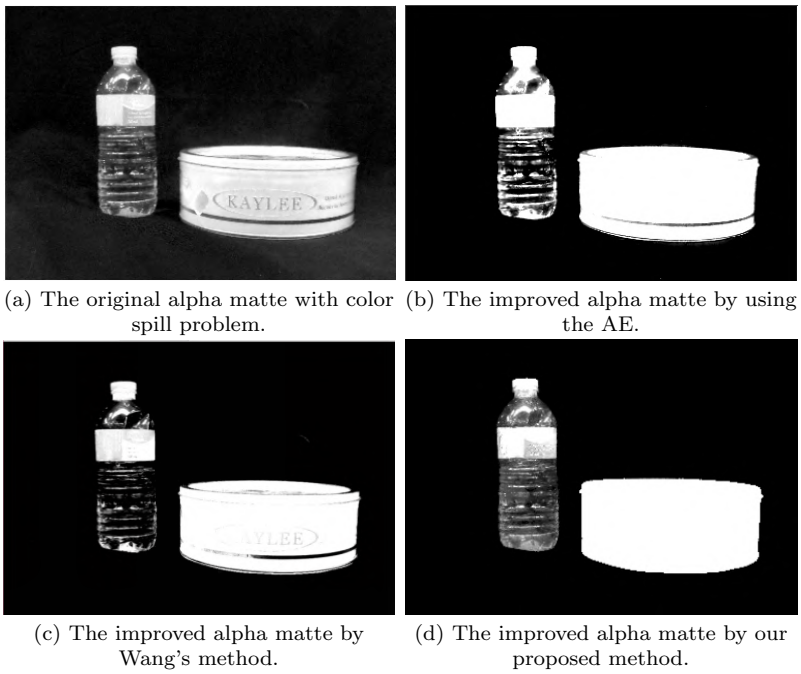


(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

**Figure 4.9:** Chroma keying of Mug image for visual comparison.

As shown in Figure 4.10, (a) is the original alpha matte without any color spill suppression, and reflections on the cracker box on the right are all considered as transparency by mistake. The matting result with color spill suppression by using the AE is shown in Figure 4.10 (b). Some reflections are corrected and processed to be foreground, while some transparency on the bottle is treated as background which is not expected. The result of Wang's color spill removal method is shown in Figure 4.10 (c). It can also reduce the color spill and has better performance in transparency on the bottle. But it still can not separate reflection and transparency. The improved alpha matte by using our proposed method with polarization is shown in Figure 4.10 (d). It can correct transparency into reflection on the cracker box and keep the transparency on the bottle at the same time. In this figure, our proposed method has better visual quality and higher accuracy than others.



(a) The original alpha matte with color spill problem.

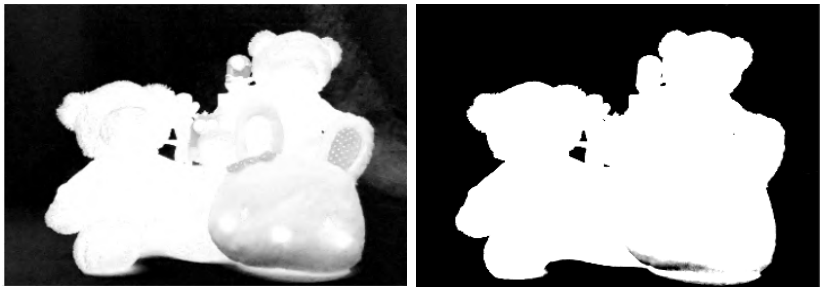
(b) The improved alpha matte by using the AE.

(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

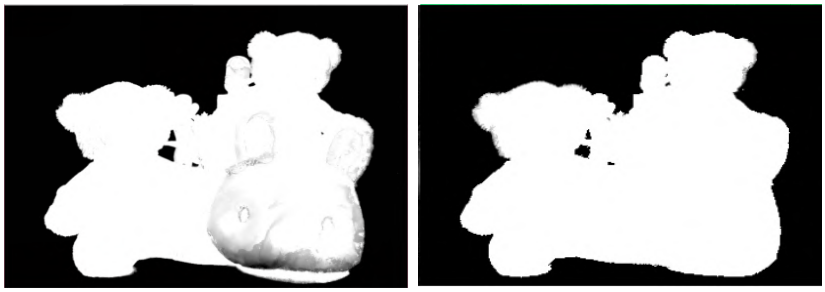
**Figure 4.10:** Chroma keying of Cracker Box image for visual comparison.

