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. In this probabilistic framework, two distinct models for the sought polarized radiances are used. In the first model, 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. In the second one, the polarization at each site is represented by the coherency matrix whose positive semidefiniteness provides a convenient way to ensure the physical admissibility of the solution. This matrix is parameterized by a set of variables assumed to be generated by a spatially varying mixture of Gaussians. Inference for both models is obtained by the Expectation-Maximization (EM) algorithm. The restored Stokes images are always physically admissible which is not the case for the naïve pseudo-inverse approach. Numerical experiments on real and synthetic images using the proposed methods assess the pertinence of the approach. Experiments on noise-degraded images confirm the robustness to noise, both in terms of visual quality as well as SNR improvement.

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 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 recently in [5]. The latter model covers spatial coherence and edge-preservation, but the related parameters have to be fine-tuned empirically and it is highly time-consumming.

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, and 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 [6]. 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 representation 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, impose 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. Both models are inferred by the Expectation-Maximization algorithm. A preliminary short version of the second model is presented in [7].

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 following 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:

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

where $\boldsymbol{g}^n$ is 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. Also, variable $\boldsymbol{\epsilon}$ 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 $\boldsymbol{\epsilon}$ on each Stokes channel as zero mean, additive, white Gaussian. Formally this translates to

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

or

$$p\left(\boldsymbol{g}|\boldsymbol{s}\right) = \prod_{n=1}^{N} p\left(\boldsymbol{g}^n|\boldsymbol{s}^n\right),$$
$$p(\boldsymbol{g}^n|\boldsymbol{s}^n) = \mathcal{N}(\boldsymbol{g}^n|\boldsymbol{H}\boldsymbol{s}^n, \boldsymbol{V}),$$

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 k^{th} variate set to 1 if the corresponding n^{th} Stokes vector $\boldsymbol{s}^n$ belongs to the k^{th} 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 [8], 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 $\pi_k^n \geqslant 0$, $\forall k \in [1, \dots K]$, $\forall n \in [1, \dots N]$ and sum-to-unity $\sum_{k=1}^K \pi_k^n = 1$, $\forall n \in [1, \dots N]$ constraints.

We assume a Markov random field (MRF) on $\boldsymbol{\pi}$, which equivalently means that $\boldsymbol{\pi}$ is governed by a Gibbs distribution [9], 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 [8], amounts to an edge-preserving line-process [9]. The probability law for local differences is thus set to the Student's-t distribution:

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

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

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 a pixel's neighborhood adjacency types and $\gamma_d(n)$ is the index set of neighbors of pixel indexed by n, with respect to the d^{th} 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 [10].

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

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

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

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 [11], 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

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

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

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

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

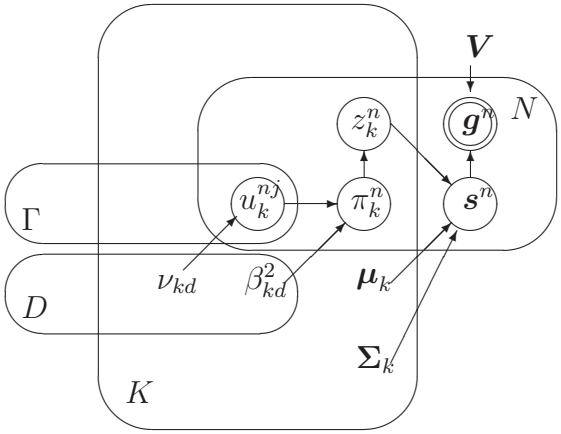

Fig. 1. Graphical model for the mixture of truncated Gaussians (GMM-T) Stokes image restoration / segmentation model (see section 3). Stokes vectors $\boldsymbol{s}^n$ constitute the estimated restoration, produced by observations $\boldsymbol{g}^n$. Random variable sets u_k^{nj}, π_k^n constitute the 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 $d \in [1, \dots D]$ describes the neighborhood direction type. At most Γ neighbors are defined for each site.

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

This may be addressed considering the hidden variables $\boldsymbol{z}$ (the labels) and $\boldsymbol{u}$ (the edge map) and using the EM algorithm [12]. 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 [12]:

E step: Evaluate the posterior probabilities of the hidden varibles $\boldsymbol{z}$ and $\boldsymbol{u}$ and compute the cost function Q defined by

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

M step: update the parameters as

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

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

A summary of our proposed restoration / segmentation algorithm steps can be surveyed 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

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

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:

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

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.

