Recovery of polarimetric Stokes images by spatial mixture models

Giorgos Sfikas,[1,2,♦] Christian Heinrich,[1] Jihad Zallat,[1] Christophoros Nikou,[2,*]
and Nikos Galatsanos[3]

[1] *University of Strasbourg, Laboratoire des Sciences de l'Image, de l'Informatique et de la*

Télédétection / LSIIT, UMR CNRS-UDS 7005, BP 10413-67412, Illkirch cedex, France.

[2] *Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece.*

[3] *Department of Electrical and Computer Engineering,*

University of Patras, 26500 Rio, Greece.

[*] *Corresponding author: cnikou@cs.uoi.gr*

A Bayesian approach for joint restoration and segmentation of polarization encoded images is presented with emphasis on both physical admissibility and smoothness of the solution. Two distinct models for the sought polarized radiances are used: i) the polarized light at each site of the image is described by its Stokes vector which directly follows a mixture of truncated Gaussians explicitly assigning zero probability to inadmissible configurations, ii) polarization at each site is represented by the coherency matrix which is parameterized by a set of variables assumed to be generated by a spatially varying mixture of Gaussians. Application on real and synthetic images using the proposed methods assess the pertinence of the approach. © 2011 Optical Society of America

OCIS codes: (120.5410) Polarimetry; (110.2960) Image analysis; (100.3190) Inverse problems; (000.3860) Mathematical methods in physics.

1

1. Introduction

Exploiting the polarization of light has been shown to be a useful and powerful technique, overcoming many limitations that arise in classical radiance measurement-based imagery. There is increasing evidence that polarization imaging provides rich information about the local nature of inhomogeneous objects. This imaging modality requires the development of efficient imaging systems that can record spatially distributed polarization patterns across a scene, and the development of appropriate techniques for handling and processing the issued multicomponent images while preserving the physical integrity of the unknowns to be estimated.

In this paper we address the processing of images issued by imaging Stokesmeters. Stokes imaging consists in estimating the four Stokes parameters of each pixel in an image. This is traditionally achieved by placing a Polarization State Analyzer (PSA) in front of a camera. This configuration allows acquiring polarized radiance images $\boldsymbol{g}$ that are used to estimate the multi-component Stokes image $\boldsymbol{s}$. Images $\boldsymbol{g}$ and $\boldsymbol{s}$ are linked by the Polarization Measurement Matrix (PMM) (matrix $\boldsymbol{H}$) that depends on the PSA configuration ($\boldsymbol{g} = \boldsymbol{H}\boldsymbol{s}$, pixelwise). Classically, the Stokes parameters are obtained using a pseudo-inverse approach, which is sensitive to noise that degrades the acquired intensity images. Moreover, at each location, the Stokes parameters must satisfy physical constraints that can be infringed when using pseudo-inversion.

Image processing literature is abundant with restoration estimators and methodologies [1–3]. However, most of them are not readily adaptable to the polarimetric imaging context. Specifically, polarimetric image restoration methodologies have already been proposed, like the method in [4], where the Stokes image is assumed piecewise-constant. One of the objectives of the present work is to relax this hypothesis. Stokes restoration estimators were also proposed in [5]. While the latter model covers spatial coherence and edge-preservation, a number of related parameters have to be fine-tuned empirically. This work uses a combination of general-purpose optimization methods including Nelder-Mead, BFGS and brute force search. All this makes the method time-consuming. Besides, physical admissibility constraints are not accounted for.

Much effort has been devoted to the problem of jointly estimating Stokes images and the blur, which is a blind deconvolution problem (see in particular the work of Valenzuela and Fessler [6] and LeMaster and Cain [7]). The corresponding estimation problem is much more complicated and difficult. Estimation of the point spread function is outside the scope of the current article. Let us also mention that LeMaster and Cain [7] propose a parameterization of the Stokes vector, accounting for the physical admissibility constraints. However, their work does not incorporate prior information, corresponding to considering a regularization term, which is the core of the present work.

In this paper, we present an edge-preserving spatially-smoothing Bayesian model, capable of simultaneously performing restoration and segmentation of Stokes images. Intuitively, we conjecture that the problem of recovering the images should be intertwined with the problem of segmenting the images. In other words, a good restoration should lead to a good segmentation and *vice versa* which is the motivation of performing the two operations jointly. Also, all important model parameters are computed automatically, with a low computational burden. Eventually, Stokes channels are inherently treated as correlated. This makes sense, since the Stokes channels correspond to the same given object.

Polarimetric imaging gives rise to intricate estimation problems because of the associated underlying physical admissibility conditions [8]. Stokes images exhibit the particularity that, while they are comprised of four separate channels, only a subset of $\mathbb{R}^4$ constitutes admissible Stokes 4-variate vectors. To work around this problem, we propose two variations of the proposed approach, resulting in two distinct algorithmic schemes.

In the first approach we consider that a Stokes vector is attached to each pixel in the image. We assume further that the image vectors follow a truncated Gaussian mixture distribution, such that restorations comprised of inadmissible vectors are tagged with zero probability. In the second approach, we choose to work with the coherency matrix to represent the light polarization. From a physical point of view, the two representations are strictly equivalent since there is a one-to-one correspondence between the two entities. However, working with the coherency matrix allows to account easily for the constraints that arise from the physics of light polarization. This variant of the algorithm imposes a spatially varying mixture of

Gaussian distributions on a suitable transformation of the Stokes image. Furthermore, with a properly chosen prior set on the probabilities of the underlying segmentation class labels, we achieve to produce smooth edge-preserving segmentations that in turn produce smooth image restorations. A preliminary short version of the second model is presented in [9].

Both models are inferred using the Expectation-Maximization (EM) algorithm. EM, widely used in a wide range of problems since its introduction in [10], is well-known to provide numerically stable and elegant solutions [11].

The paper is organized as follows. In section 2, we introduce the observation model and general stochastic assumptions. We define the notations and quantities used in the subsequent developments. Sections 3 and 4 introduce the Bayesian model for joint segmentation and restoration in its two different versions, both yielding a valid restoration. We show how to handle the model for both versions using the EM algorithm. In section 5, we present results corresponding to sets of noise-degraded synthetic images and to sets of real data. Conclusion and discussion are given in section 6.

2. General model

The polarization state of light coming from a scene is fully described by either the Stokes vector $\boldsymbol{s} = [s_1 \ s_2 \ s_3 \ s_4]^T$ or the coherency matrix $\boldsymbol{\Phi}$ which is the variance-covariance matrix of the components $E_{x,y}$ of the complex electric field vector of the wave. While $\boldsymbol{\Phi}$ is a complex 2×2 matrix and $\boldsymbol{s}$ is a real 4×1 vector, there is a one-to-one correspondence between $\boldsymbol{\Phi}$ and $\boldsymbol{s}$ as:

$$\boldsymbol{\Phi} = \begin{bmatrix} \langle E_x E_x^* \rangle & \langle E_x E_y^* \rangle \\ \langle E_x^* E_y \rangle & \langle E_y E_y^* \rangle \end{bmatrix} = \begin{bmatrix} s_1 + s_2 & s_3 - i\,s_4 \\ s_3 + i\,s_4 & s_1 - s_2 \end{bmatrix}.$$

In Stokes imaging polarimetry, the main concern is to acquire the distributed polarization pattern of light across the scene. This is accomplished by inserting a complete state analyzer in front of the camera. The imaging model may be described by the following equation:

$$\boldsymbol{g}^n = \boldsymbol{H}\boldsymbol{s}^n + \boldsymbol{\epsilon}^n,$$

where $\boldsymbol{g}^n$ is the vector of the raw measured radiances, $\boldsymbol{H}$ is the PMM, while $\boldsymbol{s}^n$ is the

Stokes vector to be estimated and n indexes the pixel of interest. Variable $\boldsymbol{\epsilon}^n$ represents the image noise. The problem can be stated as follows: given the radiances $\boldsymbol{g}^n$ measured through a polarization sensitive detector characterized by its PMM $\boldsymbol{H}$ and specific measurement noise characteristics inherently related to physical systems, find the best polarization pattern estimate that can describe correctly the raw measured data $\boldsymbol{g}^n$. At least four independent states of the analyzer are required to estimate all elements of the Stokes vector for each pixel location.

We emphasize that polarization state estimates must satisfy:

$$s_1^n \geq 0, \quad (s_1^n)^2 \geq (s_2^n)^2 + (s_3^n)^2 + (s_4^n)^2. \tag{1}$$

Equivalently, the constraints (1) are satisfied if and only if the corresponding coherency matrix $\boldsymbol{\Phi}$ of the estimated Stokes vector is positive semidefinite.

Let $\boldsymbol{g} = \{\boldsymbol{g}^n\}_{n=1}^N$ denote the set of vectors $\boldsymbol{g}^n$. We model the noise $\boldsymbol{\epsilon}^n$ on each Stokes channel as zero mean, additive, white Gaussian. Formally this translates to

$$p(\boldsymbol{g}|\boldsymbol{s};\boldsymbol{V}) = \prod_{n=1}^N p(\boldsymbol{g}^n|\boldsymbol{s}^n;\boldsymbol{V}),$$
$$\boldsymbol{g}^n|\boldsymbol{s}^n;\boldsymbol{V} \sim \mathcal{N}(\boldsymbol{H}\ \boldsymbol{s}^n, \boldsymbol{V}), \tag{2}$$

where the covariance matrix is $\boldsymbol{V} = \sigma^2 \boldsymbol{I}$, and the 4×4 PMM $\boldsymbol{H}$ is supposed to be known *a priori*. Let us note at this point that some extensions are easily applicable to our model, namely extending the noise model to consider colored noise, or considering a spatially variant PMM, the latter case being not rare in practice. These extensions will not be considered here for the sake of simplicity.