For diffuse reflection, our proposed method also has better performance than others. As shown in Figure 4.11, (a) is the original alpha matte without color spill suppression. As we can see in the alpha matte, the diffuse reflection on stuffed toys is considered as transparency. The AE's color spill method can separate most of the reflection regions and process them as foreground as shown in Figure 4.11 (b). As shown in Figure 4.11 (c), the improved alpha matte for Wang's method with color spill suppression can reduce color spill and keep the transparency on the rough edge to improve the quality and accuracy. However, there are still some diffuse reflection on the bunny's face that is treated as transparency incorrectly. The experimental result for our proposed method by using stereo matching with optical flow is shown in Figure 4.11 (d). It can process the reflection as foreground and keep the transparency on the boundary, which is also better than others.



(a) The original alpha matte with color spill problem.

(b) The improved alpha matte by using the AE.



(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

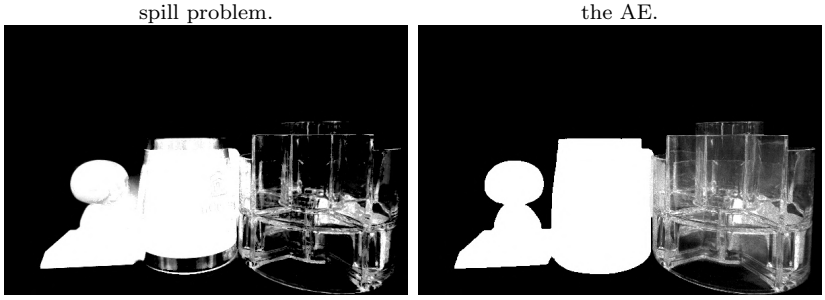
**Figure 4.11:** Chroma keying of Bear image for visual comparison.

For a more complicated situation, the visual comparison for color spill suppression is shown in Figure 4.12, where (a) is the original alpha matte without any operation. As we expressed in Figure 4.8, we can suppress color spill on Toy image with our proposed method. When using the AE to suppress color spill, it's hard to keep transparency while processing reflection as foreground by adjusting parameters. The color spill suppression result of the AE is shown in Figure 4.12 (b). Similarly, the result of Wang's suppression method is shown in Figure 4.12 (c). In order to modify the diffuse reflection on the toy from transparency to foreground, some reflections on the mug and some transparency are processed as background incorrectly. The result from our proposed method by combining polarization and stereo matching is shown in Figure 4.12 (d). Obviously, our method has better performance than others, since

we can generate an improved and accurate alpha matte with separated reflections and transparency.

(a) The original alpha matte with color spill problem.

(b) The improved alpha matte by using the AE.

