PolarMatte: Fully Computational Ground-Truth-Quality Alpha Matte Extraction for Images and Video using Polarized Screen Matting

Kenji Enomoto TJ Rhodes Brian Price Gavin Miller
Adobe Research

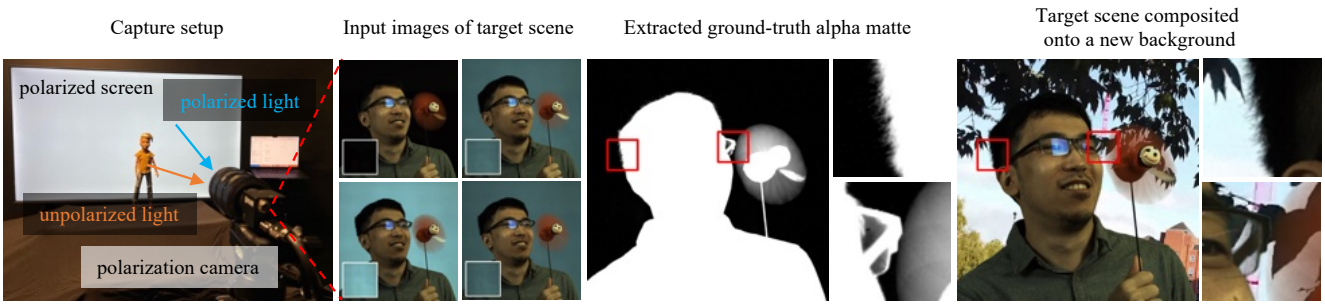

Figure 1. PolarMatte extraction method. A modern polarization camera with a color polarization filter array (CPFA) is used to capture four polarization states of a target scene and polarized background separately as input images. Our method extracts the alpha matte and foreground color which can be composited onto an arbitrary background. The highlighted regions show fine-details, glasses, and motion blur. We can computationally extract alpha mattes for dynamic scenes containing various materials without manual corrections.

Abstract

The creation of high-quality alpha mattes as ground-truth data for video matting is typically a laborious task. The trade-off between accuracy, manual corrections, and capture constraints often produces erroneous results or is cost prohibitive. We propose PolarMatte, a fully computational alpha matte extraction method for images and video without compromise between quality and practicality. A single polarization camera is used to capture dynamic scenes backlit by an off-the-shelf LCD monitor. PolarMatte exploits the polarization channel to compute the per-pixel opacity of the target scene, including the transparency of fine-details, translucent objects, and optical/motion blur. We leverage polarization clues to robustly detect indistinguishable pixels, and extract the alpha matte value at polarized foreground reflections with a polarimetric matting Laplacian. Quantitative and qualitative evaluation demonstrate our ability to computationally extract ground-truth-quality alpha mattes without human labour.

1. Introduction

Video matting has progressed immensely in recent history with the adoption of large neural networks [9, 11] trained on user-annotated datasets. However, the quantity and quality of dataset assets in the matting community restricts their accuracy and generalization ability [12]. The considerable effort required to accurately label alpha mattes for *ground-truth* is the primary cause for the lack of assets. This paper proposes a solution for the root cause and provides a pathway towards building large-scale video matting datasets.

Ground-truth alpha mattes for prior matting datasets have been created by photographing objects against multiple backgrounds [8, 21], extracting the mattes manually using existing matting methods [30, 31], or with chroma keying [8, 13, 28, 32]. These methods have a trade-off between quality and manual post-processing to produce the final matte. Such methods are impractical at scale and are prone to inherent errors. A recent work [24] can achieve high-quality results without manual keying, but the constrained lighting environment makes this less accessible for widespread use. These trade-offs impede the creation of accurate ground-truth for video matting datasets at scale.

Inspired by an earlier work using polarization for matting [15, 16], we propose an improved video matte capture method that eliminates the compromise between quality, effort, and setup. We present PolarMatte, a fully computational method for extracting ultra high-quality alpha mattes of dynamic scenes using polarization. A camera with a color polarization filter array (CPFA) is used to capture polarization images of a target object backlit with polarized light. While the foreground reflects the ambient light, the background light is polarized. We formulate the matting problem in the polarization channel since less polar-

ized light is observed from the foreground compared to the polarized background.

Our PolarMatte method accounts for polarized light reflected from the foreground, *polarized foreground reflections*, that cause perceptible alpha matte errors in a previous work that uses polarization [15]. Our method first detects well-posed pixels in a robust manner, and then optimizes the complete alpha matte using polarization clues and detected high-confidence alpha values. Pixels that show reflected light from the foreground are defined as *well-posed* if (1) the reflected light is unpolarized, or (2) the reflection is polarized but the phase angle differs sufficiently from the polarized backlight. Pixels are defined as *ill-posed* if the reflected light from the foreground is polarized and the phase angle of the reflection is comparable to the backlight.

The output of our alpha matte extraction process can directly be used as ground-truth for training datasets without manual post-processing. Our minimal setup uses a single polarization camera with a CPFA to capture dynamic scenes backlit by an off-the-shelf LCD monitor. This method allows us to extract ultra high-quality alpha mattes including fine-details, optical/motion blur, and translucent objects (see Fig. 1) without overly constrained capture conditions. We qualitatively and quantitatively compare our method with standard practices for alpha matte creation.

2. Related Work

High-quality ground-truth data is required for training accurate matting models. Matting datasets have been generated in several ways. The alphamatting.com dataset [21] was created by photographing static objects against multiple colored backgrounds to triangulate alpha values [25]. This provides accurate alpha mattes, though thresholding edges can be seen in their ground-truth mattes. While this has been applied to video [8], it requires stop-motion photography, which inherently does not support motion blur.

Manual extraction using existing matting methods has also been used. Closed-form matting [10] and KNN matting [6] were used to label real images [23]. The Deep Image Matting (DIM) dataset [31] extracted objects from images using Adobe Photoshop that can be composited onto new backgrounds. Semantic Image Matting [27] expanded DIM using the same method. While DIM has been used by many learning-based matting methods (*e.g.*, [5, 7, 14, 22, 29]), manually extracting mattes requires significant user time and skill, is error prone, and is biased by the algorithms used. Manual extraction was used for video to label 711 frames [30], an insufficient quantity for large-scale learning.