We take advantage of our probabilistic generative model formulation and assume a prior distribution on the Stokes vectors. Such prior knowledge accounts for the intuitive fact that vectors with spatially neighboring coordinates are likely to have values close to one another. Therefore, we consider an underlying segmentation of the polarimetric image in K segments. The segmentation is defined by the set of $K \times 1$ vectors $\boldsymbol{z} = \{\boldsymbol{z}^n\}_{n=1}^N$. Each member $\boldsymbol{z}^n$ is defined as a vector with its kth variate set to 1 if the corresponding n^{th} Stokes vector $\boldsymbol{s}^n$ belongs to the kth segment; otherwise, it is set to zero. Correspondingly, every Stokes vector is assumed to belong to exactly one segment.

Recurring to a prior proposed originally in a natural image segmentation context [12], we suppose that label vectors $\boldsymbol{z}$ are multinomially *i.i.d.* distributed. This distribution is parameterized by the *contextual mixing proportions* set $\boldsymbol{\pi}$:

$$\boldsymbol{z}^n | \boldsymbol{\pi}^n \sim Mult(\boldsymbol{\pi}^n).$$

The prior probability vectors $\boldsymbol{\pi} = \{\boldsymbol{\pi}^n\}_{n=1}^N$ are subject to the positiveness $\pi_k^n \geqslant 0, \forall k \in [1, \ldots K], \forall n \in [1, \ldots N]$ and sum-to-unity $\sum_{k=1}^K \pi_k^n = 1, \forall n \in [1, \ldots N]$ constraints.

We assume a Markov random field (MRF) on $\boldsymbol{\pi}$, which equivalently means that $\boldsymbol{\pi}$ is governed by a Gibbs distribution [13], generally expressed as:

$$p(\boldsymbol{\pi}) \propto \prod_C e^{-\psi_c(\boldsymbol{\pi})}, \tag{3}$$

where ψ_c is a function on clique c, called *clique potential* function in the literature, and the product is over all minimal cliques of the MRF.

A clique function assuming that the local differences of contextual mixing proportions follow a Student-t distribution, with its peak set at zero, is a pertinent choice. This choice, proposed in a natural image segmentation context in [12], amounts to an edge-preserving line-process [13]. The probability law for local differences is thus set to the Student's-t distribution:

$$\pi_k^n - \pi_k^j \sim \mathcal{S}t(0, \beta_{kd}^2, \nu_{kd}),$$

$$\forall n \in [1, \ldots N], \forall k \in [1, \ldots K], \forall d \in [1, \ldots D], \forall j \in \gamma_d(n). \tag{4}$$

The parameters β_{kd} control how tightly smoothed we need the Stokes vectors of segment k to be. In (4), D stands for the number of neighborhood adjacency types and $\gamma_d(n)$ is the index set of neighbors of pixel indexed by n, with respect to the dth adjacency type. In our model, we assume 4 neighbors for each pixel (first-order neighborhood), and partition the corresponding adjacency types into horizontal and vertical, thus setting $D = 2$. This variability of parameter aims to capture the intuitive property that smoothness statistics may vary along clusters and spatial directions [14].

One can see that the assumption in (4) is equivalent to

$$\pi_k^n - \pi_k^j \sim \mathcal{N}(0, \beta_{kd}^2 / u_k^{nj}),$$

6

$$u_k^{nj} \sim \mathcal{G}(\nu_{kd}/2, \nu_{kd}/2), \quad \forall n, k, d, \quad \forall j \in \gamma_d(n),$$

where $\mathcal{N}$ and $\mathcal{G}$ represent a Gaussian and a Gamma distribution respectively. This breaking-down of the Student's-t distribution allows clearer insight on how our implicit edge-preserving line-process works. Since u_k^{nj} depends on data indexed by n, each weight difference in the MRF can be described by the realization of a particular Gaussian distribution. On the one hand, as $u_k^{nj} \to +\infty$, the distribution tightens around zero and forces neighboring contextual mixing proportions to be smooth. On the other hand, $u_k^{nj} \to 0$ signifies the existence of an edge and consequently no smoothness.

By imposing smoothness constraints for the segmentation, Stokes vectors spatially close to one another are more likely to belong to the same class and share the same statistics. Consequently, smoothness is implicitly enforced on the resulting restoration as well.

In model terms, restoration and segmentation means estimating respectively $\boldsymbol{s}$ and $\boldsymbol{z}$ given $\boldsymbol{g}$.

3. Spatially varying mixture of truncated Gaussians on the Stokes vectors

In this section, we present a complete generative model, building on the basic structure defined in section 2, and based on the Stokes vectors.

3.A. Model definition

For each image segment k we define a probability distribution (*kernel*) generating the Stokes vectors belonging to the corresponding segment. While in image segmentation problems the chosen distribution is typically Gaussian [15], in the present problem we need a kernel choice that will assign zero probability mass density to vectors not complying with the constraints given in (1). To this end, the Gaussian distribution is not convenient. To work around this particularity, we can instead consider a truncated Gaussian probability kernel, assumed to generate the Stokes vectors belonging to the corresponding segment. This writes

$$p(\boldsymbol{s}|\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{n=1}^{N} p(\boldsymbol{s}^n|\boldsymbol{z}^n; \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with

$$p\left(\boldsymbol{s}^{n}|\boldsymbol{z}^{n};\boldsymbol{\mu},\boldsymbol{\Sigma}\right) \propto \prod_{k=1}^{K}\left[\mathcal{I}\left(\boldsymbol{s}^{n}\right)\ \mathcal{N}\left(\boldsymbol{s}^{n};\boldsymbol{\mu}_{k},\boldsymbol{\Sigma}_{k}\right)\right]^{z_{k}^{n}}. \tag{5}$$

The indicator function $\mathcal{I}(\cdot)$ is equal to unity if the Stokes vector constraints (1) are met, otherwise it is equal to zero. In this manner, restorations comprising any inadmissible Stokes vector are automatically assigned a vanishing density, and are thus impossible.

We call this model mixture of truncated Gaussians (GMM-T) and its generative mechanism over the distribution of $\boldsymbol{s}$ may be examined in detail in the graphical model in figure 1.

3.B.   Model inference

The inference problem formulates as: given observations $\boldsymbol{g}$, estimate the true Stokes image $\boldsymbol{s}$.

We resort to generalized maximum likelihood estimation, where the estimates are obtained by optimization of the generalized log-likelihood

$$\ln p\left(\boldsymbol{g},\boldsymbol{s},\boldsymbol{\pi};\boldsymbol{\Psi}\right) \tag{6}$$

with respect to $\boldsymbol{s}$, $\boldsymbol{\pi}$ and $\boldsymbol{\Psi}=\{\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{V}\}$.

This may be addressed considering the hidden variables $\boldsymbol{z}$ (the labels) and $\boldsymbol{u}$ (the edge map) and using the EM algorithm [10, 16]. While we do not need to determine estimates for the hidden variables $\boldsymbol{z}$ and $\boldsymbol{u}$, they play an important role in the model inference in an indirect way. The choice as to which variables are hidden or not is made according to the tractability of the model.

Writing

$$\ln p\left(\boldsymbol{g},\boldsymbol{s},\boldsymbol{\pi};\boldsymbol{\Psi}\right) = \ln p\left(\boldsymbol{g}|\boldsymbol{s},\boldsymbol{\pi};\boldsymbol{\Psi}\right) + \ln p\left(\boldsymbol{s},\boldsymbol{\pi};\boldsymbol{\Psi}\right),$$

the EM algorithm reads [16]:

E-step: compute the complete likelihood function Q defined by

$$Q\left((\boldsymbol{s},\boldsymbol{\pi},\boldsymbol{\Psi}),(\boldsymbol{s},\boldsymbol{\pi},\boldsymbol{\Psi})^{(t)}\right) = \mathcal{E}_{\boldsymbol{z},\boldsymbol{u}|\boldsymbol{g},(\boldsymbol{s},\boldsymbol{\pi})^{(t)};\boldsymbol{\Psi}^{(t)}}\left[\ln p\left(\boldsymbol{g},\boldsymbol{z},\boldsymbol{u}|\boldsymbol{s},\boldsymbol{\pi};\boldsymbol{\Psi}\right)\right]$$
$$= \int_{\boldsymbol{u}}\sum_{\boldsymbol{z}}\ \ln p\left(\boldsymbol{g},\boldsymbol{z},\boldsymbol{u}|\boldsymbol{s},\boldsymbol{\pi};\boldsymbol{\Psi}\right)\ p\left(\boldsymbol{z},\boldsymbol{u}|\boldsymbol{g},(\boldsymbol{s},\boldsymbol{\pi})^{(t)};(\boldsymbol{\Psi})^{(t)}\right)\ \mathrm{d}\boldsymbol{u}, \tag{7}$$

M-step: update the parameters as

$$(\boldsymbol{s},\boldsymbol{\pi},\boldsymbol{\Psi})^{(t+1)} = \arg\max_{\boldsymbol{s},\boldsymbol{\pi},\boldsymbol{\Psi}}\left[Q\left((\boldsymbol{s},\boldsymbol{\pi},\boldsymbol{\Psi}),(\boldsymbol{s},\boldsymbol{\pi},\boldsymbol{\Psi})^{(t)}\right) + \ln p\left(\boldsymbol{s},\boldsymbol{\pi};\boldsymbol{\Psi}\right)\right].$$

Iterating the E and M steps (iteration is indexed by t) yields a local optimum of the generalized log-likelihood (see eq. 6). The corresponding computations are detailed in Appendices A and B.

