A Bayesian approach for polarimetric data reduction: the Mueller imaging case

Jihad Zallat, Christian Heinrich, and Matthieu Petremand
Laboratoire des Sciences de l’Image, de l’Informatique et de la Télédétection (LSIIT, UMR CNRS-ULP 7005)
Parc d’Innovation, Bd Sébastien Brant, B.P. 10413, F-67412 Illkirch cedex, France.
zallat@lsiit.u-strasbg.fr

Abstract: In this paper, we extend to the Mueller imaging framework a formerly introduced Bayesian approach dealing with polarimetric data reduction and robust clustering of polarization encoded images in the piecewise constant case. The extension was made possible thanks to a suitable writing of the observation model in the Mueller context that relies on the system’s coherency matrix and Cholesky decomposition such that the admissibility constraints are easily captured. This generalization comes at the cost of nonlinearity with respect to the parameters that have to be estimated. This estimation-clustering problem is tackled in a Bayesian framework where a hierarchical stochastic model based on a Markov random field proposed by Potts is used. This fully unsupervised approach is extensively tested over synthetic data as well as real Mueller images.

© 2008 Optical Society of America

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

References and links

1. Introduction
In a previous paper [1], we reported on a new method addressing polarimetric data reduction in the Stokes imaging framework. Our main contribution there concerned the use of a Bayesian approach ensuring the physical authenticity of polarization encoded images obtained from raw intensity acquisitions while providing a robust clustering of the imaged scene. We showed also that our approach preserved the Stokes channels from image noises and errors induced by the
imaging system. The problem of the more complex Mueller imaging case remained pending. It is addressed in the present article.

The importance of Mueller imaging comes from the rich set of information it provides about the local nature of the target which may be of great use in studying number of physical phenomena that arise in many fields of science. Hence, an increasing interest is devoted to the development of new imaging systems that record spatially dependent patterns of polarized light diffusely scattered or transmitted by scenes or samples.

Mueller imaging is defined as the distributed measurement of Mueller matrices attached to each pixel. This is made possible by the use of effective polarization state analyzers (PSAs) that acquire the Stokes vectors corresponding to each pixel in the image while a source lights up the scene through a polarization state generator (PSG). Different configurations using different polarization modulation techniques are used in imaging polarimetry. Whatever the technique used, the principle remains the same: a set of raw intensity images have to be inverted to yield polarization encoded images. This procedure called "polarimetric data reduction" (PDR) is achieved through a linear observation model that relates the two sets of images via the system dependent PSA and PSG matrices.

It is worth noting that a proper choice of the PSA and the PSG matrices plays an important role in PDR [2]. It ensures a good conditioning that allows the inversion of the system and minimizes the contamination of polarimetric channels by image noise. However, if one uses the classical PDR techniques, even for well behaved imaging polarimeters [2], the image noise propagates into the estimated Mueller matrices and may prevent their physical admissibility. Indeed, a $4 \times 4$ real matrix must satisfy some physical constraints to be bona fide.

The contribution of the present work is twofold. Firstly, we propose an original parameterization of the forward model in order to account for the physical admissibility conditions. This leads to a nonlinear observation model. Secondly, we propose a fully unsupervised Bayesian framework (model and algorithm) addressing the estimation problem. We suppose that the sought object is piecewise constant, which corresponds to many situations of practical interest. The Markov random field model originally proposed by Potts will account for the hidden segmentation map holding the spatial prior knowledge. We will also assume the Mueller matrices to be class wise constant. Estimation will take place in the generalized maximum likelihood framework where the a posteriori probability law is optimized jointly with respect to the parameters and the hyperparameters.

The article is organized as follows: the nonlinear parameterization of the forward problem and the stochastic model are presented in section 2. The estimation procedure is detailed in section 3. Simulation results and real data processing are presented and discussed in section 4. Conclusion and perspectives are given in section 5.