The ground-truth alpha matte in video matting datasets [13, 32] is often created by labeling green screen footage using commercial keying tools. However, chroma keying is an underconstrained problem that requires careful setup of the capture environment and considerable expertise for manual post-processing in the form of tweaking parameters, rotoscoping, and further adjustments to obtain a high-quality alpha matte.

While keying is less studied in the research community [4], a new keying technique was recently proposed called magenta green screen [24]. Similar to our method, magenta green screen uses an emissive backlight to extract the alpha matte. However, their constrained capture environment sacrifices the green channel, so studio dependent learning-based colorization is required to recover the foreground color. In contrast, PolarMatte uses the polarization channel, which does not modify the original color channels.

Our extraction method uses polarized screen matting, which extracts the alpha matte of a target scene backlit with a polarized screen. This setup was originally considered in polarimetric triangulation matting (PTM) [15, 16]. This approach uses a polarization camera system with a beamsplitter to capture two polarization images of a target scene in front of a polarized background screen. This formulates the matting problem as a simple arithmetic operation by assuming that all the light reflected from the foreground is unpolarized. However, in practice, the foreground can reflect polarized light depending on the material composition and surface orientation [20], which can yield erroneous alpha matte values for the foreground. In contrast to PTM, our method explicitly considers these polarized foreground reflections using four polarization images captured with a modern polarization camera that enables us to computationally extract an ultra high-quality alpha matte.

3. Background

Our method is motivated by a limitation of PTM [15]. To describe the issue theoretically, we briefly describe the research thread of PTM.

Triangulation Matting Image matting for an ordinary color image can be formulated as an inverse composition problem [19] described in

$$\mathbf{I} = \alpha\mathbf{F} + (1-\alpha)\mathbf{B}, \tag{1}$$

where the image intensities $\mathbf{I} \in \mathbb{R}^3_+$ are known, and foreground intensities $\mathbf{F} \in \mathbb{R}^3_+$, background intensities $\mathbf{B} \in \mathbb{R}^3_+$, and alpha matte $\alpha \in [0, 1]$ are unknown. Smith and Blinn proposed triangulation matting [25] that takes two images, $\mathbf{I}$ and $\mathbf{I}'$, of the same foreground against different known backgrounds $\mathbf{B}$ and $\mathbf{B}'$. It allows solving for the triangulation matte α_T as

$$\alpha_T = 1 - \frac{[\mathbf{I} - \mathbf{I}']}{[\mathbf{B} - \mathbf{B}']}, \tag{2}$$

where $[\cdot]$ is the averaging operator of a vector. While the triangulation matting problem is well-posed if $[\mathbf{B} - \mathbf{B}'] \neq 0$,

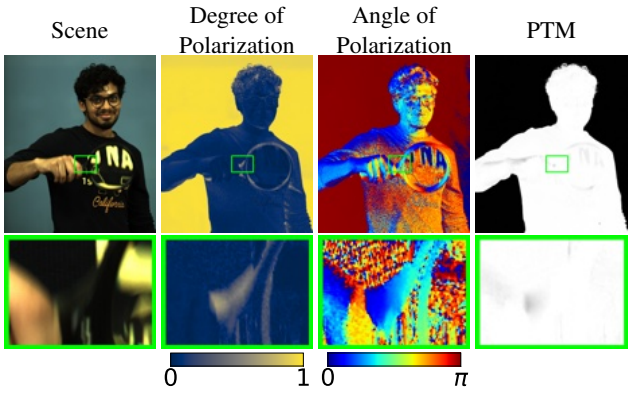

Figure 2. The alpha matte of PTM is inaccurately fractional if the foreground reflects polarized light, and the phase angle of the reflection is comparable to the polarized backlight. Zero and one of DoP indicate completely unpolarized and polarized, respectively.

it is only applicable for static scenes since the two images must be taken at different moments in time.

Polarimetric Triangulation Matting McGuire and Matusik [15] introduced passive polarization to acquire triangulation mattes in real-time using a camera with a polarizing beam-splitter and a polarized background. They assume that all the light reflected from the foreground is unpolarized under natural unpolarized illumination, and hence, the polarization image intensity $\mathbf{I}_\phi$ is described as

$$\mathbf{I}_\phi = \alpha \mathbf{F} + (1-\alpha)\,\mathbf{B}_\phi, \tag{3}$$

where $\mathbf{F}$ is the unpolarized foreground intensity, $\mathbf{B}_\phi$ is the polarized background intensity, and ϕ is the angle of the polarizing filter in front of the camera. Their capture system typically captures two $\frac{\pi}{2}$-shifted polarization images, and the triangulation matte can be computed as

$$\alpha_T = 1 - \frac{\left[\mathbf{I}_{\phi+\frac{\pi}{2}} - \mathbf{I}_\phi\right]}{\left[\mathbf{B}_{\phi+\frac{\pi}{2}} - \mathbf{B}_\phi\right]}, \tag{4}$$

similar to Eq. (2), but in a single-shot.

The assumption, that all the light reflected from the foreground, is unpolarized is not observed in practice. The reflected light can be partially polarized depending on the surface orientation and material composition [20] of the subject. This omission of polarized foreground reflections produces inaccuracies in the matte estimate that require correction and limit the practicality of the approach.

4. PolarMatte

We formulate the polarized screen matting problem using a more detailed compositing equation than Eq. (3):

$$\begin{aligned}\mathbf{I}_\phi &= \alpha \mathbf{F}_\phi + (1-\alpha)\,\mathbf{B}_\phi \\ &= \alpha\left(\mathbf{F} + \mathbf{P} \odot \cos^2\left(\boldsymbol{\theta}_f - \phi\mathbf{1}\right)\right) \\ &\quad + (1-\alpha)\left(\mathbf{B} + \mathbf{Q} \odot \cos^2\left(\boldsymbol{\theta}_b - \phi\mathbf{1}\right)\right),\end{aligned} \tag{5}$$

where $\mathbf{F}_\phi \in \mathbb{R}^3_+$ is the partially polarized foreground intensity. The partially polarized foreground and background intensities are further decomposed into unpolarized and polarized foreground intensities $\mathbf{F}, \mathbf{P} \in \mathbb{R}^3_+$, and unpolarized and polarized background intensities $\mathbf{B}, \mathbf{Q} \in \mathbb{R}^3_+$, respectively. The $\boldsymbol{\theta}_f \in \mathbb{R}^3$ and $\boldsymbol{\theta}_b \in \mathbb{R}^3$ are the phase angles of the light from the foreground and background. The operation of $\odot$ indicates element-wise multiplication. We call Eq. (5) the *polarimetric compositing equation*.