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 $\boldsymbol{\varphi}$.

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 (B6, B7) to compute initial estimates 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 $\boldsymbol{\beta}^{(t)}$ parameters using equation (A7).
- **3.2** Update $\boldsymbol{V}^{(t)}$ parameters using equation (A8).
- **3.3** Update $\boldsymbol{\pi}^{(t)}$ parameters by optimizing (A9).
- **3.4** Update $\boldsymbol{\nu}^{(t)}$ parameters by optimizing (A10).
- **3.5** Update $\boldsymbol{\mu}^{(t)}$, $\boldsymbol{\Sigma}^{(t)}$ parameters using as initial approximations eq. (B3, B4) and correcting on them as described in the text.
- **3.6** Update Stokes estimates $\boldsymbol{s}^{(t)}$. For each $n \in [1, N]$ do
  - **3.6.1** $\kappa \leftarrow 0$. Set $\phi^{(\kappa)}$ to pseudoinverse of g^n (eq. B6) and correct it if necessary using eq. (B7).
  - **3.6.2** Let $\rho^{(\kappa)}(h) = (1-h)\phi^{(\kappa)} + h[\boldsymbol{A}^{n(t)}]^{-1}\boldsymbol{b}^{n(t)}$.
  - **3.6.3** Find $h^\star \in [0,1]$ such that $h^\star$ as large as possible and such that $\rho^{(\kappa)}(h)$ satisfies the Stokes constraints (1).
  - **3.6.4** Let $\phi^{(\kappa+1)} \leftarrow \rho^{(\kappa)}(h^\star)$. $\kappa \leftarrow \kappa + 1$.
  - **3.6.5** Line search for $\phi^{(\kappa+1)}$ on a random direction. $\kappa \leftarrow \kappa + 1$.
  - **3.6.6** If $\phi^{(\kappa)}$ has converged, let $s^{n(t)} \leftarrow \phi^{(\kappa)}$ 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.

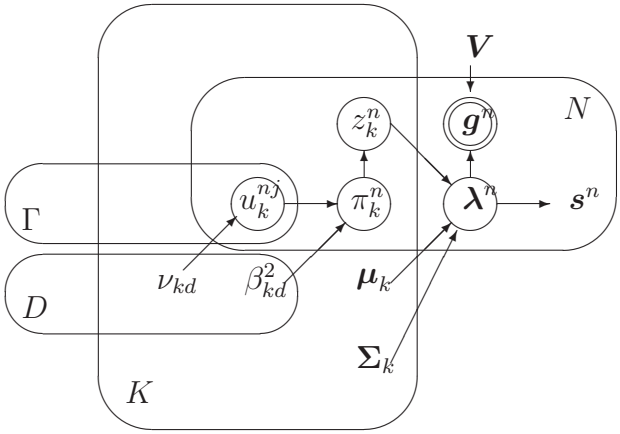

Fig. 2. Graphical model for spatially variant GMM on intrinsic parameters (GMM-IP) Stokes image restoration / segmentation model (see section 4). Stokes vectors $\boldsymbol{s}^n$ constitute the estimated restoration, produced by observations $\boldsymbol{g}^n$. Random variable sets u_k^{nj}, π_k^n constitute 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 $d \in [1, \dots D]$ describes the neighborhood direction type. At most Γ neighbors are defined for each site .

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 $\boldsymbol{\Psi} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\nu}\}$.

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

A summary of our proposed restoration / segmentation algorithm steps can be surveyed in Algorithm 2.

5. Numerical experiments

We have applied the proposed recovery algorithm to four test Stokes images, two artificial and two real images. On the artificial images, size 64×64, the experiment was conducted by reproducing the indirect noisy model (2) and by applying varying levels of noise. We also used different assumed numbers of underlying segments K. The obtained results are

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 (B6, B7) and (8) for initial $\boldsymbol{\lambda}^{(0)}$.