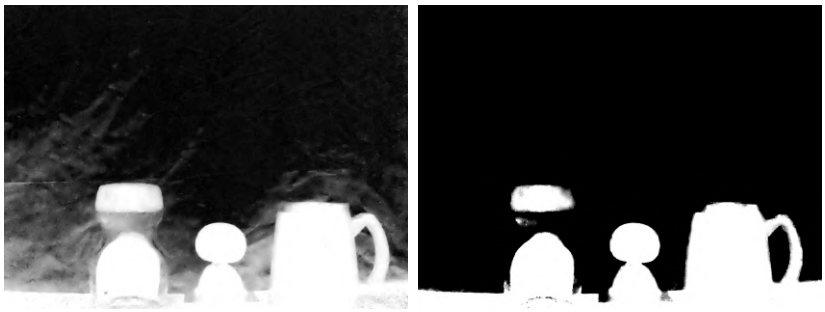


(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

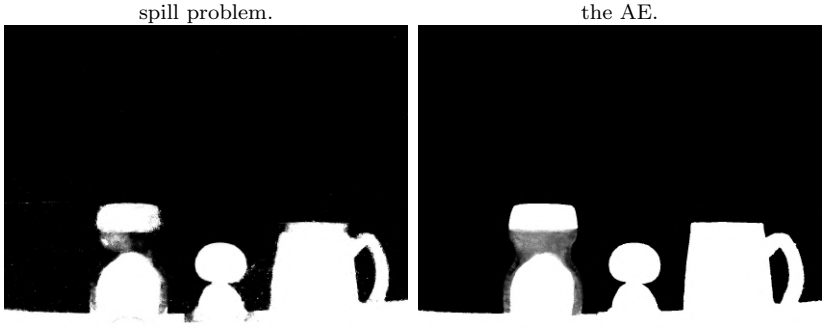
**Figure 4.12:** Chroma keying of Toy image for visual comparison.

Another complicated example is shown in Figure 4.13. As shown in Figure 4.13 (a), there is specular reflection on the mug to the right, a few diffuse reflections on the middle and transparency on the bottle to the left. By using the AE's suppression method, the result is shown in Figure 4.13 (b). The alpha matte with color spill suppression from Wang's method is shown in Figure 4.13 (c). Obviously, both of them can not separate reflection and transparency accurately. Our method combined stereo matching and polarization can generate a better alpha matte with separated reflection and transparency as shown in Figure 4.13 (d).



(a) The original alpha matte with color spill problem.

(b) The improved alpha matte by using the AE.

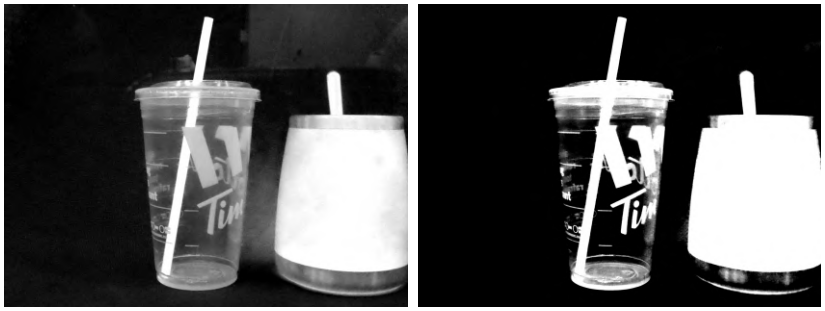


(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

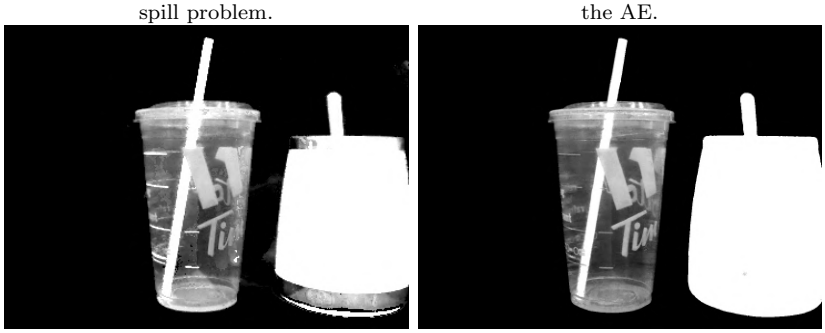
**Figure 4.13:** Chroma keying of Bottle image for visual comparison.

Another test image is similar to Mug image but with different positions and placement, which is called Cup image. As shown in Figure 4.14 (a), there is only specular reflection and transparency in the Cup image. When testing Mug image, we only used polarization to conduct color spill suppression since the reflection in the image is specular. In order to test if the combined method could also solve the simple situation like only specular reflection, we use the combined method to process this Cup image. Our result is shown in Figure 4.14 (d), where reflection and transparency still can be separated by using the combined method. The AE's suppression result is shown in Figure 4.14 (b) and Wang's method result is shown in (c). They can not remove reflection while keep transparency at the same time. Our method still has better performance than theirs.



(a) The original alpha matte with color spill problem.

(b) The improved alpha matte by using the AE.