The proposed restoration / segmentation procedure is summed up in Algorithm 1.

4. Spatially varying mixture of Gaussians on intrinsic parameters

In this section, we present a complete generative model, building on the basic structure defined in section 2, and based on intrinsic parameters.

4.A. Model definition

To assign zero probability density to vectors not complying with the constraints given in (1), we consider here a special parameterization of the Stokes vectors. Instead of considering a prior structure on the $\boldsymbol{s}$ vectors, we shall work with this intermediate layer in the present approach.

Let $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \lambda_3 \ \lambda_4]^T$ be a parameterization $\boldsymbol{\varphi}$ of the Stokes vector $\boldsymbol{s} = [s_1 \ s_2 \ s_3 \ s_4]^T$, with $\boldsymbol{s} = \boldsymbol{\varphi}(\boldsymbol{\lambda})$. Let matrix $\boldsymbol{\Lambda}$ be

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_3 + i \ \lambda_4 & \lambda_2 \end{bmatrix}.$$

The transformation $\boldsymbol{\varphi}$ is defined to comply with

$$\frac{1}{2}\boldsymbol{\Phi} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H, \tag{8}$$

where $\boldsymbol{\Phi}$ is the coherency matrix defined in section 2. It can be easily seen that any real vector $\boldsymbol{\lambda}$ will yield, according to transformation (8), a Stokes vector $\boldsymbol{s}$ that will necessarily satisfy the constraints (1). We can conveniently assume a Gaussian $i.i.d.$ distribution hypothesis on the $\boldsymbol{\lambda}$ vectors:

$$\boldsymbol{\lambda}^n | z_k^n = 1; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \quad \sim \quad \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{9}$$

We call this model spatially varying Gaussian mixture model on intrinsic parameters (GMM-IP). The generative mechanism exploiting the described transformation for $\boldsymbol{s}$ can be examined in detail in the graphical model in figure 2.

<center>9</center>

Algorithm 1 Proposed restoration / segmentation algorithm (section 3, GMM-T)

1 Initialize $\boldsymbol{\Psi}^{(0)}, \boldsymbol{\pi}^{(0)}, \boldsymbol{s}^{(0)}$. $t \leftarrow 1$. Use the pseudo-inverse (B7, B8) to compute initial estimate for $\boldsymbol{s}$.

2 Compute E-step estimates. Namely

    2.1 Update $\boldsymbol{z}^{(t)}$ expectations using equation (A1).

    2.2 Update $\boldsymbol{u}^{(t)}$ expectations using equation (A2).

    2.3 Update $\ln \boldsymbol{u}^{(t)}$ expectations using equation (A3).

3 Compute M-step estimates. Namely

    3.1 Update parameters $\boldsymbol{\beta}^{(t)}$ using equation (A9).

    3.2 Update parameters $\boldsymbol{V}^{(t)}$ using equation (A10).

    3.3 Update parameters $\boldsymbol{\pi}^{(t)}$ by solving (A11).

    3.4 Update parameters $\boldsymbol{\nu}^{(t)}$ by solving (A12).

    3.5 Update $\boldsymbol{\mu}^{(t)}$, $\boldsymbol{\Sigma}^{(t)}$ parameters using as initial approximations eq. (B4, B5) and correcting them as described in the appendix.

    3.6 Update Stokes estimates $\boldsymbol{s}^{(t)}$. For each $n \in [1, N]$ do

        3.6.1 $\kappa \leftarrow 0$. Set $\boldsymbol{\phi}^{(\kappa)}$ to pseudoinverse of $\boldsymbol{g}^n$ (eq. B7) and correct it if necessary using eq. (B8).

        3.6.2 Let $\boldsymbol{\rho}^{(\kappa)}(h) = (1-h)\boldsymbol{\phi}^{(\kappa)} + h[\boldsymbol{A}^{n(t)}]^{-1}\boldsymbol{b}^{n(t)}$.

        3.6.3 Find $h^\star \in [0,1]$ such that $h^\star$ is as large as possible and such that $\boldsymbol{\rho}^{(\kappa)}(h)$ satisfies the Stokes constraints (1).

        3.6.4 Let $\boldsymbol{\phi}^{(\kappa+1)} \leftarrow \boldsymbol{\rho}^{(\kappa)}(h^\star)$. $\kappa \leftarrow \kappa + 1$.

        3.6.5 Line search for $\boldsymbol{\phi}^{(\kappa+1)}$ on a random direction. $\kappa \leftarrow \kappa + 1$.

        3.6.6 If $\boldsymbol{\phi}^{(\kappa)}$ has converged, let $\boldsymbol{s}^{n(t)} \leftarrow \boldsymbol{\phi}^{(\kappa)}$ and move to next index n of step 3.6; else go back to step 3.6.1.

4 $t \leftarrow t + 1$. Back to step 2 unless model generalized likelihood (eq. (6)) has converged.

5 End. Vector set $\boldsymbol{s}^{(t)}$ constitutes the desired Stokes vector reconstruction; vector set $\boldsymbol{z}^{(t)}$ constitutes the desired segmentation.

4.B.   Model inference

The problem of inference formulates in our case as: given observations $\boldsymbol{g}$, estimate the intrinsic parameters $\boldsymbol{\lambda}$. The Stokes parameters $\boldsymbol{s}$ are then deduced from $\boldsymbol{\lambda}$ using the parameterization $\boldsymbol{\varphi}$.

We resort to generalized maximum likelihood estimation, where the estimates are obtained by optimization of the generalized log-likelihood

$$\ln p\left(\boldsymbol{g}, \boldsymbol{\lambda}, \boldsymbol{\pi}; \boldsymbol{\Psi}\right) \tag{10}$$

with respect to $\boldsymbol{\lambda}$, $\boldsymbol{\pi}$ and $\boldsymbol{\Psi} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{V}\}$.

The general EM approach considered to handle this model is similar to the one considered for the preceding model. The corresponding computations are detailed in Appendices A and C.

The proposed restoration / segmentation procedure is summed up in Algorithm 2.

5.   Simulation results

We applied the proposed recovery algorithm to three Stokes test images, one artificial image and two real images. We pre-processed all observed images using the scatter plot method [17, 18] which proved beneficial, especially in the case of low SNRs. We used a small 3×3 pixels sliding window and robust estimators to compute local statistics (mean and standard deviation). We obtain at the end of the noise-removal process an estimate of the noise-free image corresponding to the acquired image.

5.A.   Application to an artificial test image

To illustrate the relevance and efficiency of the proposed approach, we synthesized a 64×64 pixels Stokes image composed of a 32×32 pixels square region with a smoothly varying completely polarized Stokes signature surrounded by a 32 pixels border with a constant signature corresponding to the vector $\left(1, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)^{T}$.

The results obtained are shown in figure 3. The simulations were carried out with a noise variance value that corresponds to signal-to-noise ratios over the mean radiance image (first

Algorithm 2 Proposed restoration / segmentation algorithm (section 4, GMM-IP)

---

1 Initialize $\boldsymbol{\Psi}^{(0)}$, $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, $\boldsymbol{s}^{(0)}$. $t \leftarrow 1$. Use the pseudoinverse (B7, B8) and (8) for $\boldsymbol{\lambda}^{(0)}$.

2 Compute E-step estimates. Namely

   2.1 Update $\boldsymbol{z}^{(t)}$ expectations using equation (A1).

   2.2 Update $\boldsymbol{u}^{(t)}$ expectations using equation (A2).

   2.3 Update $\ln \boldsymbol{u}^{(t)}$ expectations using equation (A3).

3 Compute M-step estimates. Namely

   3.1 Update $\boldsymbol{\mu}^{(t)}$, $\boldsymbol{\Sigma}^{(t)}$ using equations (A7, A8).

   3.2 Update parameters $\boldsymbol{\beta}^{(t)}$ using equation (A9).

   3.3 Update parameters $\boldsymbol{V}^{(t)}$ using equation (A10).

   3.4 Update parameters $\boldsymbol{\pi}^{(t)}$ by solving (A11).

   3.5 Update parameters $\boldsymbol{\nu}^{(t)}$ by solving (A12).

   3.6 Update parameters $\boldsymbol{\lambda}^{(t)}$. In particular, for each $n \in [1, N]$, do

      3.6.1 $\kappa \leftarrow 1$.

      3.6.2 Optimize (C1) for variate κ of $\boldsymbol{\lambda}^{n(t)}$.

      3.6.3 $\kappa \leftarrow \kappa + 1$. If $\kappa > 4$ and if $\boldsymbol{\lambda}^{n(t)}$ has converged, continue to step 3.7; else go back to step 3.6.1.

   3.7 Update Stokes vectors estimates $\boldsymbol{s}^{(t)}$ using (C2).

4 $t \leftarrow t + 1$. Back to step 2 unless model generalized log-likelihood (eq. (10)) has converged.