Suppose we have two polarization images $\mathbf{I}_\phi$ and $\mathbf{I}_{\phi+\frac{\pi}{2}}$, and corresponding polarized backgrounds $\mathbf{B}_\phi$ and $\mathbf{B}_{\phi+\frac{\pi}{2}}$. Let α_T and α be matte estimates based on Eq. (3) and (5), respectively. Then, the relation between α_T and α is

$$\alpha_T = \alpha\left(1 + \frac{\left[\mathbf{P} \odot \cos 2\left(\boldsymbol{\theta}_f - \phi\mathbf{1}\right)\right]}{\left[\mathbf{B}_{\phi+\frac{\pi}{2}} - \mathbf{B}_\phi\right]}\right). \tag{6}$$

Thus, if the foreground subject reflects polarized light, and the phase angle of the reflection is in the range of $\phi + \frac{\pi}{4} < \boldsymbol{\theta}_f < \phi + \frac{3\pi}{4}$, then the triangulation matte α_T for the foreground pixel will be less than 1, as highlighted in Fig. 2. This theoretical analysis motivates us to explicitly consider polarized foreground reflections for the extraction of ground-truth-quality alpha mattes.

Problem Statement A modern polarization camera with a color polarization filter array (CPFA) can capture four polarization images $\mathbf{I}_\phi, \phi \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ in a single-shot. These polarization images are often transformed into the Stokes parameters [26] for convenience. This paper only considers linear polarization, and the Stokes parameters can be described in the form of compositing equations:

$$\mathbf{s}_0 = \frac{1}{2}\sum_\phi \mathbf{I}_\phi = \alpha \underbrace{(2\mathbf{F} + \mathbf{P})}_{\mathbf{f}_0} + (1-\alpha)\underbrace{(2\mathbf{B} + \mathbf{Q})}_{\mathbf{b}_0} \tag{7}$$

$$\mathbf{s}_1 = \mathbf{I}_0 - \mathbf{I}_{\frac{\pi}{2}} = \alpha \underbrace{\mathbf{P} \odot \cos 2\boldsymbol{\theta}_f}_{\mathbf{f}_1} + (1-\alpha)\underbrace{\mathbf{Q} \odot \cos 2\boldsymbol{\theta}_b}_{\mathbf{b}_1} \tag{8}$$

$$\mathbf{s}_2 = \mathbf{I}_{\frac{\pi}{4}} - \mathbf{I}_{\frac{3\pi}{4}} = \alpha \underbrace{\mathbf{P} \odot \sin 2\boldsymbol{\theta}_f}_{\mathbf{f}_2} + (1-\alpha)\underbrace{\mathbf{Q} \odot \sin 2\boldsymbol{\theta}_b}_{\mathbf{b}_2} \tag{9}$$

Here, Eq. (7) is equivalent to Eq. (1) as the compositing equation for an ordinary image. Equations (8) and (9) are the additional equations obtained by considering polarization. Our goal is to find the optimal alpha matte α and foreground intensities $\mathbf{F}$ or $\mathbf{f}_0$, given $\mathbf{s}_i$ and $\mathbf{b}_i, i \in \{0, 1, 2\}$.

4.1. Well-posed Pixel Detection

The PolarMatte problem is ill-posed; however, the problem becomes well-posed if the foreground reflects unpolarized light or the phase angles of light from the target scene and known background differ significantly. Here, we describe a robust method for finding the well-posed pixels and extracting high-confidence alpha matte values for these pixels.

Least Squares Solution We formulate the PolarMatte problem as a minimization problem of polarized foreground intensity $\|\alpha\mathbf{P}\|_2^2$. For clarification, we denote $\mathbf{s} = \left[\mathbf{s}_1^\top, \mathbf{s}_2^\top\right]^\top$, $\mathbf{f} = \left[\mathbf{f}_1^\top, \mathbf{f}_2^\top\right]^\top$, $\mathbf{b} = \left[\mathbf{b}_1^\top, \mathbf{b}_2^\top\right]^\top$. Since $\|\alpha\mathbf{P}\|_2^2 = \|\alpha\mathbf{f}\|_2^2$, the minimization problem of $\|\alpha\mathbf{P}\|_2^2$ can be described as

$$\min_{\alpha} \|\alpha\mathbf{P}\|_2^2 = \min_{\alpha} \|\mathbf{s} - (1-\alpha)\,\mathbf{b}\|_2^2. \tag{10}$$

This is a least squares problem and has a closed-form solution for α as follows:

$$\widehat{\alpha} = 1 - \frac{\mathbf{b}^\top \mathbf{s}}{\mathbf{b}^\top \mathbf{b}}. \tag{11}$$

The value of $\|\widehat{\alpha}\widehat{\mathbf{P}}\|_2 = \|\mathbf{s} - (1-\widehat{\alpha})\,\mathbf{b}\|_2$ is zero if the problem is well-posed. Therefore, we evaluate the uncertainty of the estimated alpha matte by $\|\widehat{\alpha}\widehat{\mathbf{P}}\|_2$ and define well-posed pixels as pixels that satisfy $\|\widehat{\alpha}\widehat{\mathbf{P}}\|_2 < \tau_0$.

In theory, the thresholding described above finds all well-posed pixels. However, in practice, we found that the value of $\|\widehat{\alpha}\widehat{\mathbf{P}}\|_2$ can be non-zero even at background pixels due to imaging noise, and a large threshold degrades the reliability of the well-posed pixel detection. Therefore, we introduce an extension to a Bayesian framework to increase the robustness of well-posed pixel detection.

Extension to Robust Bayesian Framework Here, we reformulate the problem with a maximum a posteriori (MAP) estimator. Although we assume that a background $\mathbf{b}$ is pre-captured, the background intensity of $\mathbf{s}$ is always slightly deviated from the pre-captured $\mathbf{b}$ in practice. Motivated by this, we formulate the problem as the maximization problem of a probability distribution $P(\alpha, \mathbf{b}|\mathbf{s})$ described as

$$\begin{aligned}\underset{\alpha,\mathbf{b}}{\operatorname{argmax}}\, P(\alpha, \mathbf{b}|\mathbf{s}) &= \underset{\alpha,\mathbf{b}}{\operatorname{argmax}}\, \frac{P(\mathbf{s}|\alpha, \mathbf{b})\,P(\alpha)\,P(\mathbf{b})}{P(\mathbf{s})} \\ &= \underset{\alpha,\mathbf{b}}{\operatorname{argmax}}\, L(\mathbf{s}|\alpha, \mathbf{b}) + L(\mathbf{b}),\end{aligned} \tag{12}$$

where $L(\cdot) = \log P(\cdot)$ is the log likelihood. We drop the terms of $P(\mathbf{s})$ and $P(\alpha)$ because we make no assumption of the prior for $\mathbf{s}$ and α, and $P(\mathbf{s})$ and $P(\alpha)$ are constant.