2. Problem statement

2.1. Observation model in the Mueller imaging polarimetry framework

A Mueller imaging polarimeter allows the indirect measurement of the Mueller matrix attached to each pixel in the image. This is achieved by illuminating an object by $p$ independent polarization states where $p \geq 4$ through a PSG. Each of these states interacts with the object to yield an outgoing Stokes vector that is sensed through the PSA by $q$ independent probing states where $q \geq 4$. For each pixel location $(j_y, j_x)$ in the image, the $p \times q$ intensity measurements $I$ are related to the pixel’s Mueller matrix $M$ by the following linear equation:

$$I(j_y, j_x) = AM(j_y, j_x)W,$$

where $A$ is a $p \times 4$ PSA matrix whose rows constitute the $p$ probing states and $W$ is the $4 \times q$ PSG matrix whose columns represent the $q$ incident polarization states.
It can be shown [2] that Eq. (1) can be written in the following form

$$ I(j_y, j_x) = (W^t \otimes A) M(j_y, j_x) = P(\eta) M(j_y, j_x), \tag{2} $$

where the underline symbol stands for the operator which maps a \( p \times q \) matrix into an \( pq \times 1 \) vector formed by stacking up its columns and \( \otimes \) represents the Kronecker product. The known \( pq \times 16 \) matrix \( P \), also called "Polarization Measurement Matrix" (PMM), depends on a parameter vector \( \eta \) which controls the combined actions of the PSG and PSA and relates linearly the Mueller parameters vector to the raw measured intensity vector. The multi-component nature of the Mueller matrices image is treated here by seeing the Mueller image as a result of a multi-channel imaging system where the information is conveyed to the user by a multi-channel image data. These channels will be indexed in the following, where appropriate, by \( j_n \in [0, 15] \).

Regardless of the physical constraints that must be satisfied by the elements of \( M \), we have the same starting point as with the Stokes imaging framework [1]: the two problems share the same observation model. Systematic calibration errors as well as noises that arise from the use of a CCD photosensitive device as a detector (readout noise, thermal noise, dark current noise, and photon noise among others) are accounted for in the same way as in [1], which leads to the complete observation model:

$$ I(j_y, j_x) = (1 + \rho) P(\eta) M(j_y, j_x) + \epsilon(j_y, j_n), \tag{3} $$

where \( \rho \) accounts for kinds of noises and errors having their variances directly proportional to the measured intensity (see [1] for more details) while \( \epsilon \) encompasses all remaining noise terms.

The problem is now to estimate \( M \) from the observations \( I \) while ensuring the physical admissibility of the retrieved Mueller operators attached to each pixel. One can easily understand that the classical approach that relies on the PMM pseudo-inverse matrix is not the right way to achieve this task. Indeed, it necessitates a pixel-wise Mueller validity post checking which is time consuming without giving a proper estimate for pathological pixels.

We propose a new parameterization of the problem, yielding an observation model with simple constraints that allows tackling the estimation in a Bayesian framework. This comes at the cost of nonlinearity. The parameterization is achieved by observing that there is a one-to-one correspondence between a \emph{bona fide} Mueller matrix \( M \) and a Hermitian positive semidefinite matrix \( H \) called "system coherency matrix"

$$ H = \sum_{k,l=0}^{3} m_{kl} [\sigma_k \otimes \sigma_l], \tag{4} $$

where \( \{\sigma_k\}_{k=0}^{3} \), comprising the Pauli matrices and the identity matrix, are the elements for a basis for \( \mathbb{C}^{2 \times 2} \) and \( m_{kl} \) are the elements of \( M \) [3]. The linear equation (4) leads to the following expression:

$$ M = TH, \tag{5} $$

where \( T \) is a \( 16 \times 16 \) invertible complex transformation matrix relating the stacked Mueller matrix to the stacked system’s coherency matrix.

It was stated that \( M \) is physically acceptable if and only if \( H \) is positive semi-definite [3]. Furthermore, if \( H \) is positive semidefinite, it can be written in terms of its Cholesky decomposition as [4]

$$ H = \Lambda \Lambda', \tag{6} $$
where \( \Lambda \) is a lower triangular matrix composed of 16 real parameters \( \{ \lambda_i \} \)

\[
\Lambda = \begin{bmatrix}
\lambda_0 & 0 & 0 \\
\lambda_4 + i \lambda_5 & \lambda_1 & 0 \\
\lambda_{10} + i \lambda_{11} & \lambda_6 + i \lambda_7 & \lambda_2 \\
\lambda_{14} + i \lambda_{15} & \lambda_{12} + i \lambda_{13} & \lambda_8 + i \lambda_9 & \lambda_3
\end{bmatrix}.
\]

(7)

