Adaptive Polarization-Difference Transient Imaging for Depth Estimation in Scattering Media

RIHUI WU¹,³, ADRIAN JARABO†⁴, JINLI SUO²,⁵,*, FENG DAI³, YONGDONG ZHANG³, QIONGHAII DAI²,⁵, and DIEGO GUTIERREZ‡

¹University of Chinese Academy of Sciences, Beijing 100049, China
²Department of Automation, Tsinghua University, Beijing 100084, China
³Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
⁴Universidad de Zaragoza - IBA, Spain
⁵Beijing Key Laboratory of Multi-dimension & Multi-scale Computational Photography (MMCP), Tsinghua University, Beijing 100084, China
*Corresponding author: jlsuo@tsinghua.edu.cn

Compiled May 4, 2018

Introducing polarization into transient imaging improves depth estimation in participating media, by discriminating reflective from scattered light transport, and calculating depth from the former component only. Previous works have leveraged this approach, under the assumption of uniform polarization properties. However, the orientation and intensity of polarization inside scattering media is non-uniform, both in the spatial and temporal domains. As a result of this simplifying assumption, the accuracy of the estimated depth worsens significantly as the optical thickness of the medium increases. In this letter, we introduce a novel adaptive polarization-difference method for transient imaging, taking into account the non-uniform nature of polarization in scattering media. Our results demonstrate a superior performance for impulse-based transient imaging over previous unpolarized or uniform approaches. © 2018 Optical Society of America

OCIS codes:  (010.1350) Backscattering; (110.5405) Polarimetric imaging; (150.0155) Machine vision optics; (110.1758) Computational imaging.

http://dx.doi.org/10.1364/ao.XXXXX

Transient imaging has recently allowed researchers to capture and visualize large-scale light transport effects at extreme temporal resolutions. Representative transient imaging setups include a streak camera [1], single-photon avalanche detectors (SPADs) [2], Michelson interferometer [4], amplitude modulated continuous wave (AMCW) systems [3, 5], laser-gated system [6], or hybrid sensor systems [7]. This development of transient imaging has in turn triggered many applications in scene understanding, computer vision, or computer graphics (see [8] for an in-depth survey on the field). For example, it is possible to estimate scene depth from the first peak of the temporal profiles that describe surface reflection. However, these transient imaging setups fail in the presence of participating media, due to the large number of scattering effects.

As light travels through participating media, the light reflected from surfaces and the light scattered in the medium (including single and multiple scattering) become mixed in a complex manner; it thus becomes difficult to extract the surface reflection component, which encodes scene depth. Several setups have been proposed to improve visibility in the presence of scattering media for steady-state [11, 12], or transient imaging [6, 9, 13]. Within these setups, the use of polarization has been proven to be a powerful approach for improving visibility in media. Polarization-difference imaging (PDI) takes two frames in mutually orthogonal polarization states (I_{max} and I_{min} respectively) to analyze the scattering effect of the medium [14, 15]. Ni and Alfano [10] used both circularly and linearly polarized light to improve contrast in time-resolved imaging in turbid media. Wu et al. [16] proposed to incorporate polarization into time-of-flight AMCW sensors, achieving robust depth estimation. However, since their method relies on Treibitz and Schechner’s [15] method for dehazing, it assumes radiance integrated in time, and therefore cannot be used with direct recording setups, e.g. streak cameras or SPADs. More importantly, the authors assume uniform linear polarization across the whole image. This assumption does not hold in most real scenes [17], due to spatially-varying view and light directions, different amount of scattering, etc. This is especially noticeable in the transient regime, where each pixel-frame has a different degree and orientation of linear polarization. As a result, it fails to estimate depth in the presence of optically-thick media.

Fig. 1 illustrates the non-uniformity of polarization in a scattering medium. For the highlighted region (yellow rectangle) in the displayed frame, we show the degree of polarization (DOP)¹ and orientation of linear polarization, for a set of points uniformly distributed in the image plane. It can be clearly seen how both the DOPs and polarization direction are spatially varying. Moreover, the polarization properties are also time-varying.

In this letter, we take this spatio-temporal, non-uniform nature of polarization into account, and propose a novel approach

¹Joint first author

¹Throughout the text we use “DOP” as “linear DOP” for simplicity.
Polarization in scattering media is non-uniform. Left: Intensity image. Right: Polarization for a sparse set of points in the region highlighted in yellow. The longer double arrows indicate the state of linear polarization (intensity and direction, computed using the Q and U components of the per-pixel Stokes vector) ; the ratio of the module of longer arrow to the sum of modules of all arrows indicates the degree of polarization.