The log likelihood $L(\mathbf{s}|\alpha, \mathbf{b})$ measures the fitness of the estimates $\{\alpha, \mathbf{b}\}$ to an observed polarization intensity $\mathbf{s}$ by

$$L(\mathbf{s}|\alpha, \mathbf{b}) = -\frac{1}{2\sigma_s^2}\|\mathbf{s} - (1-\alpha)\,\mathbf{b}\|_2^2, \tag{13}$$

where σ_s is a noise variance of $\mathbf{s}$. We model $P(\mathbf{b})$ with a multivariate Gaussian distribution centered at the pre-captured background, hence, the log likelihood can be described as

$$L(\mathbf{b}) = -\frac{1}{2}\left(\mathbf{b} - \bar{\mathbf{b}}\right)^\top \boldsymbol{\Sigma}_b^{-1}\left(\mathbf{b} - \bar{\mathbf{b}}\right), \tag{14}$$

where $\bar{\mathbf{b}} \in \mathbb{R}^6$ and $\boldsymbol{\Sigma}_b \in \mathbb{R}^{6\times 6}$ are the mean vector and diagonal covariance matrix. We pre-capture background images several times and compute the element-wise mean and variance of $\mathbf{b}$ for $\bar{\mathbf{b}}$ and the diagonal elements of $\boldsymbol{\Sigma}_b$, respectively. Taking the partial derivative of Eq. (12) with respect to α and $\mathbf{b}$ and equating them to zero results in

$$\alpha = 1 - \frac{\mathbf{b}^\top \mathbf{s}}{\mathbf{b}^\top \mathbf{b}}, \tag{15}$$

$$\mathbf{b} = \left(\frac{(1-\alpha)^2}{\sigma_s^2}\mathbf{E}_6 + \boldsymbol{\Sigma}_b^{-1}\right)^{-1}\left(\frac{1-\alpha}{\sigma_s^2}\mathbf{s} + \boldsymbol{\Sigma}_b^{-1}\bar{\mathbf{b}}\right), \tag{16}$$

where $\mathbf{E}_6$ is the 6×6 identity matrix. We iteratively compute α and $\mathbf{b}$ using Eqs. (15) and (16) and estimate alpha matte $\widehat{\alpha}$. We then evaluate the value of $\|\widehat{\alpha}\widehat{\mathbf{P}}\|_2$ to find well-posed pixels.

Phase Angle Thresholding The thresholding of minimized polarized foreground intensity finds well-posed pixels with approximately unpolarized foreground reflection, and the remaining pixels likely contain polarized reflection.

Here, we introduce another thresholding based on the phase angle deviation. When the phase angles of an observed image $\mathbf{I}_\phi$ and pre-captured background $\mathbf{B}_\phi$ differ significantly, the pixel is likely to be foreground. Therefore, foreground pixels can be found by the thresholding,

$$\frac{1}{2}\cos^{-1}\frac{\mathbf{s}^\top \mathbf{b}}{\|\mathbf{s}\|_2\|\mathbf{b}\|_2} > \tau_1, \tag{17}$$

where τ_1 is a threshold. We make $\widehat{\alpha} = 1$ for the pixels that satisfy this thresholding.

4.2. Alpha Matte Unification

We have discussed well-posed pixel detection thus far. This section describes the complete alpha matte extraction by unifying the alpha matte estimates for well-posed pixels, polarization clues, and classic color clues using our proposed *polarimetric matting Laplacian*.

Polarimetric Matting Laplacian Our polarimetric matting Laplacian is motivated by the original matting Laplacian [10] that is derived from a local linear model:

$$\alpha_i = \mathbf{a}_k^\top \mathbf{I}_i + b_k, \quad \forall i \in \mathcal{W}_k, \tag{18}$$

where α_i and $\mathbf{I}_i \in \mathbb{R}^3_+$ are the alpha matte and image intensities of the i-th pixel in a local window $\mathcal{W}_k$. The coefficient vector $\mathbf{a}_k \in \mathbb{R}^3$ and bias $b_k \in \mathbb{R}$ are constant over the local window. By minimizing the squared error of the linear model, Eq. (18), over every local window in an image using the linear ridge regression model, Levin *et al*. proved that the cost function can be summarized in the quadratic form,

$$J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \mathbf{L} \boldsymbol{\alpha}. \tag{19}$$

Here, $\boldsymbol{\alpha} \in \mathbb{R}^N$ is a vector of alpha matte values for all N pixels in an image and $\mathbf{L} \in \mathbb{R}^{N \times N}$ is called a matting Laplacian matrix, which can be computed solely from observed image intensities.

To leverage the polarization clues, we design a Laplacian matrix using the vector $\mathbf{s}$. Similarly to the original matting Laplacian, we assume the local linear model Eq. (18) with $\mathbf{s}_i$, which is the vector $\mathbf{s}$ of the i-th pixel in a local window, instead of the vector $\mathbf{I}_i$. Hence, our cost function can also be described as

$$J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \mathbf{L}_p \boldsymbol{\alpha}. \tag{20}$$

We refer to the matrix $\mathbf{L}_p \in \mathbb{R}^{N \times N}$ as polarimetric matting Laplacian, whose (i,j)-th element is

$$\sum_{k:(i,j)\in\mathcal{W}_k} \left(\delta_{ij} - \frac{1}{|\mathcal{W}_k|} \left(1 + (\mathbf{s}_i - \boldsymbol{\mu}_k)^\top \left(\boldsymbol{\Sigma}_k + \frac{\epsilon}{|\mathcal{W}_k|} \mathbf{E}_6 \right)^{-1} (\mathbf{s}_j - \boldsymbol{\mu}_k) \right) \right), \tag{21}$$

where $\boldsymbol{\mu}_k \in \mathbb{R}^6$ and $\boldsymbol{\Sigma}_k \in \mathbb{R}^{6\times 6}$ are the mean vector and covariance matrix of $\mathbf{s}$ in a local window $\mathcal{W}_k$, respectively. δ_{ij} is the Kronecker delta, and ϵ is a small positive constant.