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 $\boldsymbol{\beta}^{(t)}$ parameters using equation (A7).
- **3.2** Update $\boldsymbol{V}^{(t)}$ parameters using equation (A8).
- **3.3** Update $\boldsymbol{\pi}^{(t)}$ parameters by optimizing (A9).
- **3.4** Update $\boldsymbol{\nu}^{(t)}$ parameters by optimizing (A10).
- **3.5** Update $\boldsymbol{\mu}^{(t)}$, $\boldsymbol{\Sigma}^{(t)}$ using equations (A5, A6).
- **3.6** Update $\boldsymbol{\lambda}^{(t)}$ parameters. 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** $\kappa \leftarrow \kappa + 1$ unless $\kappa = 4$. If $\kappa = 4$, and $\lambda^{n(t)}$ 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 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.

shown in table 1 and figures 3 and 4. In that table, we recall that the method relying on the GMM on the intrinsic parameters is represented by GMM-IP and the method relying on the mixture of truncated Gaussians is denoted by GMM-T. 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 degraded (observed) image respectively. Higher values for the ISNR correspond to better restorations. The pseudo-inverse estimates also computed for the restoration using (B6). The results clearly demonstrate that the proposed method gives a consistently better restoration compared to the general purpose pseudo-inverse estimator. Note also that the pseudo-inverse estimate will not necessarily yield values that satisfy the constraints (1). The main advantage of our method is that it takes explicitly into account the Stokes admissibility constraints (1), which is not the case for standard restoration methods.

Table 1. *Restoration error results on the simulated Stokes data of (a) fig. 3, (b) fig. 4. 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, as well as the result of the pseudo-inverse estimate (PI).*

	PI	GMM-IP			GMM-T		
SNR		$K=3$	$K=5$	$K=7$	$K=3$	$K=5$	$K=7$
$20dB$	21.4	**22.4**	22.3	22.7	17.8	20.5	21.1
$10dB$	11.4	14.7	14.5	14.5	16.4	16.9	**17.4**
$5dB$	6.3	10.9	9.9	9.9	14.0	14.3	**14.4**
$1dB$	2.3	7.3	7.6	7.6	**13.2**	13.0	12.3

(a)

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

(b)

Representative visual results for the artificial images are shown in figures 3 & 4. The pseudo-inverse estimate in all tests provides a solution of bad quality that infringes physical constraints on considerable number of pixel locations. Under the heavy-noise scenario of fig. 3, this figure goes up to 1316 inadmissible vectors out of a total of 4096 - i.e. almost one third of the image. On the other hand, both of our methods ensure always admissibility for *all* recovered vectors.

The first of the real images tested (courtesy of Dr. A. de Martino, fig.5) consists of an assembly of a number of objects with -as a first approximation- homogeneous polarimetric responses (background, metallic, dichroic and birefringent patches). 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 image (fig. 6) represents a metallic target-car toy with a circular dichroic patch imaged by an inline rotating quarter-wave plate Stokes polarimeter under natural illumination through a $10nm$ bandwidth interferential filter (low SNR). The two images were selected to evaluate the interest of our methods to handle piecewise constant signatures as well as varying ones under different illumination conditions. Figure 6 shows the pseudo-inverse solution as well as the outputs of our algorithms corresponding to the first image. For the second target, we present only the DOP (degree of polarization) images estimated from the pseudo-inverse Stokes solution and the GMM-T (fig.6) solution. Again, not only does our approach give a neat solution, but compared to the pseudo-inverse output which here yields about 20% non-physical pixels, it preserves the physical nature of the signal.

Note that in both artificial and real cases, the approach making use of truncated Gaussians (alg. 1) yields a smoother result. In the artificial image, where there is extra noise added, this is desirable. On the other hand the real image seems oversmoothed in comparison to our other proposed model (alg. 2), where this latter model seems to give the better result in that case. In view of table 1, it is confirmed that the model of alg. 1 does not work well in the low-noise scenario (20 dB), where a spatial grouping constraint is not that necessary as in cases where we have heavier noise.

In table 2 we see 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 $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ (section 3). In practice however, we found that keeping the initial values (B3), (B4) 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 dual core 1.8 GHz PC workstation with 2 Gb of RAM.