Fig. 1. Polarization in scattering media is non-uniform. Left: Intensity image. Right: Polarization for a sparse set of points in the region highlighted in yellow. The longer double arrows indicate the state of linear polarization (intensity and direction, computed using the Q and U components of the per-pixel Stokes vector) ; the ratio of the module of longer arrow to the sum of modules of all arrows indicates the degree of polarization.

for robust, high quality depth estimation in the presence of scattering media, targeting imaging systems capable to directly capture time-resolved light transport. Different from traditional PDI methods, we fix a linear polarizer in front of the sensor and acquire a set of three images, each at different arbitrary orientation. From them, we obtain the polarization properties of each pixel at each instant in time. Last, we use both the images and polarization properties to separate the scattering component from the reflection component, from which we estimate depth of the scene.

We first introduce our adaptive de-scattering algorithm for steady-state imaging, and then extend it to transient state. Assuming that surface interreflections are not dominant and can be neglected, the total incoming radiance \( I(x) \) in the sensor’s pixel at each instant in time. Last, we use both the images and polarization properties to separate the scattering component from the reflection component, from which we estimate depth of the scene.

In the following, we skip the spatial dependency \( x \) for clarity. The direct component \( I_D \) is

\[
I_D = I_C \cdot T(d). \tag{2}
\]

Here \( I_C \) represents the clear image (without the scattering medium) with direct reflection from a surface at distance \( d \), and \( T(d) \) is the transmittance. Assuming a homogeneous medium, \( T(d) \) follows the Beer-Lambert law as \( T(d) = \exp(-\mu_t \cdot d) \), with \( \mu_t \) being the extinction coefficient of such medium. In steady-state, \( T(d) \) can be approximated as [15]

\[
T(d) \approx 1 - \frac{I_S}{I_{S\infty}}, \tag{3}
\]

where \( I_S \) is the scattering intensity of the medium without objects (i.e. assuming \( I_D = 0 \), and therefore \( I = I_S \)), which needs to be acquired as a post-process in the empty medium.

If we assume Lambertian reflection at surfaces, then the degree of polarization of \( I_C \) is \( p_C = 0 \) (i.e. the surfaces are depolarizers). Thus, all observed polarization is due to scattering, and we can approximate the scattering component \( I_S \) as

\[
I_S \approx I \cdot \frac{p}{p_{S\infty}}, \tag{4}
\]

where \( p \) and \( p_{S\infty} \) are the degrees of polarization of \( I \) and \( I_{S\infty} \), respectively.

Unfortunately, we cannot disambiguate the contribution of the direct and the scattering terms in Eq. 1 from Eqs. 2 and 4 alone. Following Treibitz and Schechner [15], we can define the DOP of \( I \) and \( I_S \) respectively as

\[
p = I^{-1}(I_{\text{max}} - I_{\text{min}}) \quad \text{and} \quad p_{S\infty} = I_{S\infty}^{-1}(I_{S\infty}^{\text{max}} - I_{S\infty}^{\text{min}}). \tag{5}
\]

The terms \( (\cdot)_{\text{min}} \) and \( (\cdot)_{\text{max}} \) are acquired from two orthogonal polarizer orientations with minimum and maximum intensity respectively, where \( I = I_{\text{max}} + I_{\text{min}} \) and \( I_{S\infty} = I_{S\infty}^{\text{max}} + I_{S\infty}^{\text{min}} \). Combining Eqs. 1 to 5, we obtain the clear image \( I_C \) as

\[
I_C = I(1 - \frac{p}{p_{S\infty}})(1 - \frac{I - p}{I_{S\infty} \cdot p_{S\infty}})^{-1}. \tag{6}
\]

This formulation implicitly assumes both steady-state light transport, and that the polarization direction is the same for all pixels. Unfortunately, this assumption does not hold in real scenes (see [17] and Fig. 1).