5 End. Vector set $\boldsymbol{s}^{(t)}$ constitutes the desired Stokes vector reconstruction; vector set $\boldsymbol{z}^{(t)}$ constitutes the desired segmentation.

Stokes channel) of about 10 dB (this corresponds to global SNR of about 16 dB). The values presented in table 1 correspond to the improvement in signal to noise ratio (ISNR) defined as

$$ISNR = 20 \log_{10} \frac{\| \boldsymbol{s}^\star - \boldsymbol{g} \|}{\| \boldsymbol{s}^\star - \hat{\boldsymbol{s}} \|}, \tag{11}$$

where $\boldsymbol{s}^\star$, $\hat{\boldsymbol{s}}$ and $\boldsymbol{g}$ represent the ground truth, the estimate, and the observed image respectively. Higher values for the ISNR correspond to better restorations. The pseudo-inverse estimates are computed using (B7).

The results clearly demonstrate that the proposed methods give better restorations than the general purpose pseudo-inverse. The pseudo-inverse approach yields a bad quality solution, which infringes the physical admissibility constraints (1) for a considerable number of pixels. Both of our methods ensure physical admissibility for each recovered Stokes vector.

5.B. *Application to real images*

The two real images were selected to evaluate the ability of our methods to handle piecewise constant signatures as well as varying signatures under different illumination conditions.

The first of the real images (fig. 4) consists of an assembly of a number of objects with – as a first approximation – homogeneous polarimetric responses (background, metallic, dichroic and birefringent patches). These data were provided by Prof. A. de Martino, LPICM, Ecole Polytechnique, France. The scene was imaged by a complete imaging Stokesmeter under a very good polarized illumination (high SNR). We found that our approach performs quite well and provides a solution that is consistent with the physics of the problem (constant responses).

The second of the real images (figs. 5 and 6) corresponds to a metallic car toy partially occluded by a circular dichroic patch. Imaging is performed by an inline rotating quarter-wave plate Stokes polarimeter under natural illumination through a 10 nm bandwidth interferential filter (low SNR). For the sake of conciseness, only the degree of polarization (DOP) images computed from the pseudo-inverse (PI) solution and from the GMM-T solution are presented. The PI DOP map is too rough and has 12% of non physically admissible pixels, whereas the DOP maps provided by our approaches are smooth and physically admissible.

13

Estimation has been done using $K = 5$.

In all cases, synthetic data and real data, the GMM-T model yields slightly smoother results than the GMM-IP model. This could have been expected: in the GMM-IP model, the Stokes channels are computed via the sum of intermediate variables (the $\boldsymbol{\lambda}$'s), whose oscillations are amplified, by summation, and transmitted to the Stokes maps. Moreover, the smoother results (GMM-T model) are the more satisfying ones for the simulations considered here.

For spatially varying gray level signatures, the choice of K plays no crucial role if one is not primarily interested in the segmentation since several values of K lead to the same Stokes-vector estimates. For example, in the metallic toy car case, Stokes estimates corresponding to $K = 4, 5, 6$ are nearly the same. However, the choice of K is more crucial in the case of piecewise constant images. Processing of such data and the choice of K are addressed in [19]. In the general case, the choice of K is still a largely open topic.

In table 2 we indicate the runtime necessary for each iteration, for each of the proposed algorithms. We found that our algorithm converged in each case in about 10 EM iterations. Note also that while GMM-T turns out to be much more computationally expensive than GMM-IP, a large percentage of the former is spent on performing the Nelder-Mead optimization step for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ (see Appendix B). In practice however, we found that keeping the initial values (B4), (B5) results in a 3 to 5-fold improvement in runtime, with negligible loss in quality of results.

All computations were done using Matlab, running on a 1.8 GHz PC workstation with 2 Gb of RAM.

6. Conclusion

The modeling of polarimetric Stokes images using two different spatially varying mixtures of Gaussians has been proposed. In each case, the estimation procedure is unsupervised and relies on the EM algorithm, which is well-known to provide an elegant and numerically stable solution [11]. Both proposed models assume an underlying segmentation for the input image.

This estimated segmentation is provided as a byproduct of the procedures.

In the first proposed approach, the Stokes vector was modeled as a spatially varying mixture of truncated Gaussians. The truncation accounted for the polarimetric constraints. In the second approach, each Stokes vector was parametrized by its coherency matrix, and the random variables corresponding to the coherency matrices were generated by spatially varying Gaussian mixtures. The parameterization implicitly accounted the polarimetric constraints. The first model gave slightly better results than the second one, with no conclusive advantage for one method over the other.

We emphasize that our solutions are obtained at a computational cost lower than the ones of concurrent approaches. Moreover, those approaches do not handle physical admissibility conditions. Our computation time should easily be reduced, for example by choosing another implementation language.

Matlab code

The Matlab code is available upon request for academic users from the corresponding author.

Appendix A: Derivation of the EM updates for common variables and parameters of the GMM-T and GMM-IP models

In this appendix, we detail the derivation of the E-step and M-step concerning the variables and parameters common to both models.

E-step

The E-step consists in computing the expected log-likelihood, the expectation being taken over the hidden variables $\boldsymbol{z}$ and $\boldsymbol{u}$ conditioned on the current iteration parameters $\boldsymbol{\Psi}^{(t)}$, $\boldsymbol{\lambda}^{(t)}$ (or $\boldsymbol{s}^{(t)}$), $\boldsymbol{\pi}^{(t)}$ and observation $\boldsymbol{g}$, where t indexes the current iteration. According to the graphical models of figures 1 and 2, the hidden variables $\boldsymbol{z}$ and $\boldsymbol{u}$ are conditionally independent given the parameters and the observed variables. Therefore $\mathcal{E}_{\boldsymbol{z},\boldsymbol{u}|\boldsymbol{g},\boldsymbol{\lambda},\boldsymbol{\pi};\boldsymbol{\Psi}}(\cdot) = \mathcal{E}_{\boldsymbol{z}|\boldsymbol{g},\boldsymbol{\lambda},\boldsymbol{\pi};\boldsymbol{\Psi}}\{\mathcal{E}_{\boldsymbol{u}|\boldsymbol{g},\boldsymbol{\lambda},\boldsymbol{\pi};\boldsymbol{\Psi}}(\cdot)\}$ for GMM-IP, meaning that we can compute these expectations separately. An analogous independence relation holds for GMM-T, where $\boldsymbol{\lambda}$ is replaced by $\boldsymbol{s}$

15

in the previous formula. Substituting the posteriors of the hidden variables in the complete likelihood function (7) results in an expression where functions of the expected values of the hidden variables need to be computed. In the current model, this results in $Q(\cdot)$ being a function of $<\boldsymbol{z}>$, $<\boldsymbol{u}>$ and $<\ln \boldsymbol{u}>$.

The induced E-step updates for GMM-IP read, for all n, k, d and $j \in \gamma_d(n)$:

$$<z_k^n>^{(t)} = \frac{\pi_k^{n(t)} \, \mathcal{N}\left(\boldsymbol{\lambda}^{n(t)}; \boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Sigma}_k^{(t)}\right)}{\sum_{l=1}^{K} \pi_l^{n(t)} \, \mathcal{N}\left(\boldsymbol{\lambda}^{n(t)}; \boldsymbol{\mu}_l^{(t)}, \boldsymbol{\Sigma}_l^{(t)}\right)}, \tag{A1}$$

$$<u_k^{nj}>^{(t)} = \zeta_k^{nj(t)} / \eta_k^{nj(t)}, \tag{A2}$$

$$<\ln u_k^{nj}>^{(t)} = \mathcal{F}\left(\zeta_k^{nj(t)}\right) - \ln \eta_k^{nj(t)}, \tag{A3}$$

where $\mathcal{F}(\cdot)$ stands for the digamma function, and parameters ζ, η are defined by:

$$\forall j \in \gamma_d(n) \qquad \zeta_k^{nj(t)} = \frac{1}{2}\left(\nu_{kd}^{(t)} + 1\right),$$

$$\forall j \in \gamma_d(n) \qquad \eta_k^{nj(t)} = \frac{1}{2}\left(\nu_{kd}^{(t)} + \frac{\left(\pi_k^{n(t)} - \pi_k^{j(t)}\right)^2}{\beta_{kd}^{2(t)}}\right). \tag{A4}$$

For GMM-T, eq. A1 becomes:

$$<z_k^n>^{(t)} = \frac{\pi_k^{n(t)} \left(\Xi\left(\boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Sigma}_k^{(t)}\right)\right)^{-1} \mathcal{N}\left(\boldsymbol{s}^{n(t)}; \boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Sigma}_k^{(t)}\right)}{\sum_{l=1}^{K} \pi_l^{n(t)} \left(\Xi\left(\boldsymbol{\mu}_l^{(t)}, \boldsymbol{\Sigma}_l^{(t)}\right)\right)^{-1} \mathcal{N}\left(\boldsymbol{s}^{n(t)}; \boldsymbol{\mu}_l^{(t)}, \boldsymbol{\Sigma}_l^{(t)}\right)}, \tag{A5}$$

$\Xi(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ being defined in (B2).