It is now possible to parameterize the elements of \( H \) by \( \{ \lambda_i \} \) in Eq. (5) and report the expression of \( M \) in the observation model of Eq. (3) to obtain

\[
L(j_y, j_x) = (1 + \rho) P(\eta) T H(\{ \lambda_i \}; j_y, j_x) + \varepsilon(j_y, j_x).
\]

(8)

We must note that the observation model is now a nonlinear function of \( \{ \lambda_i \} \) and that \( \{ \lambda_i, j \in [0, 3] \} \) must obey a positivity constraint to respect the positive definiteness of \( H \). The estimation problem turns now to the estimation of the real parameters \( \{ \lambda_i \} \) that are used to construct \( H \), hence yielding a physically acceptable Mueller matrix. We note also that generalization to the case where the matrix \( P \) depends on the site \( (j_y, j_x) \) is straightforward.

2.2. Underlying segmentation model

The underlying segmentation map will be modeled as a Markov random field, and more precisely using Pott’s model [5]. Let

\[
I_0 = \{ I_0 (j_y, j_x); I_0 (j_y, j_x) \in [1, K], j_y \in [1, n], j_x \in [1, m] \}
\]

(9)

denote the segmentation image, where \( K \) is the number of classes. The prior probability associated to \( I_0 \) reads

\[
p(I_0|\beta) = \frac{1}{Z(\beta)} \exp (-\beta U(I_0)),
\]

(10)

where

\[
U(I_0) = -\sum_{t=3} \mathbb{I}(i_t(i_t)=i_{q(t)})
\]

(11)

is the abstract energy measure associated to the field and \( Z \) the partition function. Variables \( t \) and \( s \) denote pixel sites of the image, the \( \mathbb{I} \) denotes neighboring sites (we used eight neighbors in the present work) and \( \mathbb{I}(\cdot) \) is the indicator function taking values 1 or 0 whether the condition between parentheses is true or false. This model favors piecewise constant images, the size of the patches being governed by the abstract inverse temperature parameter \( \beta \). A key problem in using Eq. (10) as a penalty term is the computation of \( Z \) as a function of \( \beta \). We rely here on thermodynamic integration, i.e., on the relation

\[
\frac{d \log Z}{d \beta} = \frac{1}{Z} \frac{dZ}{d\beta} = \sum_{I_0} -U(I_0) \frac{\exp (-\beta U(I_0))}{Z} = E[-U(I_0)],
\]

(12)

where the expectation is taken with respect to \( p(I_0|\beta) \).

From a theoretical point of view, we have

\[
\log Z = nm \log K + \int_{0}^{\beta} E[-U(I_0)] d\beta'.
\]

(13)

From a practical point of view, values of \( E[-U(I_0)] \) are computed using Markov chain Monte Carlo approximations for different values of \( \beta \). The corresponding set of points is linearly interpolated, thus defining an approximation of \( \frac{d \log Z}{d \beta} \) [6] which is further integrated.
thus giving an approximation of $\log Z$. Let us mention that the samples of $I_0$ are generated by alternating Gibbs sampling (pixel-flipping) raster scans of the image with Swendsen-Wang iterations (cluster-flipping). The role of the latter iterations is to improve the mixing properties of the Markov chain. We will not further detail this standard procedure, the reader being referred to [5].

Let $\Phi$ be the parameter matrix defined as

$$\Phi = \begin{bmatrix} \lambda_0 (1) & \ldots & \lambda_0 (K) \\ \vdots & \ddots & \vdots \\ \lambda_{15} (1) & \ldots & \lambda_{15} (K) \end{bmatrix},$$  \hspace{1cm} (14)$$

where the numbers between parentheses stand for the class index. In the estimation procedure to be described, the role of the segmentation map $I_0$ will be to index the appropriate column $k$ of candidate Mueller class parameters $\{\tilde{\lambda}_i (k)\}$ for pixels labeled as class $k$: $I_0 (j_y, j_x) = k$. This straightforwardly yields the vector $H$ attached to such pixel.