Table 2. *Runtime in seconds per each* EM *iteration, for the proposed algorithms. Results are shown for the artificial images (*64×64*) of fig.3 & fig.4, and the real image (*256×256*) of fig.5.*

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

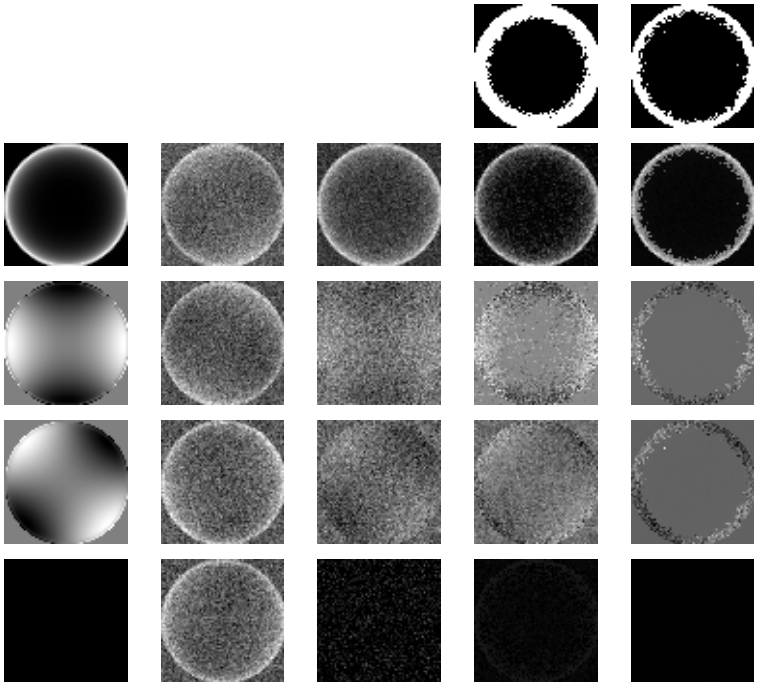

Fig. 3. Recovery result for simulated Stokes data under significant degradation. From left to right, each column shows the four channels of (a) the original Stokes image $\boldsymbol{s}^{\star}$, (b) the degraded image $\boldsymbol{g}$ (SNR of 5 dB), (c) the non-complying to Stokes constraints pseudo-inverse recovery estimate. In (d) and (e) we see the restoration result $\hat{\boldsymbol{s}}$ obtained with our method, in (d) using the Stokes image transformation (section 4) and (e) without the Stokes image transformation (section 3). The corresponding segmentations of the degraded image into $K = 2$ classes are shown at the top of the (d) and (e) columns.

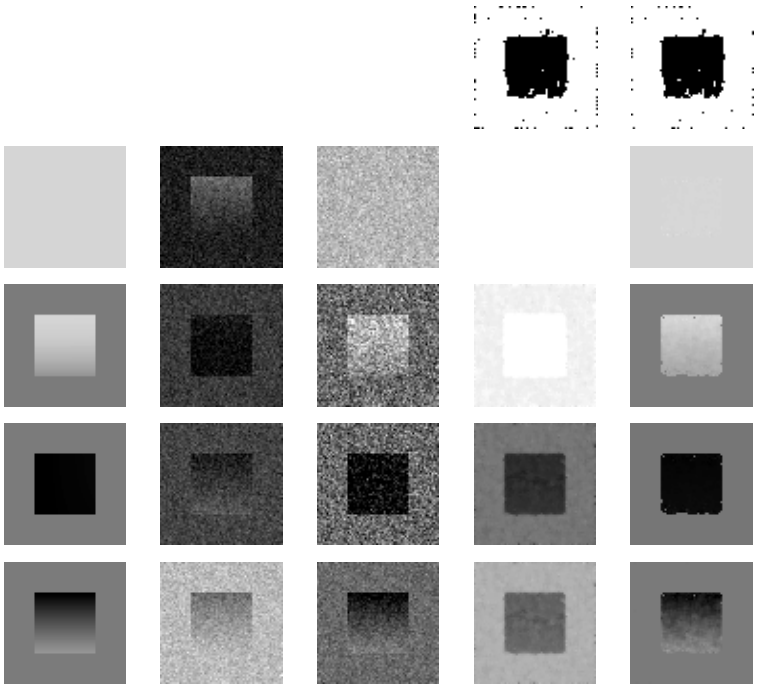

Fig. 4. Recovery result for simulated Stokes data. From left to right, each column shows the four channels of (a) the original Stokes image $\boldsymbol{s}^{\star}$, (b) the degraded image $\boldsymbol{g}$ (SNR of 10 dB), (c) the non-complying to Stokes constraints pseudo-inverse recovery estimate. In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

6. Conclusion

The modeling of polarimetric Stokes images as spatially varying mixtures of Gaussians was proposed. In this framework, two algorithmic schemes were presented. In the first approach, a Stokes image was parameterized by random variables generated by spatially varying Gaussian mixtures. The parameterization implicitly modeled the polarimetric constraints. In the second approach, the polarimetric constraints were considered through a mixture of truncated Gaussians. While the first model showed to be accurate in low SNR and low number of Gaussian components, the second model showed, in general, higher performances.