M-step

Maximization of the current MAP cost function (that is, the complete likelihood function augmented with prior information) must be driven with respect to the model parameters $\boldsymbol{\Psi}$, $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ (or $\boldsymbol{s}$). The MAP cost function may be written, after straightforward computation:

$$Q\left((\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\Psi}), (\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\Psi})^{(t)}\right) + \ln p\left(\boldsymbol{\lambda}, \boldsymbol{\pi}; \boldsymbol{\Psi}\right) =$$

$$= \mathcal{E}_{\boldsymbol{z}, \boldsymbol{u} | \boldsymbol{g}, (\boldsymbol{\lambda}, \boldsymbol{\pi})^{(t)}; \boldsymbol{\Psi}^{(t)}} \left\{ \ln p\left(\boldsymbol{g}, \boldsymbol{z}, \boldsymbol{u} | \boldsymbol{\lambda}, \boldsymbol{\pi}; \boldsymbol{\Psi}\right) + \ln p\left(\boldsymbol{\lambda}, \boldsymbol{\pi}; \boldsymbol{\Psi}\right) \right\}$$

$$= \mathcal{E}_{\boldsymbol{z}, \boldsymbol{u} | \boldsymbol{g}, (\boldsymbol{\lambda}, \boldsymbol{\pi})^{(t)}; \boldsymbol{\Psi}^{(t)}} \left\{ \ln p\left(\boldsymbol{g}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\lambda}, \boldsymbol{\pi}; \boldsymbol{\Psi}\right) \right\} \quad \text{(A6)}$$

$$= \mathcal{E}_{\boldsymbol{z} | \boldsymbol{\lambda}^{(t)}, \boldsymbol{\pi}^{(t)}} \left\{ \ln p(\boldsymbol{g} | \boldsymbol{\lambda}; \boldsymbol{V}) \right\} + \mathcal{E}_{\boldsymbol{z} | \boldsymbol{\lambda}^{(t)}, \boldsymbol{\pi}^{(t)}} \left\{ \ln p(\boldsymbol{\lambda} | \boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \right\} + \mathcal{E}_{\boldsymbol{z} | \boldsymbol{\lambda}^{(t)}, \boldsymbol{\pi}^{(t)}} \left\{ \ln p(\boldsymbol{z} | \boldsymbol{\pi}) \right\}$$

$$+ \mathcal{E}_{\boldsymbol{u} | \boldsymbol{\pi}^{(t)}} \left\{ \ln p(\boldsymbol{\pi} | \boldsymbol{u}; \boldsymbol{\beta}) \right\} + \mathcal{E}_{\boldsymbol{u} | \boldsymbol{\pi}^{(t)}} \left\{ \ln p(\boldsymbol{u}; \boldsymbol{\nu}) \right\}$$

Optimizing with respect to each of the parameters in turn yields the update equations for all $n, k, d, j \in \gamma_d(n)$, which make up the M-step:

$$\boldsymbol{\mu}_k^{(t+1)} = \frac{\sum_{n=1}^N <z_k^n>^{(t)} \boldsymbol{\lambda}^{n(t)}}{\sum_{n=1}^N <z_k^n>^{(t)}} \quad \text{(A7)}$$

$$\boldsymbol{\Sigma}_k^{(t+1)} = \frac{\sum_{n=1}^N <z_k^n>^{(t)} (\boldsymbol{\lambda}^{n(t)} - \boldsymbol{\mu}_k^{(t+1)})(\boldsymbol{\lambda}^{n(t)} - \boldsymbol{\mu}_k^{(t+1)})^T}{\sum_{n=1}^N <z_k^n>^{(t)}} \quad \text{(A8)}$$

$$\beta_{kd}^{2(t+1)} = \frac{\sum_{n=1}^N \sum_{j \in \gamma_d(n)} <u_k^{nj}>^{(t)} (\pi_k^{n(t)} - \pi_k^{j(t)})^2}{\sum_{n=1}^N |\gamma_d(n)|} \quad \text{(A9)}$$

$$\boldsymbol{V}^{(t+1)} = (4N)^{-1} \sum_{n=1}^N \left(\boldsymbol{g}^n - \boldsymbol{H}\boldsymbol{s}^{n(t)}\right)^T \left(\boldsymbol{g}^n - \boldsymbol{H}\boldsymbol{s}^{n(t)}\right) \boldsymbol{I}. \quad \text{(A10)}$$

For GMM-T, $\boldsymbol{\lambda}$ should be replaced by $\boldsymbol{s}$ in eqs. (A7) and (A8).

Moreover, the *contextual mixing proportions* π_j^n are computed as the roots of a quadratic equation:

$$a_k^n \left(\pi_k^{n(t+1)}\right)^2 + b_k^n \left(\pi_k^{n(t+1)}\right) + c_k^{n(t+1)} = 0 \quad \text{(A11)}$$

with coefficients:

$$a_k^n = -\sum_{d=1}^D \left\{ \beta_{kd}^{-2(t)} \sum_{j \in \gamma_d(n)} <u_k^{nj}>^{(t)} \right\},$$

$$b_k^n = \sum_{d=1}^D \left\{ \beta_{kd}^{-2(t)} \sum_{j \in \gamma_d(n)} <u_k^{nj}>^{(t)} \pi_k^{j(t)} \right\},$$

$$c_k^n = \frac{1}{2} <z_k^n>^{(t)}.$$

The form of the coefficients guarantees that there is always a non negative solution [20]. However, the solutions of eq. (A11) for a given pixel indexed by n will in general not satisfy

the constraints $\sum_{k=1}^{K} \pi_k = 1$, $\pi_k \geqslant 0, \forall k \in [1..K]$. Hence we have to perform a projection onto the constraints space.

Setting the derivative of (A6) with respect to $\nu_{kd}^{(t+1)}$ equal to zero, we obtain $\nu_{kd}^{(t+1)}$ as the solution of the equation:

$$\ln\left(\nu_{kd}^{(t+1)}/2\right) - F\left(\nu_{kd}^{(t+1)}/2\right) +$$

$$+\left[\frac{\sum_{n=1}^{N}\sum_{j\in\gamma_d(n)}(<\ln u_k^{nj}>^{(t)} - <u_k^{nj}>^{(t)})}{\sum_{n=1}^{N}|\gamma_d(n)|}\right] + 1 = 0 \tag{A12}$$

with $F(\cdot)$ being again the digamma function, and $|\mathcal{C}|$ the cardinal of set $\mathcal{C}$.

Appendix B: Derivation of the M-step updates for the GMM-T model

In this appendix, we detail the derivation of the M-step concerning the parameters of the GMM-T model.

Updates of kernel means and kernel covariance matrices

Optimization with respect to kernel means $\boldsymbol{\mu}$ and kernel covariance matrices $\boldsymbol{\Sigma}$ requires special treatment. It involves optimizing $\mathcal{E}_{\boldsymbol{z}|\boldsymbol{s}^{(t)},\boldsymbol{\pi}^{(t)}}\{\ln p(\boldsymbol{s}|\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma})\}$, which cannot be done in closed form. This is due to the fact that $p(\boldsymbol{s}|\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma})$, defined as a truncated Gaussian, contains a normalization constant not readily tractable. Let us rewrite (5) in a more convenient manner as

$$p(\boldsymbol{s}^n|\boldsymbol{z}^n;\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{k=1}^{K}\left[\,\Xi(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)^{-1}\,\,\mathcal{I}(\boldsymbol{s}^n)\,\,\mathcal{N}(\boldsymbol{s}^n;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\right]^{z_k^n}, \tag{B1}$$

where

$$\Xi(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) = \int_{\mathbb{R}^4}\mathcal{I}(\boldsymbol{\sigma})\,\,\mathcal{N}(\boldsymbol{\sigma};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\,\,d\boldsymbol{\sigma}, \tag{B2}$$

so that the distribution normalizes correctly to unity. Noticing that this latter expression cannot be computed analytically but can be written as the expectation $\mathcal{E}_{\boldsymbol{\sigma};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k}\{\mathcal{I}(\boldsymbol{\sigma})\}$, we resort to Monte Carlo approximation. Thus $\Xi(\cdot)$ approximates as

$$\Xi(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) \simeq \frac{1}{T}\sum_{\tau=1}^{T}\mathcal{I}(\boldsymbol{\sigma}^{(\tau)}), \tag{B3}$$

where $\{\boldsymbol{\sigma}^{(\tau)}\}_{\tau=1}^T$ are samples from a Gaussian of mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$. From a computational point of view, we can use a fixed set of samples from a standard Gaussian distribution $\mathcal{N}(0, I)$ and transform them using standard matrix operations to have them distributed as desired [16]. Thence, approximation (B3) can be readily seen as the percentage of samples satisfying the Stokes constraints (1). As $\mathcal{E}_{\boldsymbol{z}|\boldsymbol{s}^{(t)},\boldsymbol{\pi}^{(t)}}\{\ln p\,(\boldsymbol{s}|\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma})\}$ is not continuous everywhere, to optimize it with regard to $\{\boldsymbol{\mu}_k\}_{k=1}^K$ and $\{\boldsymbol{\Sigma}_k\}_{k=1}^K$ we choose the Nelder-Mead method [21].

As initial values $\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{\Sigma}}$ of means and covariances for Nelder-Mead we use

$$\tilde{\boldsymbol{\mu}}_k^{(t+1)} = \frac{\sum_{n=1}^N <z_k^n>^{(t)} \boldsymbol{s}^n}{\sum_{n=1}^N <z_k^n>^{(t)}}, \tag{B4}$$

$$\tilde{\boldsymbol{\Sigma}}_k^{(t+1)} = \frac{\sum_{n=1}^N <z_k^n>^{(t)} (\boldsymbol{s}^n - \boldsymbol{\mu}_k^{(t+1)})(\boldsymbol{s}^n - \boldsymbol{\mu}_k^{(t+1)})^T}{\sum_{n=1}^N <z_k^n>^{(t)}}, \tag{B5}$$

which would in fact be the correct optimizers if we used standard Gaussian distributions instead of truncated ones in our model for the Stokes vectors $\boldsymbol{s}$.