2.3. Noise model

The noise will be supposed to be independent and Gaussian distributed according to

$$\varepsilon (j_y, j_x | I_0 (j_y, j_x)) \sim N \left( \mu^{(\varepsilon)}_{I_0 (j_y, j_x)}, \left( \sigma^{(\varepsilon)}_{I_0 (j_y, j_x)} \right)^2 \right),$$  \hspace{1cm} (15)$$

where $\left( \mu^{(\varepsilon)}_{I_0 (j_y, j_x)}, \sigma^{(\varepsilon)}_{I_0 (j_y, j_x)} \right)$ are the noise parameters (mean and standard deviation) that affect the $k^{th}$ class in the $j_n$ channel image.

Let us define $\Theta$ as the set $\Theta = \{ \Theta^{(j_n)}; j_n \in [0, 15] \}$, where

$$\Theta^{(j_n)} = \begin{bmatrix} \mu^{(j_n)}_1 \\ \mu^{(j_n)}_2 \\ \vdots \\ \mu^{(j_n)}_K \\ \sigma^{(j_n)}_1 \\ \sigma^{(j_n)}_2 \\ \vdots \\ \sigma^{(j_n)}_K \end{bmatrix}.$$  \hspace{1cm} (16)$$

All these parameters are unknown and will have to be estimated. This general model will in the sequel be restricted to the case of vanishing means. From a modelling point of view, the model with non vanishing means suffers from an identifiability problem: modifying some or all the $\tilde{\lambda}_i$’s amounts to modifying the $\mu_j$’s and we have, at least for now, no information (prior or measured) to be able to make a decision. Furthermore, this assumption does not penalize at all the algorithm performances as observed in practice.

2.4. General probabilistic model

The overall probabilistic model may be depicted using a Directed Acyclic Graph (DAG) (see Fig. 1).

For the sake of clarity, we drop the dependency on $K$, since the only role of $K$ is in defining dimensions. Considering the DAG, we may write:

$$p (I, \Phi, I_0 | \rho, \Theta, \beta, \sigma_\lambda) = p (I | \Phi, I_0, \rho, \Theta) \ p (\Phi | \sigma_\lambda) \ p (I_0 | \beta).$$  \hspace{1cm} (17)$$

The component terms on the right-hand-side are reckoned as follows:

- $p (I_0 | \beta)$ is established by Eq. (10) and section 2.2;
Fig. 1. The directed acyclic graph representing the dependence relations between the variables. According to usual conventions, square boxes represent fixed or observed quantities whereas circles represent quantities to be estimated. More precisely, $\beta$ is the abstract inverse temperature parameter of the Markov random field, $\sigma_\lambda$ is the standard deviation of the $\lambda_i$'s, $\rho$ is related to the quantum efficiency of the camera, $I_0$ is the underlying class segmentation map, $\Phi$ is the Mueller parameter matrix encompassing the $\lambda_i(\eta)$'s, $\Theta$ is the matrix encompassing the parameters of the noise distribution, and $I$ is the intensity observations matrix.

- $p(\Phi|\sigma_\lambda)$ is defined in a class-independent manner as follows:

\[
\begin{align*}
\lambda_i & \sim \mathcal{N}(0, \sigma_\lambda^2) \quad \mathbb{I}(\lambda_i \geq 0) \quad \forall i \in [0, 3], \\
\lambda_i & \sim \mathcal{N}(0, \sigma_\lambda^2) \quad \forall i \in [4, 15].
\end{align*}
\]

The variance $\sigma_\lambda^2$ is chosen sufficiently large in order to have no influence. In other words, the prior $p(\Phi|\sigma_\lambda)$ is flat compared to the likelihood $p(I|\Phi, I_0, \rho, \Theta)$ when the $\lambda_i(\eta)$ vary. Thus the prior $p(\Phi|\sigma_\lambda)$ has no influence besides imposing the positivity of $\lambda_0, \ldots, \lambda_3$ when estimating $\Phi$. The variance $\sigma_\lambda^2$ may be estimated using physical considerations or using a set of $I$'s (from this set of $I$'s, we compute the $H$'s in a pseudo-inverse framework and estimate the $\lambda_i$'s by minimizing $\|H - \Lambda \lambda\|^2$). This would give a set of $\lambda_i$'s from which we could derive $\sigma_\lambda^2$.