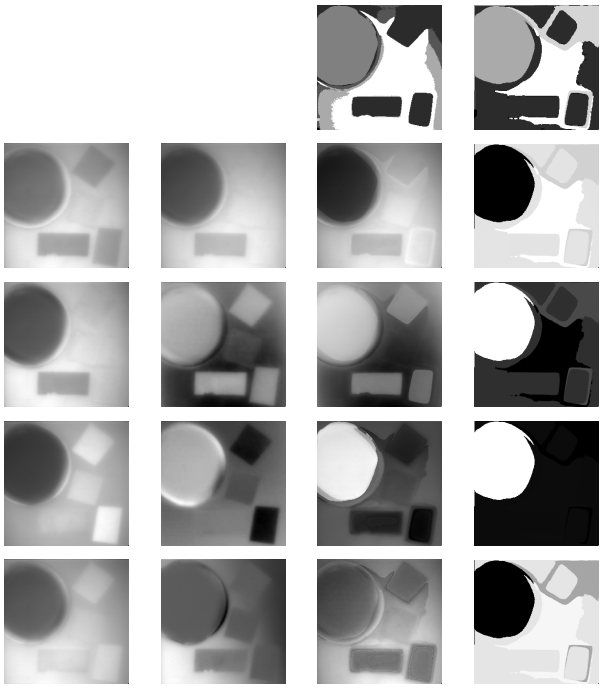

Fig. 5. Recovery result for real Stokes data. From left to right, each column shows the four channels of (a) the observed image $\boldsymbol{g}$, (b) the non-complying to Stokes constraints pseudo-inverse recovery estimate. In (c) and (d) we see the restoration result $\hat{\boldsymbol{s}}$ obtained with our method, in (c) using the Stokes image transformation (section 4) and (d) without the Stokes image transformation (section 3). The corresponding segmentations of the observed image into $K = 4$ classes are shown at the top of the (c) and (d) columns.

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

In this appendix, we detail the derivation of the E-step and M-step concerning the variables and parameters that the two models have in common.

E-step

The E-step consists in computing the expected log likelihood - expectation taken over the hidden variables $\boldsymbol{z}$ and $\boldsymbol{u}$ - with respect to current iteration parameters $\boldsymbol{\Psi}^{(t)}$, $\boldsymbol{\lambda}^{(t)}$, $\boldsymbol{\pi}^(t)$ where t indexes the current iteration. Observing the graphical model in figures 1 and 2, we can see that given $\boldsymbol{\pi}$ and the other observed variables, $\boldsymbol{z}$ and $\boldsymbol{u}$ are conditionally independent.