To address this spatial dependence, we rely on the Stokes formulation of polarized light transport. The Stokes vector compactly describes polarized light as \( S = [I \ Q \ U \ V]^T \) [19], where \( I \), \( Q \), \( U \) and \( V \) are the Stokes parameters, and \( V^T \) is the transpose operator for vector \( v \). Specifically, \( I \) is the intensity of light, while \( Q \) and \( U \) represent the linear polarization at two different rotations: \( Q \) is aligned with the axis of the optical system, while \( U \) is rotated 45 degrees with respect to \( Q \). Finally, \( V \) models the amount of circular polarization. Note that the Stokes vector is dependent on the wavelength; for clarity, we consider monochromatic light and remove this dependence. We can now compute the linear DOP of \( S \) as

\[
p = \frac{\sqrt{Q^2 + U^2}}{I}. \tag{7}
\]

We compute the per-pixel DOP \( p \) from a set of measurements taken using a linear polarizer in front of the camera with different arbitrary orientations. For each polarizer orientation \( \alpha \), we capture an image \( I(\alpha) \) defined as [20]

\[
I(\alpha) = \frac{1}{2}(I + Q \cos(2\alpha) + U \sin(2\alpha)). \tag{8}
\]

Taking three different images \( I(\alpha) \) under different orientations \( \alpha \) allows us to build a linear equation system for every pixel. From it, we can calculate \( I \), \( Q \) and \( U \), which are used to compute \( p \) following Eq. 7. Note that the direction of linear polarization can also be obtained since \( I \), \( Q \) and \( U \) are known, although only the DOP is necessary in our work. Fig. 2 shows three measurements \( I(\alpha) \) for the scene shown in Fig. 1, measured through a linear polarizer in three orientations \( \alpha = (\pi/18, 8\pi/18, 3\pi/4) \).

So far we have defined our problem in steady-state. However, the adaptivity is even more challenging for time-resolved imaging, where the direction of linear polarization is also time-varying, so that an exact DOP for each pixel-frame is required.
To extend our pixel-wise adaptive method to transient-state, we need to also take into account the time dependency of the captured data, turning Eq. 8 into

\[ I(t,a) = \frac{1}{2} (I(t) + Q(t) \cos(2\alpha) + U(t) \sin(2\alpha)). \]  

(9)

This allows us to compute the DOP for every pixel-frame \( p(t) \), by capturing three transient images and building a linear system for each pixel-frame. With \( p(t) \) we can now solve the pixel-frame version of Eq. 1, and remove the scattering term \( I_S(x,t) \), which allows us to compute the direct term as

\[ I_D(t) = I(t) (1 - \frac{p(t)}{P_{S\alpha}(t)}). \]  

(10)

Note that, as opposed to Eq. 6, we are not extracting the clear term \( I_C(t) \), but the attenuated one \( I_D(t) = I_C(t) T(d) \). The reason is that Eq. 3 does not hold on transient state, and therefore cannot be applied here. However, \( I_D(t) \) is sufficiently good for recovering high-quality depth, as we will show later.

In theory, Eq. 10 will give us the attenuated surface reflection, from which we can obtain the depth by selecting the first strong peak \( \tau(x) \) in the temporal profile. However, we have found that real measurements might present some noise resulting in numerical problems when solving Eq. 10. To reduce such numerical artifacts, we threshold \( I_D(t) \) as

\[ I_D(t) = \begin{cases}  I(t) \left( 1 - \frac{p(t)}{P_{S\alpha}(t)} \right), & \text{if } \frac{p(t)}{P_{S\alpha}(t)} \geq \epsilon \\ 0, & \text{otherwise} \end{cases} \]  

(11)

where \( \epsilon \) represents the threshold value. In our tests we use \( \epsilon = 0.3 \), which we found to give good results.

Finally, knowing the positions of the camera \( O \) and the light source \( S \), we can compute the depth of a surface point \( X \), by tracking the time of arrival \( \tau(x) \) of the first reflection at \( X \) in pixel \( x \) [21], and solving

\[ \frac{\vec{S} \times \vec{X}}{\cos \theta - 2(\epsilon \tau(x))} = \tau(x), \]  

(12)

where \( \epsilon \) is the speed of light in the medium. By using simple trigonometry, the depth \( d(x) \) at pixel \( x \) can be calculated as

\[ d(x) = \sqrt{\frac{\vec{O} \times \vec{X}}{\cos \theta - 2(\epsilon \tau(x))}}, \]  

(13)

with \( \theta = \vec{S} \times \vec{X} \) the angle between \( \vec{S} \) and \( \vec{X} \).