While the original matting Laplacian implicitly assumes foreground and background color vectors in a local window lie in a low-dimensional space, *e.g*., color line model [18], our model implicitly assumes the foreground and background Stokes parameters, $\mathbf{f}$ and $\mathbf{b}$, in a local window lie in a low-dimensional space, *e.g*., their color satisfies the color line model, and phase angles are uniform (see the proof in Sec. C). Although the matting Laplacian matrices are similar in concept, the matting accuracy of our polarimetric matting Laplacian is superior. The differences will be discussed in Sec. 5.5 with a spectral analysis of the Laplacian matrices.

Total Optimization We define our total cost function with the polarimetric matting Laplacian, the constraint of the well-posed alpha matte value estimates, and a weak supervision of the least squares objective in Eq. (10), which can be described as

$$\begin{aligned}\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\, & \boldsymbol{\alpha}^\top \mathbf{L}_p \boldsymbol{\alpha} + \lambda (\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}})^\top \mathbf{D} (\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}) \\ & + \boldsymbol{\alpha}^\top \mathbf{W}\mathbf{V}\boldsymbol{\alpha} + 2\mathbf{v}^\top \mathbf{W}\boldsymbol{\alpha}.\end{aligned} \tag{22}$$

The first term is the cost function with the polarimetric matting Laplacian. The second term constrains the solution to match the well-posed alpha matte value estimates obtained in Sec. 4.1, where λ is a large positive number, $\widehat{\boldsymbol{\alpha}}$ is a vector of well-posed alpha matte value estimates, and $\mathbf{D} \in \mathbb{R}^{N\times N}$ is a diagonal matrix whose diagonal elements are 1 for well-posed pixels and 0 for other pixels. The third term weakly supervises the solution by the least squares solution in case there are no clues from the first and second terms. Here, $\mathbf{W} \in \mathbb{R}^{N\times N}$ is a diagonal weighing matrix whose diagonal elements are $\exp\left(-\kappa \|\widehat{\alpha}\widehat{\mathbf{P}}\|_2\right)$ per pixel and κ is a large number. $\mathbf{V}$ is a diagonal matrix whose diagonal elements are $\mathbf{b}^\top\mathbf{b}$, and $\mathbf{v}$ is the vector of $\mathbf{b}^\top(\mathbf{s}-\mathbf{b})$ for each pixel. The solution to this total cost function is obtained by solving the linear system,

$$(\mathbf{L}_p + \lambda\mathbf{D} + \mathbf{W}\mathbf{V})\,\boldsymbol{\alpha} = \lambda\mathbf{D}\widehat{\boldsymbol{\alpha}} - \mathbf{W}\mathbf{v}. \tag{23}$$

4.3. PolarIC Adjustment

Our method is based on the polarization image described in Eq. (5), which implicitly assumes a linear constraint,

$$I_0 - I_{\frac{\pi}{4}} + I_{\frac{\pi}{2}} - I_{\frac{3\pi}{4}} = 0. \tag{24}$$

However, this constraint is rarely satisfied in real images even after linearization of the camera's response function as illustrated in Fig. 3. Deviation from this constraint causes errors in matte estimation. We introduce an adjustment step called polarimetric intensity correction (PolarIC) to minimally adjust the input polarization image intensities so that the implicit constraint from Eq. (24) is satisfied.

We employ a pixel-wise linear correction model,

$$I'_\phi = c_\phi I_\phi, \tag{25}$$

where c_ϕ is the correction coefficient for each polarization state, and I'_ϕ is the corrected image intensity. The desirable solution for the coefficients makes the minimum correction on the image intensities to satisfy the constraint. With denoting $\mathbf{c} = \left[c_0, c_{\frac{\pi}{4}}, c_{\frac{\pi}{2}}, c_{\frac{3\pi}{4}}\right]^\top$ and $\mathbf{m} = \left[I_0, -I_{\frac{\pi}{4}}, I_{\frac{\pi}{2}}, -I_{\frac{3\pi}{4}}\right]^\top$, we define the objective function as

$$\mathbf{c}^* = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{1} - \mathbf{c}\|_2^2 \quad \text{s.t. } \mathbf{m}^\top\mathbf{c} = 0. \tag{26}$$

This objective is a linear regression problem with an equality constraint and has the closed-form solution described as

$$\mathbf{c}^* = \mathbf{1} - \left(\mathbf{1}^\top \widetilde{\mathbf{m}}\right)\widetilde{\mathbf{m}} \tag{27}$$

with denoting $\widetilde{\mathbf{m}} = \mathbf{m}/\|\mathbf{m}\|_2$. The proof is described in Sec. B. We perform this PolarIC adjustment on all color channels independently.

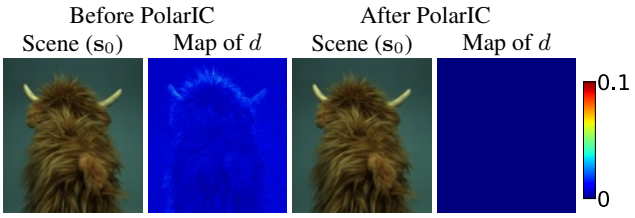

Figure 3. The values of $d = |I_0 - I_{\frac{\pi}{4}} + I_{\frac{\pi}{2}} - I_{\frac{3\pi}{4}}|$ before and after applying our PolarIC adjustment.

5. Experimental Evaluation

We present visual results of our PolarMatte method along with quantitative and qualitative comparisons with existing matting and keying techniques on various subjects. We further discuss the role of each step in our process, including a spectral analysis of our polarimetric matting Laplacian. Additional experimental results (Sec. G, H, and I), three-channel alpha matte extraction (Sec. J), and video results can be found in the supplementary materials.

5.1. Implementation Details

Hardware Setup We use the Blackfly® S USB3 color polarization camera [1] and an LCD monitor [3] as a polarized screen as shown in Fig. 1. The camera's configuration was set to 30 frames per second, $1/30$ second exposure time, and $\mathrm{f}/1.4$. We measured the camera's response function once using a color checker board [2] to linearize the response function.

Initial Adjustment Given a RAW image, we perform polarization demosaicing tailored for CPFAs [17] and our PolarIC adjustment. We capture background images 30 times and compute the mean vector $\bar{\mathbf{b}}$ and covariance matrix Σ_b for the Bayesian framework before capturing a target scene.