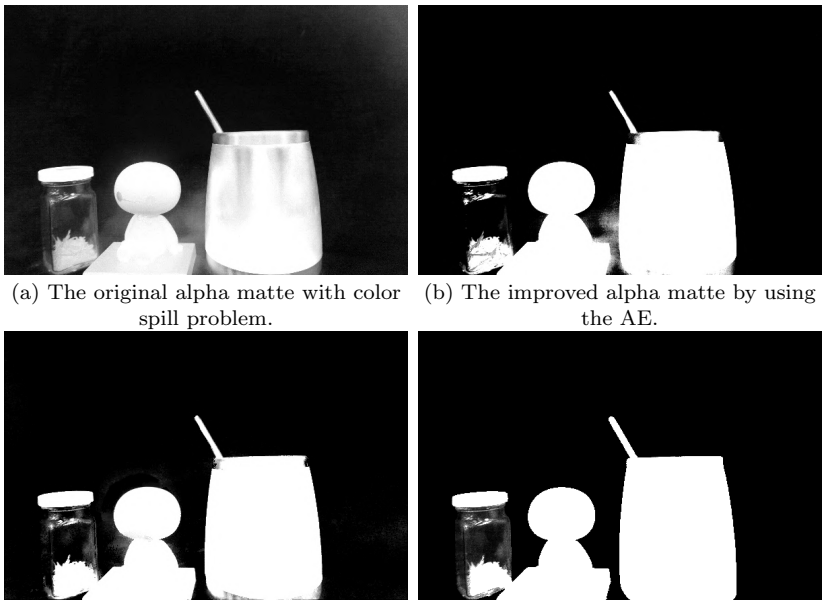
(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

**Figure 4.14:** Chroma keying of Cup image for visual comparison.

Figure 4.15 (a) is an original alpha matte without conducting color spill suppression in a complicated situation. On the left, there is a transparent jar with some tea inside. In the middle, there is some diffuse reflection on the toy's head and some specular reflection on the bottom. The reflection on the mug to the right is specular. When conducting color spill suppression, as shown in Figure 4.15 (b), the AE can not separate reflection and transparency. There is still reflection left on the mug and the transparency is treated as background incorrectly. Wang's method has similar performance as shown in Figure 4.15 (c), it can keep some transparency while remove the reflection in the middle. However, reflection on the right still remains. Our method as shown in Figure 4.15 (d) has better performance than theirs. The transparency on the left is kept the same while both diffuse reflection and specular reflection are pro-

cessed as foreground accurately. Our method can improve the quality of the original alpha matte by conducting color spill suppression.



(a) The original alpha matte with color spill problem.

(b) The improved alpha matte by using the AE.

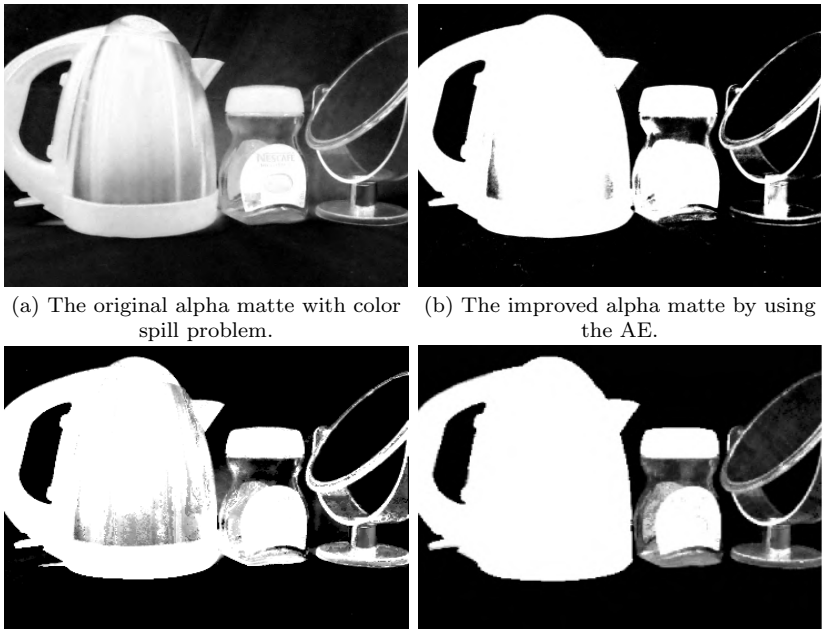
(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