Updates of Stokes estimate

Optimization with respect to $\boldsymbol{s}$ involves the terms $\mathcal{E}_{\boldsymbol{z}|\boldsymbol{s}^{(t)},\boldsymbol{\pi}^{(t)}}\{\ln p\,(\boldsymbol{g}|\boldsymbol{s};\boldsymbol{V})\} + \mathcal{E}_{\boldsymbol{z}|\boldsymbol{s}^{(t)},\boldsymbol{\pi}^{(t)}}\{\ln p\,(\boldsymbol{s}|\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma})\}$ of the objective function (eq. A6). After some manipulation, it appears that we have to maximize for $\boldsymbol{s}^{n(t+1)}$ the quadratic equations $\forall n \in [1, N]$:

$$\boldsymbol{s}^{n(t+1)^T}\boldsymbol{A}^{n(t+1)}\boldsymbol{s}^{n(t+1)} - 2\boldsymbol{b}^{n(t+1)^T}\boldsymbol{s}^{n(t+1)}, \tag{B6}$$

under the constraints (eq. (1)). Matrices $\boldsymbol{A}^{n(t+1)}$ and vectors $\boldsymbol{b}^{n(t+1)}$ are given by:

$$\boldsymbol{A}^{n(t+1)} = \boldsymbol{H}^T\boldsymbol{V}^{n(t+1)^{-1}}\boldsymbol{H} + \sum_{k=1}^K <z_k^n>^{(t)} \boldsymbol{\Sigma}_k^{(t)^{-1}},$$

$$\boldsymbol{b}^{n(t+1)} = \boldsymbol{g}^{nT}\boldsymbol{V}^{n(t+1)^{-1}}\boldsymbol{H} + \sum_{k=1}^K <z_k^n>^{(t)} \boldsymbol{\mu}_k^{(t)^T}\boldsymbol{\Sigma}_k^{(t)^{-1}}.$$

In order to deal with the optimization of the expression (B6), which is constrained, we employ a line search method. As initial estimates $\boldsymbol{s}^{(0)}$ of the restored Stokes vectors, we use the pseudo-inverse of $\boldsymbol{g}$ given by equation

$$\boldsymbol{s}^{n(0)} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{g}^n. \tag{B7}$$

19

If $\boldsymbol{s}^n$ does not satisfy the constraints in (1) its first component is set to

$$(s_1^n)^2 = (s_2^n)^2 + (s_3^n)^2 + (s_4^n)^2 \tag{B8}$$

To find proper directions of search, we take advantage of the fact that (B6) is quadratic and concave, and has an unconstrained optimum in $[\boldsymbol{A}^{n(t+1)}]^{-1}\boldsymbol{b}^{n(t+1)}$. Each step thus comprises a line search towards the direction of this latter point, alternating with a line search on a random direction so as to avoid getting caught on the border of the domain. All line search steps result in points that are feasible under the Stokes constraints (1).

Appendix C: Derivation for the M-step updates for the GMM-IP model

In this appendix, we detail the derivation of the M-step concerning the parameters of the GMM-IP model.

Here we concentrate on the M-step which differs from standard EM approaches due to the intrinsic parameters of the model.

Optimization with respect to the constraint-free parameters $\boldsymbol{\lambda}$ involves the following expression, after dropping constant terms from (A6):

$$(\boldsymbol{g}^n - \boldsymbol{H}\boldsymbol{\varphi}(\boldsymbol{\lambda}^{n(t)}))^T\boldsymbol{V}^{-1}(\boldsymbol{g}^n - \boldsymbol{H}\boldsymbol{\varphi}(\boldsymbol{\lambda}^{n(t)}))+$$

$$\sum_{k=1}^{K}(\boldsymbol{\lambda}^{n(t)} - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\lambda}^{n(t)} - \boldsymbol{\mu}_k) <z_k^n>,$$

which after some manipulation boils down to:

$$\boldsymbol{h}^T\boldsymbol{\Omega}^1\boldsymbol{h} + \boldsymbol{\lambda}^T\boldsymbol{\Omega}^2\boldsymbol{\lambda} + \boldsymbol{\omega}^3\boldsymbol{h} + \boldsymbol{\omega}^4\boldsymbol{\lambda}, \tag{C1}$$

where we have omitted the data and iteration indices n and t for brevity. Parameters $\boldsymbol{h}$, $\boldsymbol{\Omega}^1$, $\boldsymbol{\Omega}^2$, $\boldsymbol{\omega}^3$, $\boldsymbol{\omega}^4$ are given by

$$\boldsymbol{h} \equiv \boldsymbol{H}\boldsymbol{\varphi}(\boldsymbol{\lambda}), \boldsymbol{\Omega}^1 \equiv \boldsymbol{V}^{-1}, \boldsymbol{\Omega}^2 \equiv \sum_{k=1}^{K} <z_k> \boldsymbol{\Sigma}_k^{-1},$$

$$\boldsymbol{\omega}^3 \equiv -2\boldsymbol{g}^T\boldsymbol{V}^{-1}, \boldsymbol{\omega}^4 \equiv -2\sum_{k=1}^{K} <z_k> \boldsymbol{\mu}_k^T\boldsymbol{\Sigma}_k^{-1}.$$

20

In view of (8), eq. (C1) is a fourth-order polynomial over each of the variates of $\boldsymbol{\lambda}$. Setting the derivative of (C1) with respect to each of the $\boldsymbol{\lambda}$ variates to zero, we can obtain optimizers for $\boldsymbol{\lambda}$ by solving the resulting third-order polynomial equations. Thus for each $n \in [1, \ldots N]$, we solve iteratively four third-order polynomial equations, and repeat the operation until convergence of $\boldsymbol{\lambda}$.

Finally, in order compute the Stokes estimates $\boldsymbol{s}^n$ we simply make use of the $\boldsymbol{\varphi}$ transformation definition (C1) to obtain the update

$$\begin{cases} s_1^n = [(\lambda_1^n)^2 + (\lambda_2^n)^2 + (\lambda_3^n)^2 + (\lambda_4^n)^2], \\ s_2^n = [(\lambda_1^n)^2 - (\lambda_2^n)^2 - (\lambda_3^n)^2 - (\lambda_4^n)^2], \\ s_3^n = 2\lambda_1^n \, \lambda_3^n, \\ s_4^n = 2\lambda_1^n \, \lambda_4^n. \end{cases} \tag{C2}$$

References

1. R. Molina, J. Mateos, A. Katsaggelos, and M. Vega, "Bayesian multichannel image restoration using compound Gauss-Markov random fields," *IEEE Transactions on Image Processing*, vol. 12, pp. 1642–1654, 2003.

2. J. Chantas, N. Galatsanos, and A. Likas, "Bayesian restoration using a new hierarchical directional continuous edge image prior," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2987–2997, 2006.

3. L. Bar, A. Brook, N. Sochen, and N. Kiryati, "Deblurring of color images corrupted by impulsive noise," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1101–1111, 2004.

4. J. Zallat and C. Heinrich, "Polarimetric data reduction: a Bayesian approach," *Optics Express*, vol. 15, no. 1, pp. 83–96, 2007.

5. J. R. Valenzuela and J. A. Fessler, "Joint reconstruction of Stokes images from polarimetric measurements," *Journal of the Optical Society of America A*, vol. 26, no. 4, pp. 962–968, 2009.

6. J. R. Valenzuela, J. A. Fessler, and R. G. Paxman, "Joint estimation of Stokes images and aberrations from phase-diverse polarimetric measurements," *Journal of the Optical*

Society of America A, vol. 27, no. 5, pp. 1185–1193, 2010.

7. D. LeMaster and S. Cain, "Multichannel blind deconvolution of polarimetric imagery," *Journal of the Optical Society of America A*, vol. 25, no. 9, pp. 2170–2176, 2008.

8. J. Zallat, C. Collet, and Y. Takakura, "Clustering of polarization-encoded images," *Applied Optics*, vol. 43, no. 2, pp. 283–292, 2004.

9. G. Sfikas, C. Heinrich, J. Zallat, C. Nikou, and N. Galatsanos, "Joint recovery and segmentation of polarimetric images using a compound mrf and mixture modeling," in *Proceedings of the IEEE International Conference on Image Processing (ICIP09)*, 7-11 November 2009, Cairo, Egypt, 2009, pp. 3901–3904.

10. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, no. 1, pp. 1–38, 1977.

11. K. Lange, *Optimization.* Springer, 2004.

12. G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich, "Spatially varying mixtures incorporating line processes for image segmentation," *Journal of Mathematical Imaging and Vision*, vol. 36, no. 2, pp. 91–110, 2010.

13. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 721–741, 1984.

14. C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1121–1130, 2007.

15. G. McLachlan, *Finite mixture models.* Wiley-Interscience, 2000.

16. C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006.

17. B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi, "Estimating noise and information of multispectral imagery," *Optical Engineering*, vol. 41, no. 3, pp. 656–668, 2002.