Parameters For our well-posed pixel detection, we set $\sigma_s = \frac{5}{255}$, $\tau_0 = \frac{3}{255}$, and $\tau_1 = \frac{\pi}{4}$. We compute the polarimetric matting Laplacian with $\epsilon = 10^{-7}$ and 3×3 local window. The parameters λ and κ for the total optimization are set to 100 and 1000, respectively.

Test Data For quantitative evaluations, we collected 12 static scene test data including fur, solid and translucent objects, and objects with optical blur. We captured three sets of polarization images with white, black, and blue backgrounds for each scene. The baseline alpha matte was obtained with triangulation matting [25] using the white and black backgrounds. While the triangulation matte may deviate from the true alpha matte due to imaging noise and backlight reflection on the foreground, it is a well-posed solution with strong results, so we use it as the *reference* for evaluations. The polarization image with a blue background is used for chroma keying. See Sec. E in the supplementary material for more details of the test data collection.

Comparative methods We use polarimetric triangulation matting (PTM) [15], background matting (BGM) [13], and Keylight implemented in Adobe After Effects for the comparison with our PolarMatte method. For PTM, we use the two images with the brightest and darkest background from the four polarization images. For BGM, we use $\mathbf{s}_0$ and $\mathbf{b}_0$ as input. While Keylight has several interactively adjustable parameters, we only specify the key color and leave the other parameters set to their default value for comparison with our computational matting method.

Evaluation Metrics We employ three commonly used metrics for evaluating the estimated alpha matte: mean absolute error (MAE), mean squared error (MSE), and gradient error (GRAD).

Table 1. Quantitative evaluations of PTM [15], BGM [13], chroma keying in Keylight, and PolarMatte on the static scene test data.

	MAE ↓	MSE ↓	GRAD ↓
PTM	8.38	0.33	1.03
BGM	137.75	110.43	21.96
Keylight	84.45	84.45	6.13
PolarMatte	**7.65**	**0.28**	**0.72**

5.2. Comparison on Static Scene Test Data

Table 1 shows the quantitative evaluations of PTM, BGM, Keylight, and our PolarMatte method. Our method outperforms the comparative methods across all metrics. As stated in Eq. (6), the error in PTM is caused by foreground pixels where the light reflected is polarized and the phase angle of the reflection is comparable to the phase angle of the polarized backlight. This error in PTM is observed in the visual results shown in Fig. 4. In contrast to PTM, PolarMatte correctly extracts alpha matte values by explicitly considering the polarized foreground reflection. The errors in BGM are mainly caused by complex cases such as translucent objects and fine-details as shown in Fig. 4. BGM struggles with these cases because matting on an ordinary color image is heavily ill-posed at every pixel even if the background is known. In contrast, PolarMatte accurately estimates the alpha matte on these scenes by leveraging the polarization clues. A typical failure with Keylight can be observed around the boundary of objects, particularly fine-details. Since common chroma keying tools including Keylight do not use a known background, they struggle to estimate accurate fractional alpha matte values. Keylight alpha mattes may be improved with manual adjustments, particularly errors due to background non-uniformity; however, the primary advantage of PolarMatte is how the results are produced without any human interaction.

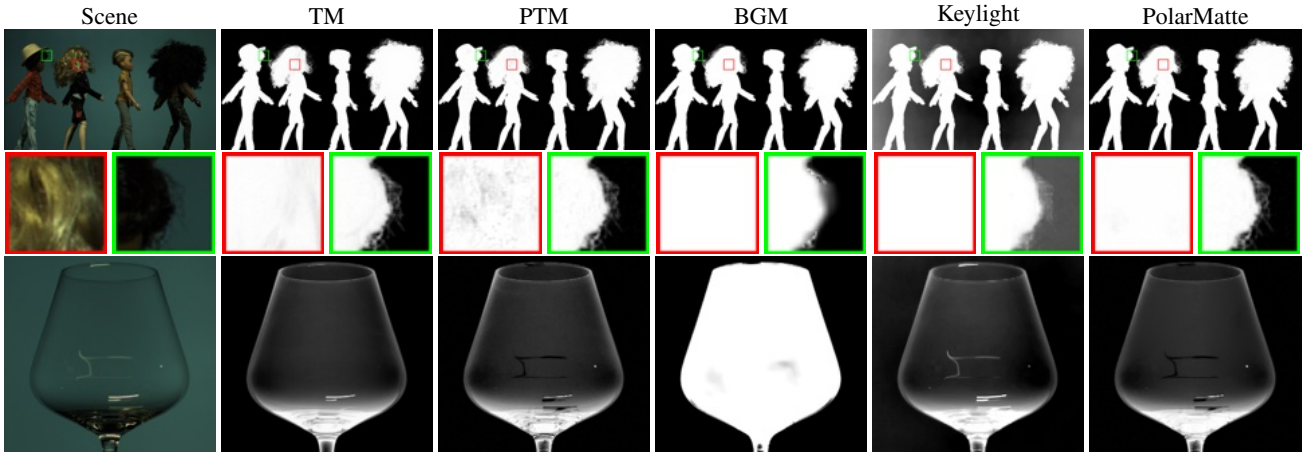

Figure 4. Visual results of the baseline alpha matte computed by triangulation matting (TM) [25], polarimetric triangulation matting (PTM) [15], background matting (BGM) [13], chroma keying in Keylight, and our PolarMatte method on the static scene test data.

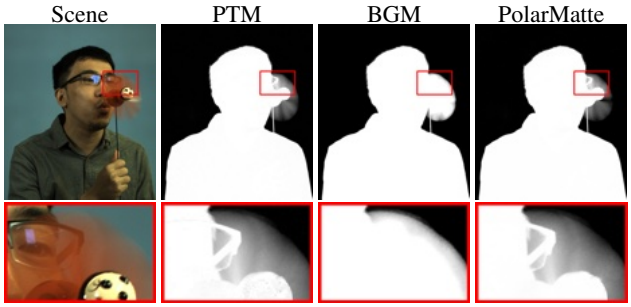

Figure 5. Visual result of PTM [15], BGM [13], and PolarMatte on a dynamic scene.