**Figure 4.15:** Chroma keying of Jar image for visual comparison.

Figure 4.16 is a problematic mirror image which is a special case we introduced in the last section. Figure 4.16 (a) is the original alpha matte for Mirror image with some color spill problems. As we can see, there is a specular reflective kettle on the left, a transparent bottle in the middle and a mirror with a transparent body on the right. We are going to keep the transparency and amend the reflection on the kettle as foreground for an improved alpha matte. However the green reflection on the mirror is reflected from the background. In other words, it means that the view on the mirror is the background. Hence the reflection should be treated as transparency to keep the view. For this problematic image, the improved result from the AE is shown in Figure 4.16 (b), where the mirror reflection is treated as background while there

are still some reflections left on the kettle by mistake. Similarly, Wang's suppression method can not separate reflection and transparency accurately either, which is shown in Figure 4.16 (c). The result by using our proposed method is shown in Figure 4.16 (d), where the reflection on the kettle can be processed as the foreground accurately while transparency can be kept. Even though our proposed method is not suitable for the mirror reflection, it is treated as transparency which is exactly what we expect. And overall, our proposed method still has better performance than others for this problematic image.



(a) The original alpha matte with color spill problem.

(b) The improved alpha matte by using the AE.

(c) The improved alpha matte by Wang's method.

(d) The improved alpha matte by our proposed method.

**Figure 4.16:** Chroma keying of Mirror image for visual comparison.

## 4.5 Summary

Since all the images are captured in our lab by ourselves and there is no ground truth for the images, we conduct visual comparison among our proposed method and other state-of-the-art matting methods with color spill suppression. Our proposal method can separate reflection and transparency on different kinds of objects in our life and generate an improved alpha matte with higher accuracy and better quality. As we can see, our method not only has better performance when only using polaroid or stereo matching, but also has better quality when using the combined method.

# Chapter 5

# Conclusion and future work

## 5.1 Conclusions

In this thesis, we proposed a novel method for color spill suppression in chroma keying to improve the quality. Matting is used to extract the accurate foreground from an image. And this technology is widely used in real life, such as the film industry. However, there are still limitations in chroma keying systems. When transparency and reflection exist in the image at the same time. It is difficult to separate the reflection and the transparency when conducting chroma keying. Reflection will be treated as the transparency. And this problem has not been effectively solved by current methods. Therefore, we proposed a novel method to separate reflection and transparency which is also called color spill suppression. In our approach, we generate a system to conduct the separation. Moreover, polarization and stereo matching with optical flow algorithm are used in our method. The image can be segmented into four regions: background, foreground, transparency, and reflection. The reflection regions

are considered as the foreground and the value of $\alpha$ is set to 1, while the value of $\alpha$ in transparency remains the same. Therefore, the obtained alpha matte by using our method is more accurate compared to other methods.

## 5.2 Future work

In this paper, we proposed a novel color spill suppression method. However, there are still some problems left in the method and can be improved. On the one hand, there are some parameters need to be adjusted along different images to separate the reflected regions and transparent regions in our method. The method can be improved by making parameters self-adaptive along different images. On the other hand, we use a manually drawn trimap in the method to remove background and keep the foreground completely at the same time. It is more accurate and can reduce the negative impact on the results by using the manually drawn trimap. However, since the trimap needs to be drafted by users, it is not convenient. Therefore, our method can be improved with a better self-adaptive algorithm to remove background and shadow.

# References

[1] Thomas Porter and Tom Duff. Compositing digital images. *ACM SIGGRAPH Computer Graphics*, 18(3):235–259, 1984.

[2] Wenyi Wang and Jiying Zhao. Robust image chroma-keying: a quadmap approach based on global sampling and local affinity. *IEEE Transactions on Broadcasting*, 61(3):356–366, 2015.