- the data-dependent term $p(I|\Phi, I_0, \rho, \Theta)$ is obtained by solving Eq. (8) for $\varepsilon$ then using Eq. (15) to express the (joint) probability density function of $\varepsilon$, assuming iid of pixels $I(j_y, j_x)$ conditioned on $I_0$:

\[
\begin{align*}
p(I|\Phi, I_0, \rho, \Theta) &= p(\varepsilon) = p\left(I - (1 + \rho) P(\eta) T H\left(\{\lambda_i\}\right)\right) \\
&= \prod_{j_y, j_x, j_n} \mathcal{N}\left(\zeta(j_y, j_x, j_n), \left(\sigma_{I_0(j_y, j_x)}\right)^2\right),
\end{align*}
\]

where

\[
\zeta(j_y, j_x, j_n) = I(j_y, j_x, j_n) - (1 + \rho) P(\eta) T H\left(\{\Phi[I_0(j_y, j_x)]\}, j_y, j_x\right)_{j_n}.
\]

In other words, the $\lambda_i(\eta)$ are indexed from $\Phi$ (per Eq. (14)) based on the class index of the pixel, which in turn is looked-up from the segmentation map $I_0$. 

#91555 - $15.00 USD
Received 9 Jan 2008; revised 6 Feb 2008; accepted 11 Feb 2008; published 2 May 2008
(C) 2008 OSA
12 May 2008 / Vol. 16, No. 10 / OPTICS EXPRESS 7124
3. Estimation procedure

All available information is contained in distribution (17), from which we will derive the estimates $\hat{\Phi}, \hat{I}_0, \hat{\rho}, \hat{\Theta}, \hat{\beta}$. Let us emphasize that this method is fully unsupervised: there is no free parameter to be chosen or tuned by the user, set aside the number $K$ of classes (see discussion below).

Let $\{\rho, \Theta, \beta\}$ be the set of hyperparameters (i.e., the parameters which are not directly of interest), whereas the parameters are $I_0$ and $\Phi$. The parameters and hyperparameters will be obtained simultaneously by joint optimization of distribution (17), thus corresponding to generalized maximum likelihood estimation whereas maximum a posteriori would correspond to optimizing the distribution with respect to the parameters for given values of the hyperparameters.

To address the optimization problem, we resort to the Gauss-Seidel scheme: in the global loop, we consider all variables in a row and optimize $-\log p(I, \Phi, I_0 | \rho, \Theta, \beta, \sigma_\lambda)$ with respect to the current variable while the other variables are held fixed. The update of $I_0$ is achieved site by site in a raster scan. The update of $\Phi$ is achieved component by component (i.e., we loop on $i$ and $k$ and optimize with respect to $\lambda_i(k)$). Convergence is reached when $I_0$ is not modified by an entire raster scan (in fact, we will monitor that $I_0$ is held constant over several (10 in the present case, see the counter < 10 condition in Algorithm 2-b) raster scans, since even if $I_0$ is held constant, the other variables may vary, if only a little, which in turn may induce a modification of $I_0$).

Optimization with respect to $\lambda_i(k)$ deserves particular comments. This optimization involves a fourth degree polynomial obtained by extracting the corresponding terms from $-\log p(I|\Phi, I_0, \rho, \Theta) - \log p(\Phi|\sigma_\lambda)$. This is handled using formal computation: the coefficients of the polynomial are analytically precomputed and the local minima are computed given the numerical values of the coefficients. The updated value of $\lambda_i(k)$ is sought among a) the arguments of the local minima when no positivity constraint is involved or b) the positive arguments of the local minima and zero when a positivity constraint is involved.

The present optimization problem being highly multimodal, particularly with respect to $I_0$ when the noise level is important, we suggest using the algorithm detailed below which behaves better than straightforward and sole Gauss-Seidel descent.