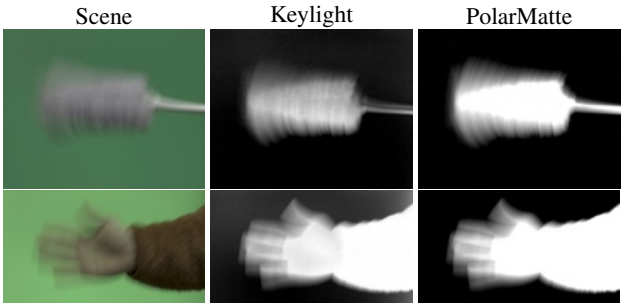

Figure 6. Visual results of chroma keying in Keylight and PolarMatte on dynamic scenes.

5.3. Comparison on Dynamic Scenes

The matting capability of our PolarMatte method is not limited to static scenes. Figure 5 shows a visual result of PTM, BGM, and PolarMatte on a dynamic scene. These results demonstrate that our method can extract accurate alpha mattes for motion blur, while BGM fails with motion blur. PTM also successfully extracts the alpha matte for motion blur, however, PTM inaccurately shows fractional alpha matte values in solid regions as shown in Figs. 4 and 5.

Figure 6 shows visual results of Keylight and PolarMatte on dynamic scenes. While Keylight offers favorable alpha mattes for motion blur, the tool struggles with balancing accuracy between foreground, background, and mixed pixels without manual post-processing. More importantly, the fractional alpha matte values from Keylight are not guaranteed to be physically correct since, like BGM, chroma keying is similarly ill-posed. The dynamic scene experiment demonstrates the superior matting ability of our PolarMatte method on fine-details, translucent objects, and motion blur for video. Additional video results can be found in the supplementary material.

5.4. Ablation Study

We investigate the role of each step in our process, including PolarIC adjustment, Bayesian framework, phase angle thresholding, and total optimization with the polarimetric matting Laplacian. We conduct an ablation study on the test data, where we run our method without each component.

Table 2 shows the results of the ablation study. These results indicate that every step of our method contributes to the matting accuracy. The phase angle thresholding significantly contributes to the MAE and MSE because a large number of pixels can be classified as foreground in this step. The use of the polarimetric matting Laplacian contributes to the GRAD score since the total optimization with polarimetric matting Laplacian can correct noisy foreground alpha matte values due to polarized foreground reflections.

5.5. Matting Laplacian Comparison

As validated in Sec. 5.4, the polarimetric matting Laplacian is a key component of our method. While the polarimetric matting Laplacian is inspired by the original matting Laplacian [10], our Laplacian produces superior results.

We first compare the matting accuracy of each Laplacian in our matting optimization Eq. (23). Table 3 shows the re-

Table 2. Ablation study of PolarMatte on the test data. The experiments are performed without one of our method's components: PolarIC adjustment, Bayesian framework, phase angle thresholding (PA thresh.), and polarimetric matting Laplacian (PML).

	MAE ↓	MSE ↓	GRAD ↓
PolarMatte w/o PolarIC	8.78	0.45	0.93
PolarMatte w/o Bayesian	8.53	0.43	0.89
PolarMatte w/o PA thresh.	8.85	0.49	1.00
PolarMatte w/o PML	8.41	0.34	1.04
PolarMatte	**7.65**	**0.28**	**0.72**

Table 3. Comparison between our polarimetric matting Laplacian $\mathbf{L}_p$ and the original matting Laplacian $\mathbf{L}$ on the test data.

	MAE ↓	MSE ↓	GRAD ↓
PolarMatte w/ $\mathbf{L}$	9.89	0.62	2.22
PolarMatte w/ $\mathbf{L}_p$	**7.65**	**0.28**	**0.72**

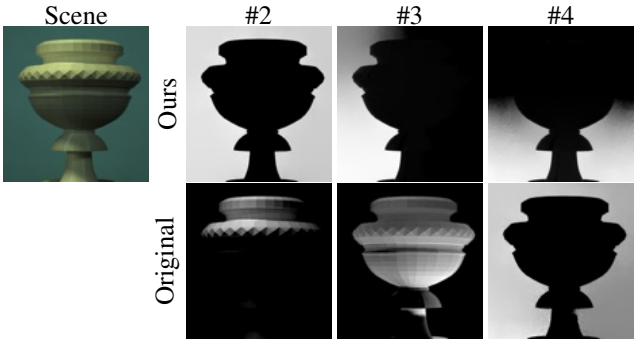

Figure 7. Spectral analysis of our polarimetric matting Laplacian and the original matting Laplacian [10]. Figures show the second, third, and fourth smallest eigenvectors of each matting Laplacian (the first smallest eigenvector is constant for both matrices).

sults on the test data, which indicate the improvement of our polarimetric matting Laplacian to the original matting Laplacian. The difference in matting accuracy can be explained by a spectral analysis of both matrices. Figure 7 shows the second, third, and fourth smallest eigenvectors of our polarimetric matting Laplacian and the original matting Laplacian that primarily affect the final matte estimate. The second smallest eigenvector of our polarimetric matting Laplacian shows nearly correct alpha matte features, which indicates the well-posedness of the PolarMatte problem. In contrast, the eigenvectors of the original matting Laplacian show the texture of the foreground subject, which suggests that the original matting Laplacian requires known alpha matte values at many pixels. This spectral analysis validates that the use of polarization clues improves matting accuracy.

5.6. Limitations

The experiments so far have shown the ability of our method to extract ultra high-quality alpha mattes for dynamic scenes containing diverse subjects without manual corrections. However, our method may estimate incorrect alpha matte values if the capturing environment is unsuitable. Specifically, if the polarized foreground reflection resembles the polarized backlight, Eq. (10) can be solved with $\|\widehat{\alpha}\widehat{\mathbf{P}}\|_2 = 0$, resulting with an incorrect alpha matte value that is less than 1. This undesirable case may occur when the foreground subject and polarized background screen are close in proximity. This type of error is common in backlight-based matting [4], including triangulation matting and PolarMatte, as shown in Fig. 8. We encourage anyone who implements PolarMatte to maintain sufficient distance between the foreground subject and polarized background screen to avoid this phenomenon.

Figure 8. A failure case of backlight-based matting methods including triangulation matting (TM) [25] and PolarMatte. The alpha matte estimate can be incorrectly fractional when background light is reflected on the foreground subject.

Another limitation of our method is its computational cost. We measured processing time on Intel® Xeon® Platinum 8275CL CPU @ 3.00 GHz, and PolarMatte required several tens of seconds to extract the alpha matte of each image in the test data, while PTM took only one second or less to process. While the computational cost of PolarMatte is acceptable for creating a matting dataset, it is unrealistic for real-time background replacement. We seek to improve the computational cost of our method in the future.