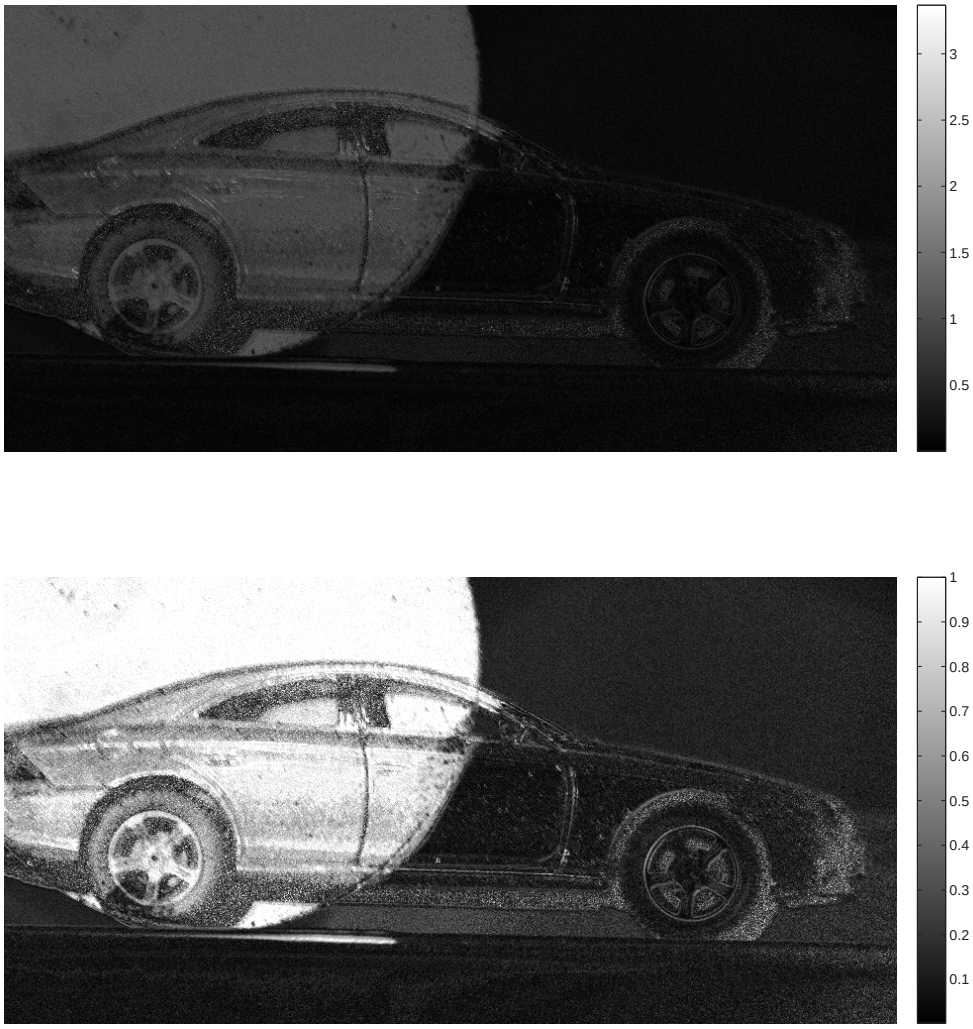

Fig. 6. Degree of polarization (DOP) maps for restorations of a real image (metallic object, toy car). Top image shows result for pseudo-inverse estimate. Bottom image shows result for GMM-T proposed algorithm. Note the abnormal value range (> 1) for the pseudo-inverse solution; on the other hand GMM-T always ensures admissibility.

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 $\boldsymbol{s}$ in the previous formula. Substituting the posteriors of the hidden variables in the cost 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 $<\ln \boldsymbol{u}>$. These expectations are computed with respect to $p(\boldsymbol{z}, \boldsymbol{u}|\boldsymbol{g}, (\boldsymbol{s}, \boldsymbol{\pi})^{(t)}; \boldsymbol{\Psi}^{(t)})$.

Below we show the induced E step updates $\forall n, k, \forall j \in \gamma_d(n)$ for GMM-IP. Likewise, the same formulae hold for GMM-T where we have to replace $\boldsymbol{\lambda}$ by $\boldsymbol{s}$, where found (that is, eq. A1):

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

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

$$<\ln u_k^{nj}>^{(t)} = \digamma(\zeta_k^{nj(t)}) - \ln \eta_k^{nj(t)} \tag{A3}$$

where $\digamma(\cdot)$ stands for the digamma function, and parameters ζ, η being:

$$\zeta_k^{nj(t)} = \frac{1}{2}\left(\nu_{kd}^{(t)} + 1\right)$$

$$\eta_k^{nj(t)} = \frac{1}{2}\left(\nu_{kd}^{(t)} + \frac{(\pi_k^{n(t)} - \pi_k^{j(t)})^2}{\beta_{kd}^{2(t)}}\right)$$

M-step

Maximization of the current complete likelihood (7) must be driven with respect to the model parameters $\boldsymbol{\Psi}$,$\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$. With some manipulation, we can rewrite it equivalently as

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

$$+\mathcal{E}_{\boldsymbol{u}|\boldsymbol{\pi}}\{\ln p(\boldsymbol{\pi}|\boldsymbol{u}; \boldsymbol{\beta})\} + \mathcal{E}_{\boldsymbol{u}|\boldsymbol{\pi}}\{\ln p(\boldsymbol{u}; \boldsymbol{\nu})\} \tag{A4}$$