At first, twenty different runs of the K-means clustering procedure are applied to the image set of $m \times n$ dimensional intensity channel vectors, yielding different class segmentation maps. The best resulting map, characterized by minimal distances between the clusters and the class’s centroids, is retained as the initial value of $I_0$ and used to initialize the DAG with the procedure detailed in Algorithm 1-a and Algorithm 1-b. In this phase, a whole scan of $I_0$ may lead to numerous modifications since the state of the DAG is far from its final (equilibrium) state. We want to avoid such a large movement without updating the rest (i.e., all variables but $I_0$) of the DAG, since it may lead to local optima far from the solution. Thus, we update the other variables of the DAG as soon as a given (reduced) number of site updates are achieved (Algorithm 1-a). This allows reaching a state much closer to the equilibrium than the initial state. We then switch to entire raster scans (Algorithm 1-b) on which we loop until convergence. The role of the mask is to improve the speed of the procedure. The new mask (see the instruction update the mask) is composed of all sites that have just been modified and of the neighbors of those sites. At this stage, the state is not optimal: a better (in the sense of a lower criterion value) solution may be found. This state cannot be reached by updating single sites at a time. This is why we resort to a multiresolution optimization scheme (Algorithm 2-a), where the sites of $I_0$ are modified blockwise. The other parameters of the DAG remain fixed in this phase, since they are close to their final values. We conclude with a standard Gauss-Seidel procedure (Algorithm 2-b) until convergence.
Algorithm 1: Optimization phase 1: initialization.

4. Algorithm application

4.1. Simulation results

The aim of this section is to compare the estimated Mueller images resulting from different approaches (K-means, pseudo-inverse, Bayesian) using a ground-truth. We synthesized a non trivial 128-pixels × 128-pixels Mueller image by using a four class label map shown in Fig. 2. A different Mueller matrix was assigned to each label as shown in Table 1. For the Bayesian approach, we considered two Markov random field models: Potts’ model and the Markov mesh (unilateral model) considered in the previous work addressing the Stokes imaging case.

The PMM of an optimal rotating-retarder Mueller polarimeter using the optimal parameter vector \( \eta_{opt} \) as defined in Ref. [2] was used to generate the corresponding intensity images by mean of Eq. (2). The hypothesis and arguments about the image noise nature introduced in our previous paper [1] are still valid here. Thus we assume that \( \varepsilon \) follows a centered Gaussian distribution: a zero mean white Gaussian noise with variance \( \sigma^2_e \) was added to each intensity.
Algorithm 2: Optimization phase 2: refinement. In algorithm (a), all pixels of a given block are forced to the same label.

For this simulation providing the necessary inputs to our model, i.e. intensity images as well as the PMM $P$. We present here simulations results carried out with two different noise variances ($\sigma^2_n = 0.1$ and $\sigma^2_n = 0.5$) which correspond respectively to signal-to-noise ratios over the mean transmittance image (first Mueller channel) about 10 dB and 3 dB.

The 16 channels of the Mueller image obtained by the classical pseudo-inverse solution are shown in Fig. 3. The results given by different approaches are summarized in Tables 2 and 3. Table 2 shows the percentage of misclassified pixels in the image. Table 3 presents the relative estimation error of the Mueller matrices associated to each class defined as

$$RE = \frac{||M_{true} - M_{estimated}||}{||M_{true}||}.$$  \hspace{1cm} \text{(21)}$$

The Frobenius norm is used in the definition of $RE$.

We calculated also the mean $RE$ value for the 3 dB case over the whole image for the classi-
Fig. 2. 128-pixels × 128-pixels mask used to generate the synthetic Mueller images. Each color (white, light gray, heavy gray and black) corresponds to one label.

Table 1. The four different matrices assigned to the classes of the label map of Fig. 2. Matrix (a) was assigned to the black pixels, matrix (b) to the heavy gray pixels, matrix (c) to the light gray pixels and matrix (d) to the white pixels in the label map.

Matrix (a):
\[
\begin{bmatrix}
1.2934 & 0.0014 & 0.0006 & 0.0002 \\
0 & 0.7043 & 0.0382 & 0.0199 \\
0.0097 & -0.0123 & -0.2619 & -0.0981 \\
0.0085 & 0.0161 & 0.1234 & -0.2832
\end{bmatrix}
\]

Matrix (b):
\[
\begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

Matrix (c):
\[
\begin{bmatrix}
1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1
\end{bmatrix}
\]

Matrix (d):
\[
\begin{bmatrix}
1 & 0.0118 & -0.0118 & 0 \\
-0.0471 & 0.4 & 0.0412 & 0.0118 \\
0.0176 & 0.0059 & -0.3 & 0.0412 \\
0.0059 & 0.0588 & -0.0353 & -0.2059
\end{bmatrix}
\]

cal pseudo-inverse PDR approach as well as for the Bayesian one (Potts’ model) and obtained respectively 110% for the first one and 2.2% for the second one. When the Mueller matrix is estimated with the pseudo-inverse, 90% (resp. 99%) of the pixels are not physically admissible for \( \sigma_n^2 = 0.1 \) (resp. \( \sigma_n^2 = 0.5 \)). Moreover, for the pseudo-inverse estimate, several Mueller channels display no information (see for example column 3 and row 3 in Fig. 3), whereas structural information appears in all channels for the Bayesian estimate.