[3] Jan Achenbach. *Wave propagation in elastic solids*. Elsevier, New York, 2012.

[4] Edward Collett. *Field guide to polarization*. SPIE, Bellingham, WA, 2005.

[5] Naejin Kong, Yu-Wing Tai, and Sung-Yong Shin. High-quality reflection separation using polarized images. *IEEE Transactions on Image Processing*, 20(12):3393–3405, 2011.

[6] Marvin R Querry. Direct solution of the generalized fresnel reflectance equations. *Josa*, 59(7):876–877, 1969.

[7] Snell's law of refraction. `https://en.wikipedia.org/wiki/Snell%27s_law`. Accessed: 2016-10-6.

[8] Fresnel equations. `https://en.wikipedia.org/wiki/Fresnel_equations`. Accessed: 2016-10-6.

[9] Ce Liu. *Beyond pixels: exploring new representations and applications for motion analysis*. Massachusetts Institute of Technology, Cambridge, Massachusetts, 2009.

[10] P. Vlahos. Composite color photography. U.S. Patent 3,158,477, November 24, 1964.

[11] P. Vlahos. Electronic composite photography. U.S. Patent 3,595,987, July 24, 1971.

[12] P. Vlahos. Electronic composite photography with color control. U.S. Patent 4,007,487, February 8, 1977.

[13] P. Vlahos. Comprehensive electronic compositing system. U.S. Patent 4,100,569, July 11, 1978.

[14] Viet-Quoc Pham, Keita Takahashi, and Takeshi Naemura. Real-time video matting based on bilayer segmentation. *Computer Vision-ACCV 2009*, pages 489–501, 2010.

[15] Yung-Yu Chuang, Aseem Agarwala, Brian Curless, David H. Salesin, and Richard Szeliski. Video matting of complex scenes. *ACM Transactions on Graphics (TOG)*, 21(3):243–248, 2002.

[16] Tai-Pang Wu, Chi-Keung Tang, Michael S. Brown, and Heung-Yeung Shum. Natural shadow matting. *ACM Transactions on Graphics (TOG)*, 26(2):8, 2007.

[17] Yung-Yu Chuang, Dan B. Goldman, Brian Curless, David H. Salesin, and Richard Szeliski. Shadow matting and compositing. *ACM Transactions on Graphics (TOG)*, 22(3):494–500, 2003.

[18] Douglas E. Zongker, Dawn M. Werner, Brian Curless, and David H. Salesin. Environment matting and compositing. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pages 205–214, 1999.

[19] Yung-Yu Chuang, Douglas E. Zongker, Joel Hindorff, Brian Curless, David H. Salesin, and Richard Szeliski. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 121–130, 2000.

[20] Mark A. Ruzon and Carlo Tomasi. Alpha estimation in natural images. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 18–25, 2000.

[21] Gunter Wyszecki and Walter Stiles Stanley. *Color Science*. Wiley, New York, 1982.

[22] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Computer Vision and Pattern Recognition*, volume 2, pages II–II, 2001.

[23] A. Berman, A. Dadourian, and P. Vlahos. Method for removing from an image the background surrounding a selected object. U.S. Patent 6,134,346, October 17, 2000.

[24] A. Berman, P. Vlahos, and A. Dadourian. Comprehensive method for removing from an image the background surrounding a selected subject. U.S. Patent 6,134,345, October 17, 2000.

[25] Leo Grady, Thomas Schiwietz, Shmuel Aharon, and Rudiger Westermann. Random walks for interactive alpha-matting. In *Proceedings of VIIP*, volume 2005, pages 423–429, 2005.

[26] Christoph Rhemann, Carsten Rother, Pushmeet Kohli, and Margrit Gelautz. A spatially varying psf-based prior for alpha matting. *Computer Vision and Pattern Recognition (CVPR)*, pages 2149–2156, 2010.

[27] Christoph Rhemann, Carsten Rother, and Margrit Gelautz. Improving color modeling for alpha matting. *British Machine Vision Conference (BMVC)*, 1(7):3, 2008.

[28] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum. Poisson matting. *ACM Transactions on Graphics (ToG)*, 23(3):315–321, 2004.