6. Conclusion

In this paper, we presented PolarMatte, a fully computational alpha matte extraction method that produces ultra high-quality alpha mattes for images and video that can directly be used as ground-truth for training datasets. By leveraging the polarization clues from our minimal setup, we can robustly detect indistinguishable pixels, and overcome the issue of polarized foreground reflections with the proposed polarimetric matting Laplacian. The evaluation of our method with standard practices and other matting techniques on real-world static and dynamic scenes demonstrates our ability to computationally extract ultra high-quality alpha mattes for fine-details, motion and optical blur, and translucent objects without any manual operations.

7. Acknowledgements

We would like to thank Scott Cohen, Nathan Carr, Michael Gharbi, and Kevin Wampler for their contributions.

References

[1] BFS-U3-51S5P-C USB 3.1. Accessed on Oct. 30, 2023. 6

[2] Calibrite ColorChecker Passport Duo. Accessed on Oct. 30, 2023. 6

[3] 55" 4K UHD Hisense Android Smart TV (2021). Accessed on Oct. 30, 2023. 6

[4] Yağiz Aksoy, Tunç Ozan Aydin, Marc Pollefeys, and Aljoša Smolić. Interactive high-quality green-screen keying via color unmixing. *ACM TOG*, 36(4):1, 2016. 2, 8

[5] Shaofan Cai, Xiaoshuai Zhang, Haoqiang Fan, Haibin Huang, Jiangyu Liu, Jiaming Liu, Jiaying Liu, Jue Wang, and Jian Sun. Disentangled image matting. In *ICCV*, 2019. 2

[6] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang. Knn matting. *IEEE TPAMI*, 2013. 2

[7] Yutong Dai, Brian Price, He Zhang, and Chunhua Shen. Boosting robustness of image matting with context assembling and strong data augmentation. In *CVPR*, 2022. 2

[8] Mikhail Erofeev, Yury Gitman, Dmitriy Vatolin, Alexey Fedorov, and Jue Wang. Perceptually motivated benchmark for video matting. In *BMVC*, 2015. 1, 2

[9] Wei-Lun Huang and Ming-Sui Lee. End-to-end video matting with trimap propagation. In *CVPR*, 2023. 1

[10] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE TPAMI*, 30(2):228–242, 2008. 2, 4, 7, 8, 9, 10

[11] Jiachen Li, Vidit Goel, Marianna Ohanyan, Shant Navasardyan, Yunchao Wei, and Humphrey Shi. VMFormer: End-to-end video matting with transformer. *arXiv preprint arXiv:2208.12801*, 2022. 1

[12] Jizhizi Li, Jing Zhang, and Dacheng Tao. Deep image matting: A comprehensive survey. *arXiv preprint arXiv:2304.04672*, 2023. 1

[13] Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian L Curless, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Real-time high-resolution background matting. In *CVPR*, 2021. 1, 2, 6, 7

[14] Hao Lu, Yutong Dai, Chunhua Shen, and Songcen Xu. Indices matter: Learning to index for deep image matting. In *ICCV*, 2019. 2

[15] Morgan McGuire and Wojciech Matusik. Real-time triangulation matting using passive polarization. In *SIGGRAPH Sketches*. 2006. 1, 2, 3, 6, 7

[16] Morgan McGuire, Wojciech Matusik, and William Yerazunis. Practical, Real-time Studio Matting using Dual Imagers. In *Symposium on Rendering*, 2006. 1, 2

[17] Miki Morimatsu, Yusuke Monno, Masayuki Tanaka, and Masatoshi Okutomi. Monochrome and color polarization demosaicking based on intensity-guided residual interpolation. *IEEE Sensors Journal*, 21(23):26985–26996, 2021. 6

[18] Ido Omer and Michael Werman. Color lines: Image specific color representation. In *CVPR*, 2004. 5

[19] Thomas Porter and Tom Duff. Compositing digital images. *SIGGRAPH*, 1984. 2

[20] Stefan Rahmann and Nikos Canterakis. Reconstruction of specular surfaces using polarization imaging. In *CVPR*, 2001. 2, 3

[21] Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, and Pamela Rott. A perceptually motivated online benchmark for image matting. In *CVPR*, 2009. 1, 2

[22] Soumyadip Sengupta, Vivek Jayaram, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman. Background matting: The world is your green screen. In *CVPR*, 2020. 2

[23] Xiaoyong Shen, Xin Tao, Hongyun Gao, Chao Zhou, and Jiaya Jia. Deep automatic portrait matting. In *BMVC*, 2016. 2

[24] Dmitriy Smirnov, Chloe LeGendre, Xueming Yu, and Paul Debevec. Magenta green screen: Spectrally multiplexed alpha matting with deep colorization. In *The Digital Production Symposium (DigiPro)*, 2023. 1, 2

[25] Alvy Ray Smith and James F Blinn. Blue screen matting. In *SIGGRAPH*, 1996. 2, 6, 7, 8

[26] George Gabriel Stokes. On the composition and resolution of streams of polarized light from different sources. *Transactions of the Cambridge Philosophical Society*, 1851. 3

[27] Yanan Sun, Chi-Keung Tang, and Yu-Wing Tai. Semantic image matting. In *CVPR*, 2021. 2

[28] Yanan Sun, Guanzhi Wang, Qiao Gu, Chi-Keung Tang, and Yu-Wing Tai. Deep video matting via spatio-temporal alignment and aggregation. In *CVPR*, 2021. 1

[29] Jingwei Tang, Yagiz Aksoy, Cengiz Oztireli, Markus Gross, and Tunc Ozan Aydin. Learning-based sampling for natural image matting. In *CVPR*, 2019. 2

[30] Tiantian Wang, Sifei Liu, Yapeng Tian, Kai Li, and Ming-Hsuan Yang. Video matting via consistency-regularized graph neural networks. In *ICCV*, 2021. 1, 2

[31] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang. Deep image matting. In *CVPR*, 2017. 1, 2

[32] Yunke Zhang, Chi Wang, Miaomiao Cui, Peiran Ren, Xuansong Xie, Xian-Sheng Hua, Hujun Bao, Qixing Huang, and Weiwei Xu. Attention-guided temporally coherent video object matting. In *ACM MM*, 2021. 1, 2