18. J. Zallat, S. Aïnouz, and M. Stoll, "Optimal configurations for imaging polarimeters: impact of image noise and systematic errors," *Journal of Optics A – Pure and Applied*

Optics, vol. 8, no. 9, pp. 807–814, 2006.

19. J. Zallat, C. Heinrich, and M. Petremand, "A Bayesian approach for polarimetric data reduction : the Mueller imaging case," *Optics Express*, vol. 16, no. 10, pp. 7119–7133, 2008.

20. G. Birkhoff and S. MacLane, *A survey of modern algebra.* New York: McMillan, 1953.

21. J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
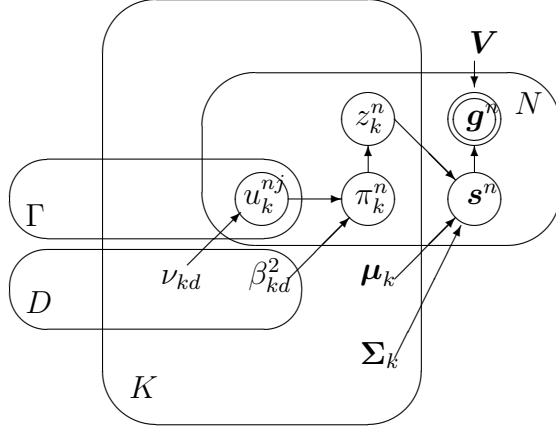
List of Figure Captions

Fig. 1. Graphical model for the mixture of truncated Gaussians (GMM-T) Stokes image restoration / segmentation model (see section 3). Stokes vectors $\boldsymbol{s}^n$ constitute the estimated restoration, produced by observations $\boldsymbol{g}^n$. Random variable sets u_k^{nj}, π_k^n constitute the smoothing prior set on the segmentation z_k^n. The Stokes vectors follow a truncated Gaussian distribution conditioned on the hidden variables $\boldsymbol{z}$. Superscripts $n \in [1, \dots N]$ and $j \in [1, \dots N]$ denote pixel indices, subscript $k \in [1, \dots K]$ denotes kernel (segment) index, subscript $d \in [1, \dots D]$ indexes the neighborhood direction type. At most Γ neighbors are defined for each site.

Fig. 2. Graphical model for spatially variant GMM on intrinsic parameters (GMM-IP) Stokes image restoration / segmentation model (see section 4). Stokes vectors $\boldsymbol{s}^n$ constitute the estimated restoration, produced by observations $\boldsymbol{g}^n$. Random variable sets u_k^{nj}, π_k^n constitute the smoothing prior set on the segmentation z_k^n. The transformed Stokes image parameters $\boldsymbol{\lambda}^n$ are assumed to be normally distributed conditioned on the hidden variables $\boldsymbol{z}$. Superscripts $n \in [1, \dots N]$ and $j \in [1, \dots N]$ denote pixel indices, subscript $k \in [1, \dots K]$ denotes kernel (segment) index, subscript $d \in [1, \dots D]$ indexes the neighborhood direction type. At most Γ neighbors are defined for each site.

Fig. 3. Recovery results for simulated Stokes data. From left to right, each column shows the four channels of (a) the degraded image $\boldsymbol{g}$ (SNR of 10 dB), (b) the original Stokes image $\boldsymbol{s}^\star$, (c) the non-complying to Stokes constraints pseudo-inverse recovery estimate. In (d) and (e) we see the restoration result $\hat{\boldsymbol{s}}$ obtained with our method, in (d) without the Stokes image transformation (section 3) and in (e) with the Stokes image transformation (section 4). The corresponding segmentations of the degraded image into $K = 2$ classes are shown at the top of columns (d) and (e). All Stokes channels share the same graylevel scale. fig3.eps.

Fig. 4. Recovery results for real Stokes data. From left to right, each column shows the four channels of (a) the observed image $\boldsymbol{g}$, (b) the non-complying to Stokes constraints pseudo-inverse recovery estimate. In (c) and (d) we see the restoration result $\hat{\boldsymbol{s}}$ obtained with our method, in (c) without the Stokes image transformation (section 3) and in (d) with the Stokes image transformation (section 4). The corresponding segmentations of the observed

image into $K = 6$ classes are shown at the top of columns (c) and (d). Columns (b), (c), (d) show from top to bottom s_1, s_2/s_1, s_3/s_1, s_4/s_1. fig4.eps.

Fig. 5. Intensity images acquired with four probing polarization states for the metallic car toy. fig5.eps.

Fig. 6. Processing results of a real image (metallic car toy). From top to bottom: degree of polarization (DOP) map computed from the pseudo-inverse (PI) solution (DOP values greater than one were truncated); non physically admissible pixels (in white, 12% of the total number of pixels) corresponding to the pseudo-inverse solution; DOP map corresponding to the GMM-T model. The GMM-T DOP-map is significantly smoother than the PI DOP-map, which is what is expected given the considered object. fig6.eps.

Fig. 1. Graphical model for the mixture of truncated Gaussians (GMM-T) Stokes image restoration / segmentation model (see section 3). Stokes vectors s^n constitute the estimated restoration, produced by observations g^n. Random variable sets u_k^{nj}, π_k^n constitute the smoothing prior set on the segmentation z_k^n. The Stokes vectors follow a truncated Gaussian distribution conditioned on the hidden variables z. Superscripts $n \in [1, \dots N]$ and $j \in [1, \dots N]$ denote pixel indices, subscript $k \in [1, \dots K]$ denotes kernel (segment) index, subscript $d \in [1, \dots D]$ indexes the neighborhood direction type. At most Γ neighbors are defined for each site.

Fig. 2. Graphical model for spatially variant GMM on intrinsic parameters (GMM-IP) Stokes image restoration / segmentation model (see section 4). Stokes vectors $\boldsymbol{s}^n$ constitute the estimated restoration, produced by observations $\boldsymbol{g}^n$. Random variable sets u_k^{nj}, π_k^n constitute the smoothing prior set on the segmentation z_k^n. The transformed Stokes image parameters $\boldsymbol{\lambda}^n$ are assumed to be normally distributed conditioned on the hidden variables $\boldsymbol{z}$. Superscripts $n \in [1, \dots N]$ and $j \in [1, \dots N]$ denote pixel indices, subscript $k \in [1, \dots K]$ denotes kernel (segment) index, subscript $d \in [1, \dots D]$ indexes the neighborhood direction type. At most Γ neighbors are defined for each site.
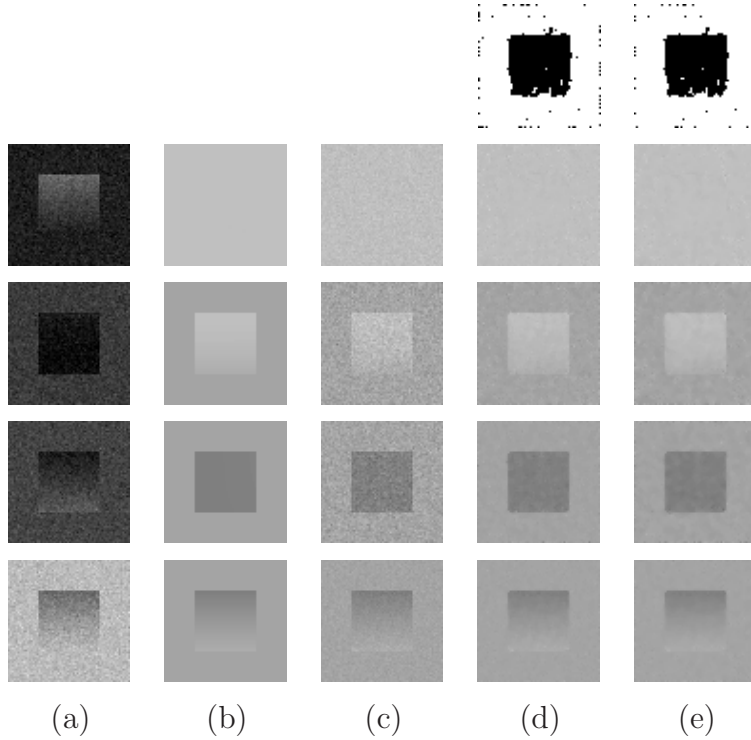
Fig. 3. Recovery results for simulated Stokes data. From left to right, each column shows the four channels of (a) the degraded image $\boldsymbol{g}$ (SNR of 10 dB), (b) the original Stokes image $\boldsymbol{s}^{\star}$, (c) the non-complying to Stokes constraints pseudo-inverse recovery estimate. In (d) and (e) we see the restoration result $\hat{\boldsymbol{s}}$ obtained with our method, in (d) without the Stokes image transformation (section 3) and in (e) with the Stokes image transformation (section 4). The corresponding segmentations of the degraded image into $K = 2$ classes are shown at the top of columns (d) and (e). All Stokes channels share the same graylevel scale. fig3.eps.
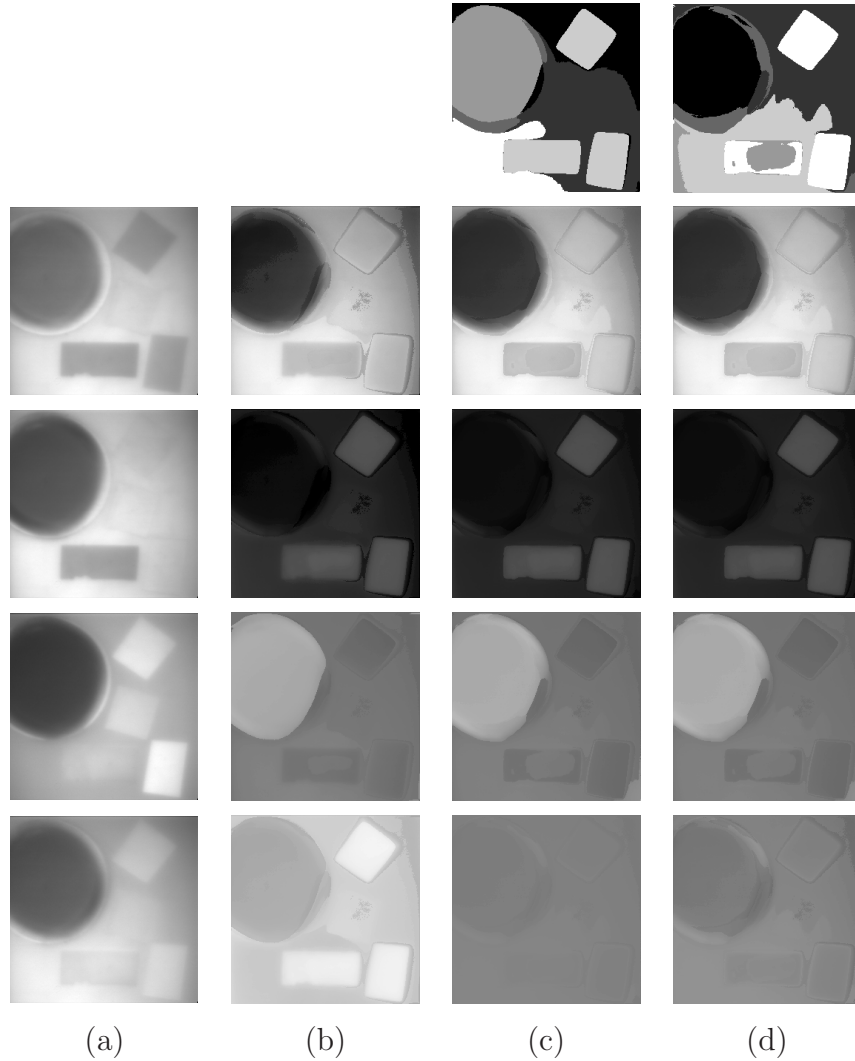
(a)   (b)   (c)   (d)