We conclude that the introduced Bayesian approach (Potts’ model) gives better results in terms of estimated quantities and structural information (composition) of the imaged scene as compared to other approaches (pseudo-inverse and K-means). Moreover, the Markov random field prior plays an important role (see Table 2, where the results significantly degrade for \( \beta = 0 \)). The two Bayesian variants (Potts’ model and the unilateral model) give nearly the same results in term of RE but Potts’ model is noticeably better in term of misclassified pixels (Table 2). Besides, further simulations showed that Potts’ model is more robust than the unilateral one with respect to increasing noise variance.

In real situations, for some limited scenario, the optimal number of classes (\( K_{opt} \)) can only be roughly estimated. Usually, it can’t be inferred easily by simple observation. Our approach is able to handle this issue. To determine \( K_{opt} \), the optimum of the law in Eq. (17) is plotted against \( K \) (i.e., we plot \(-\log \left( p \left( \tilde{I}, \tilde{J}_0 | \tilde{\rho}, \tilde{\Theta}, \tilde{\beta}, \sigma_\lambda \right) \right) + constant \) against \( K \)). We chose for \( K_{opt} \) the
Fig. 3. The 16 channels of the synthetic Mueller image obtained by the commonly used pseudo-inverse estimate obtained for the noise of variance $\sigma_n^2 = 0.5$.

Table 2. Mean percentage of misclassified pixels for the Bayesian approach (with $(\hat{\beta})$ and without $(\beta = 0)$ Markov random field prior) and for K-means clustering applied to the intensity images. Averaging has been done over noise realizations. Numbers between parentheses represents standard-deviation of related quantities.

<table>
<thead>
<tr>
<th>Noise variance $\sigma_n^2$</th>
<th>0.1</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potts’ model $(\hat{\beta})$</td>
<td>0.05</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Unilateral model $(\hat{\beta})$</td>
<td>0.40</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Potts’ and unilateral model $(\beta = 0)$</td>
<td>2.3</td>
<td>(0.2)</td>
</tr>
<tr>
<td>K-means</td>
<td>2.3</td>
<td>(0.2)</td>
</tr>
</tbody>
</table>

value for which the graph levels off as shown in Fig. 4. In a fully Bayesian framework, we would have $K_{opt} = \arg\max_K p(I|K)$. This would necessitate to resort to Markov chain Monte Carlo methods which is outside the scope of this paper. This is left for future work.

Running the method takes 4 minutes (resp. 30 minutes) for $K = 4$ (resp. $K = 8$) classes (128 × 128 image, 2.8 GHz standard desktop computer, programmation in C language). This important variation in CPU time is due to an increase in the number of iterations.

4.2. Results against measured data

The proposed model is applied to real measured images acquired by an inline rotating quarter-wave plate Mueller imaging polarimeter. The object used to carry out these measurements consists of two dichroic patches (Polaroid) and a transparent thin film (cellophane) sandwiched between two coverglasses. The overall mount is imaged through a narrow band interferential
Table 3. Mean values of the relative error $RE$ (in percent) of the four Mueller matrices corresponding to the four classes of Fig. 2. Averaging has been done over noise realizations. The first values correspond to $\sigma^2_n = 0.1$, the second values correspond to $\sigma^2_n = 0.5$. The K-means were used to derive a segmentation map over which the pseudo-inverse estimate was class-wise averaged, yielding the $\tilde{M}_k$'s. We ensured physical admissibility conditions over those averages according to $\hat{M}_k = T \hat{H}_k$, with $\hat{H}_k = \hat{\Lambda}_k \hat{\Lambda}_k^T$ and $\hat{\Lambda}_k = \arg\min\Lambda \| \tilde{M}_k - T \Lambda \Lambda^T \|^2$, the minimization being achieved under the admissibility conditions.