Given that we target direct time-resolved imaging techniques (e.g. femto-photography [1] or laser-gated techniques [6]), we validate our approach against time profiles matching these imaging devices. For self-containment and benchmarking, we resort to synthetic data. Specifically, we use a publicly available time-resolved renderer [22] with support for polarized light [23], which allows precise control on the scattering parameters as well as ground truth depth. The experimental setup is sketched in Fig. 3. We assume a homogeneous medium, filled with milk diluted in water (medium’s index of refraction \( n = 1.33 \)). Within the medium we place two planes at different depths, as well as a more complex geometry (a bunny), all of them Lambertian (perfect depolrizers) with a high surface albedo \( \Lambda = 0.8 \). We also add a floor and a large wall placed at \( d = 2 \) m. We compute the medium phase functions using Lorenz-Mie theory [24], with a log-normal particles size distribution of mean 2 m\( \mu \) 25. We assume a linearly polarized light source placed outside the medium at \( S = \{-60, 25, 0\} \) cm. The camera is also outside the medium, at \( O = \{-60, 20, 0\} \) cm and with a field of view of \( 30^\circ \), directed towards the center of the medium \( \{40, 20, 0\} \) cm. The output is a set of time-resolved Stokes images, with temporal resolution 0.5 cm \( \times \) 1 s/frame. Based on the Stokes images, we generate three measurements \( I(t,a) \) according to Eq. 9 (\( a = (\pi/18, 8\pi/18, 3\pi/4) \)), while \( I_D(t) \) per pixel-frame is calculated using Eq. 11.

From the above data, we compare with both the non-adaptive transient PDI method [15], and directly recovering the depth map from the original transient image, to demonstrate the effectiveness of the proposed approach. For a fair comparison, we adopt the same layout of the illumination and camera in all experiments, as well as the same depth estimation algorithm after separating the reflectance component in both polarization-aware methods. We evaluate the performance of the three methods for six different medium densities \( \mu_l = \{0.0213, 0.1064, 0.2128, 0.4255, 1.0638, 2.1277\} \) m\(^{-3}\), corresponding to roughly 0.1, 0.5, 1, 2, 5 and 10 times the mean free path (mfp) from the center of the scene. We determined the mean free path for each medium density by computing the mean free path for a base milk concentration, and then scaling the media density accordingly. The scattering albedo is set to 0.9966 in all cases, as given by Lorenz-Mie theory.

Figure 4 shows the depth reconstruction results of our approach compared against the non-adaptive transient PDI and the naive depth reconstruction. Visually, at low media densities (\( \mu_l \leq 1 \) mfp) all methods are able to recover nearby objects within the medium, although our method performs better recovering the details of the surfaces geometry and is significantly more robust to noise. As the density of the medium increases, the performance of both the naive and the non-adaptive method decreases, while our method remains fairly stable. Finally, for very high densities close to diffusion regime (\( \mu_l = 10 \) mfp), our method is still able to reconstruct the objects in the media, while the other two methods fail.

We also conduct a quantitative analysis to assess the reconstruction accuracy of our method across media densities, as shown in Fig. 5. The results shows the same trend shown in Fig. 4: in scenes with strongly non-uniform polarization properties (the ones with high density in our experiment), both previous transient PDI methods and the naive one fail, while our method still obtains depth within reasonable accuracy.

In conclusion, we have introduced a novel transient PDI method, which leverages the assumption of uniform polarization properties. Our method takes into account non-uniform polarization, both in the spatial and temporal domains, allow-
We have shown that our method increases the range of applicability of such methods to more challenging scenarios with optically-thick media, where existing non-adaptive or unpolarized methods failed since they were unable to disambiguate between surface reflection and scattering. Our method targets imaging devices with high temporal resolution for reconstructing depth; this limits its applicability to higher-end imaging systems such as streak cameras or SPADs. An interesting avenue of future work is to generalize our adaptive method to cheaper setups, such as time-of-flight cameras. Since our simulations come from highly accurate physical simulations, with scattering properties based on measured data, we believe they can provide a benchmark for quantitative comparisons for other future algorithms.

Fig. 4. Recovered depth maps for increasing media densities [from left to right $\mu_t = (0.0213, 0.1064, 0.2128, 0.4255, 1.0638, 2.1277)$ m$^{-1}$], with our method (top), the non-adaptive transient PDI method (middle), and the naive unpolarized approach (bottom). The ground truth depth map can be found in Fig. 3, while Fig. 5 shows the mean absolute error and SNR for each image.

Fig. 5. Mean absolute error (left) and Signal-to-Noise Ratio (right) for each method, for increasing media density. Additional details on the error as a function of distance can be found in the supplemental material.

This work is supported by the NSFC (Nos. 61327013, 61327902, 61631009, 61379084, 61402440), DARPA (REVEAL, HR0011-16-C-0025), the Spanish Ministerio de Economía y Competitividad (TIN2016-78753-P), and the European Research Council (ERC Consolidator grant: CHAMELEON project, grant agreement No. 682080).

2Both the dataset and the code will be made public.

REFERENCES

1. INFOMATIONAL FIFTH PAGE

REFERENCES