Fig. 4. Recovery results for real Stokes data. From left to right, each column shows the four channels of (a) the observed image $\boldsymbol{g}$, (b) the non-complying to Stokes constraints pseudo-inverse recovery estimate. In (c) and (d) we see the restoration result $\hat{\boldsymbol{s}}$ obtained with our method, in (c) without the Stokes image transformation (section 3) and in (d) with the Stokes image transformation (section 4). The corresponding segmentations of the observed image into $K = 6$ classes are shown at the top of columns (c) and (d). Columns (b), (c), (d) show from top to bottom s_1, s_2/s_1, s_3/s_1, s_4/s_1. fig4.eps.
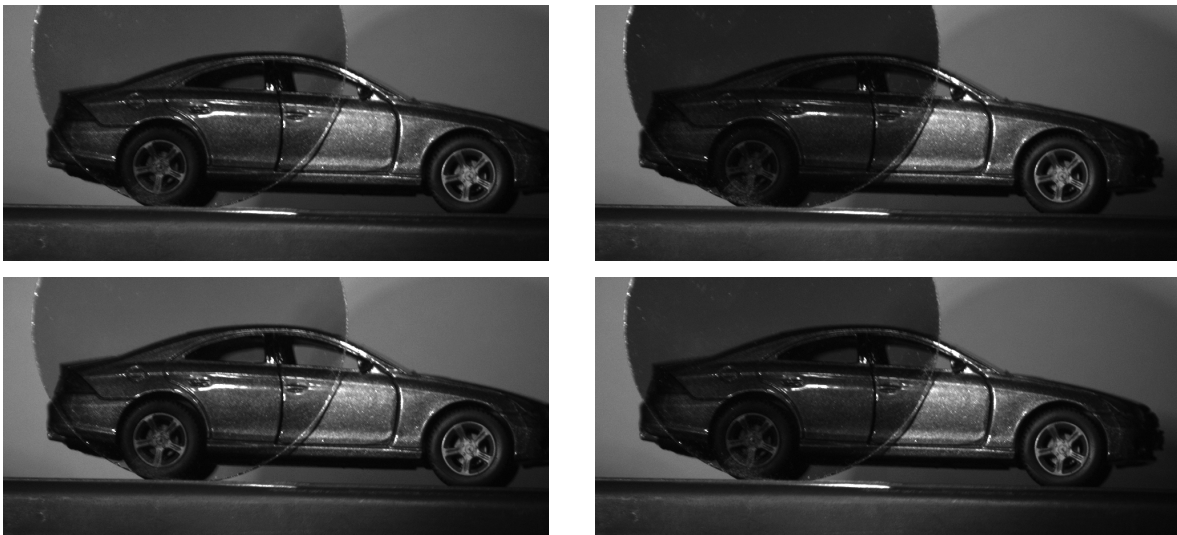
Fig. 5. Intensity images acquired with four probing polarization states for the metallic car toy. fig5.eps.
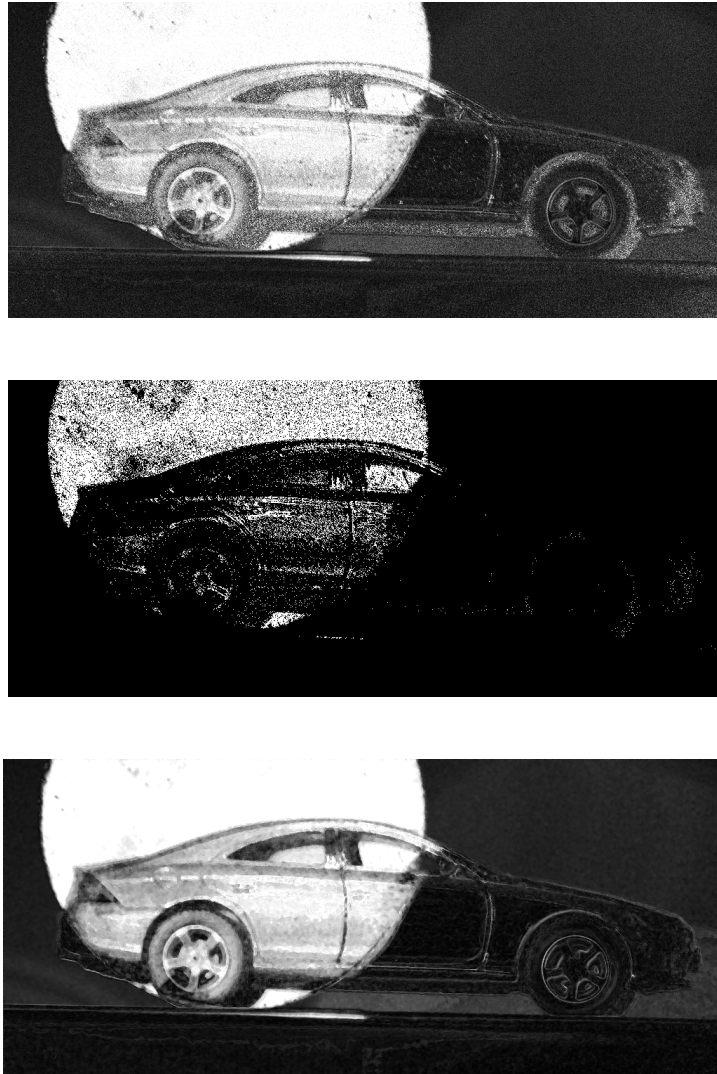
Fig. 6. Processing results of a real image (metallic car toy). From top to bottom: degree of polarization (DOP) map computed from the pseudo-inverse (PI) solution (DOP values greater than one were truncated); non physically admissible pixels (in white, 12% of the total number of pixels) corresponding to the pseudo-inverse solution; DOP map corresponding to the GMM-T model. The GMM-T DOP-map is significantly smoother than the PI DOP-map, which is what is expected given the considered object. fig6.eps.

Table 1. *Restoration error results for the simulated Stokes data of fig. 3. The image was degraded by varying noise levels. The presented values are the restoration ISNR. Higher values correspond to better restorations. Results are shown for various numbers of classes K of the underlying segmentation. PI stands for pseudo-inverse estimate.*

	PI	GMM-T			GMM-IP		
SNR		$K = 2$	$K = 3$	$K = 4$	$K = 2$	$K = 3$	$K = 4$
20 dB	18.3	**23.0**	22.0	22.4	20.6	20.2	22.7
10 dB	9.2	15.2	**15.8**	14.5	12.4	12.9	12.7
5 dB	4.4	11.9	**13.2**	12.4	9.7	10.9	9.6
1 dB	1.3	10.8	**11.9**	10.8	9.4	9.4	8.2

Table 2. *Runtime in seconds per each* EM *iteration, for the proposed algorithms. Results are shown for the artificial image (64 × 64, fig. 3), and the real image (256 × 256) of fig. 4.*

	GMM-T	GMM-IP
Artificial image	10 sec.	7 sec.
Real image	900 sec.	400 sec.