[29] Zhenlong Du, Hai Lin, Xueying Qin, and Hujun Bao. Oriented poisson matting. In *Proceedings of IEEE International Conference on Image Processing, ICIP*, volume 2, pages II–626, 2005.

[30] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):228–141, 2008.

[31] Jue Wang and Michael F. Cohen. Optimized color sampling for robust matting. *Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[32] Xiaofei He and Partha Niyogi. Locality preserving projections. In *Advances in Neural Information Processing Systems*, pages 153–160, 2004.

[33] Sebastian Thrun and Lawrence K. Saul. *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*. MIT press, Cambridge, Massachusetts, 2004.

[34] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang. Knn matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2175–2188, 2013.

[35] Dingzeyu Li. Knn matting. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 869–876, 2012.

[36] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23(3):309–314, 2004.

[37] Noboru Ohnishi, Kenji Kumaki, Tsuyoshi Yamamura, and Toshimitsu Tanaka. Separating real and virtual objects from their overlapping images. *European Conference on Computer Vision*, pages 636–646, 1996.

[38] Yoav Y. Schechner, Joseph Shamir, and Nahum Kiryati. Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 814–819, 1999.

[39] Yoav Y. Schechner, Joseph Shamir, and Nahum Kiryati. Polarization and statis-

tical analysis of scenes containing a semireflector. *Journal of the Optical Society of America (JOSA A)*, 17(2):276–284, 2000.

[40] Hany Farid and Edward H. Adelson. Separating reflections from images by use of independent component analysis. *Journal of the Optical Society of America (JOSA A)*, 16(9):2136–2145, 1999.

[41] Hany Farid and Edward H. Adelson. Separating reflections and lighting using independent components analysis. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 262–267, 1999.

[42] J-F Cardoso. Source separation using higher order moments. *Acoustics, Speech, and Signal Processing*, pages 2109–2112, 1989.

[43] Pierre Comon. Separation of stochastic processes. *Higher-Order Spectral Analysis*, pages 174–179, 1989.

[44] Alexander M. Bronstein, Michael Zibulevsky, and Yehoshua Y. Zeevi. Sparse ica for blind separation of transmitted and reflected images. *International Journal of Imaging Systems and Technology*, 15(1):84–91, 2005.

[45] Bernard Sarel and Michal Irani. Separating transparent layers of repetitive dynamic behaviors. In *Tenth IEEE International Conference on Computer Vision*, volume 1, pages 26–32, 2005.

[46] Yaron Ukrainitz and Michal Irani. Aligning sequences and actions by maximizing space-time correlations. *Computer Vision CECCV*, pages 538–550, 2006.

[47] Bernard Sarel and Michal Irani. Separating transparent layers through layer information exchange. *Computer Vision CECCV*, pages 328–341, 2004.

[48] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003.

[49] Takeo Kanade and Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, 1994.

[50] Chenxue Xu, Xingzheng Wang, and Haoqian Wang. Accurate image specular highlight removal based on light field imaging. *Visual Communications and Image Processing (VCIP)*, pages 1–4, 2015.

[51] Steven S. Beauchemin and John L. Barron. The computation of optical flow. *ACM Computing Surveys (CSUR)*, 27(3):433–466, 1995.

[52] M. Hatzitheodorou, E.A. Karabassi, G. Papaioannou, A. Boehm, and T. Theoharis. Stereo matching using optic flow. *Real-Time Imaging*, 6(4):251–266, 2000.

[53] Berthold KP Horn and Brian G. Schunck. Determining optical flow. *Real-Time Imaging*, 17(1–3):185–203, 1981.

[54] Berthold Horn. *Robot vision*. Massachusetts Institute of Technology press, Cambridge, Massachusetts, 1986.

[55] Thomas Brox, Andres Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *European conference on computer vision*, pages 25–36, 1999.

[56] George Wolberg. *Digital image warping*. IEEE computer society press, Los Alamitos, CA, 2009.

[57] Dilation (morphology). https://en.wikipedia.org/wiki/Dilation_(morphology). Accessed: 2016-3-4.