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

$$\boldsymbol{\mu}_k^{(t+1)} = \frac{\sum_{n=1}^N <z_k^n>^{(t)} \boldsymbol{\lambda}^n}{\sum_{n=1}^N <z_k^n>^{(t)}} \tag{A5}$$

$$\boldsymbol{\Sigma}_k^{(t+1)} = \frac{\sum_{n=1}^N <z_k^n>^{(t)} (\boldsymbol{\lambda}^n - \boldsymbol{\mu}_k^{(t+1)})(\boldsymbol{\lambda}^n - \boldsymbol{\mu}_k^{(t+1)})^T}{\sum_{n=1}^N <z_k^n>^{(t)}} \tag{A6}$$

$$\beta_{kd}^{2(t+1)} = \frac{\sum_{n=1}^N \sum_{j \in \gamma_d(n)} <u_k^{nj}>^{(t)} (\pi_k^{n(t)} - \pi_k^{j(t)})^2}{\sum_{n=1}^N |\gamma_d(n)|} \tag{A7}$$

$$\boldsymbol{V}^{(t+1)} = (4N)^{-1} \sum_{n=1}^{N} (\boldsymbol{H}\boldsymbol{g}^n - \boldsymbol{s}^{n(t)})^T (\boldsymbol{H}\boldsymbol{g}^n - \boldsymbol{s}^{n(t)}) I. \tag{A8}$$

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

$$a_k^n \left(\pi_k^{n(t+1)}\right)^2 + b_k^n \left(\pi_k^{n(t+1)}\right) + c_k^{n(t+1)} = 0 \tag{A9}$$

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 [13]. However, the solutions of eq. (A9) for a given pixel indexed by n will not in general 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 (7) with respect to the degrees of freedom equal to zero we obtain $\nu_{kd}^{(t+1)}$ as the solutions of the equation:

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

$$+ \left[\frac{\sum_{n=1}^{N} \sum_{j \in \gamma_d(n)} (<\ln u_k^{nj}>^{(t)} - <u_k^{nj}>^{(t)})}{\sum_{n=1}^{N} |\gamma_d(n)|} \right] + 1 = 0 \tag{A10}$$

with $\digamma(\cdot)$ being again the digamma function.

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.

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 obtained 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

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

where

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

so that the distribution normalizes correctly to unity. Noticing that this latter expression can be written as the expectation $\mathcal{E}_{\boldsymbol{\sigma}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k}\{\mathcal{I}(\boldsymbol{\sigma})\}$, we can employ a Monte Carlo approximation to estimate it. Thus $\Xi(\cdot)$ approximates as

$$\Xi(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \simeq \frac{1}{T}\sum_{\tau=1}^{T} \mathcal{I}(\boldsymbol{\sigma}^{(\tau)}) \tag{B2}$$

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 [12]. Thence, approximation (B2) 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 [14].

Nevertheless as an initial approximation $\tilde{\mu}$, $\tilde{\Sigma}$ of means and covariances for *Nelder-Mead* we use the updates

$$\tilde{\boldsymbol{\mu}}_k^{(t+1)} = \frac{\sum_{n=1}^{N} <z_k^n>^{(t)}\ \boldsymbol{s}^n}{\sum_{n=1}^{N} <z_k^n>^{(t)}}, \tag{B3}$$

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

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 $\boldsymbol{s}$.

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 conditional expectation of the complete likelihood (eq. A4). With some manipulation, it can be seen that equivalently we are to maximize for $s^{n(t+1)}$ the quadratic equations $\forall n \in [1, N]$:

$$\boldsymbol{s}^{n(t+1)^T}\boldsymbol{A}^{n(t+1)}\boldsymbol{s}^{n(t+1)} - 2\boldsymbol{b}^{n(t+1)^T}\boldsymbol{s}^{n(t+1)}, \tag{B5}$$

with $\boldsymbol{s}^{n(t+1)}$ such that Stokes constraints (eq. 1) are always respected. 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 (B5), 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

$$\boldsymbol{s}^{n(0)} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{g}^n. \tag{B6}$$

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

$$(s_1^n)^2 = (s_2^n)^2 + (s_3^n)^2 + (s_4^n)^2 \tag{B7}$$

To find proper directions of search, we take advantage of the fact that (B5) 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 of 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 local optima. 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 (7):

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

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

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

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

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

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 $n \in [1, \dots N]$, 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. P. 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, 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.
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. Zallat, C. Collet, and Y. Takakura, "Clustering of polarization-encoded images," *Applied Optics*, vol. 43, no. 2, pp. 283–292, 2004.
7. 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.
8. 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.

9. 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.
10. 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.
11. G. McLachlan, *Finite mixture models*. Wiley-Interscience, 2000.
12. C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
13. G. Birkhoff and S. MacLane, *A survey of modern algebra*. New York: McMillan, 1953.
14. 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.