<table>
<thead>
<tr>
<th>Mueller matrix</th>
<th>Table 1(a)</th>
<th>Table 1(b)</th>
<th>Table 1(c)</th>
<th>Table 1(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potts</td>
<td>0.7</td>
<td>1.7</td>
<td>0.36</td>
<td>0.85</td>
</tr>
<tr>
<td>Unilateral</td>
<td>0.72</td>
<td>2.0</td>
<td>0.35</td>
<td>0.84</td>
</tr>
<tr>
<td>K-means</td>
<td>0.9</td>
<td>6.7</td>
<td>0.4</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Fig. 4. The optimal number of classes for the simulated data. The graph corresponds to $\sigma^2_n = 0.5.$

filter (10 nm) and a PSA in front of a 12 bits radiometric CCD camera. We point out that the dichroic patches are placed such that their transmission axes are nearly orthogonal. Totally polarized incident light undergoes multiple reflections between the two coverglasses and escapes highly depolarized such that the polarization signature is expected to be relatively weak. This situation was chosen to represent a hard test case that can be encountered in many practical situations such as in polarization imaging of biological tissues.

The Mueller image obtained by classical PDR of this object is shown in Fig. 5. At first, we can observe that while the conventional intensity imaging provides little to no information about the object ($m_{00}$ channel in the Mueller image), the other Mueller channels provide a rich enhancement to classical imaging. We observe also that the Mueller image is highly contaminated by a noise that hampers the physical interpretation of the interaction mechanisms which would allow inferring the object properties.

The algorithm output is given in Fig. 6 and in Table 4. We point out that all retrieved Mueller matrices are physically valid. This can be verified easily by computing the eigenvalues spectra of their corresponding coherency matrices. Indeed, all the eigenvalues are real positive. In the other hand, Fig. 7 shows the variation of the optimum of Eq. (17) versus the number of classes. The shape of this curve indicates that the optimal number of classes is here 5. This is in complete agreement with the a priori knowledge.

Once the Mueller matrices are estimated, one can focus on the physical interpretation of the
Fig. 5. The Mueller image of the test object as obtained from classical pseudo-inverse PDR. The image in the upper left ($m_{00}$) element corresponds to a conventional intensity image.

Fig. 6. The final map obtained by the Bayesian approach (Potts’ model). Each gray level corresponds to one different class associated to a different Mueller matrix. The circular cross section delimits the footprint of the illuminating beam. Classes (2) and (3) are the dichroic patches and class (4) corresponds to the thin cellophane film.
Table 4. The five retrieved Mueller matrices that correspond to the label map output of our algorithm. These matrices have to be interpreted with the class numbering that is shown in Fig. 6.

<table>
<thead>
<tr>
<th>Background (class 1)</th>
<th>Dichroic polaroid 1 (class 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0001 -0.0019 0.0002 -0.0001</td>
<td>1.0013 -0.0499 0.0119 0.0000</td>
</tr>
<tr>
<td>0.0015 0.0078 0.0007 -0.0002</td>
<td>-0.0575 0.0446 -0.0110 -0.0005</td>
</tr>
<tr>
<td>-0.0003 0.0008 -0.0017 -0.0019</td>
<td>-0.0223 0.0151 -0.0082 -0.0022</td>
</tr>
<tr>
<td>-0.0002 -0.0001 0.0014 -0.0038</td>
<td>-0.0006 -0.0006 0.0033 -0.0066</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dichroic polaroid 2 (class 3)</th>
<th>Cellophane (class 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9996 0.0331 0.0192 -0.0010</td>
<td>0.9998 -0.0130 -0.0112 -0.0161</td>
</tr>
<tr>
<td>0.0432 0.0575 0.0213 0.0004</td>
<td>-0.0144 0.1931 0.0172 -0.0119</td>
</tr>
<tr>
<td>-0.0177 -0.0143 -0.0209 -0.0036</td>
<td>0.0027 0.0023 -0.1404 -0.1199</td>
</tr>
<tr>
<td>-0.0005 -0.0013 0.0038 -0.0148</td>
<td>0.0041 -0.0409 0.1065 -0.1423</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Internal region (class 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0003 -0.0093 -0.0215 -0.0073</td>
</tr>
<tr>
<td>-0.0196 0.2471 0.0238 0.0369</td>
</tr>
<tr>
<td>-0.0034 0.0219 -0.2440 0.0004</td>
</tr>
<tr>
<td>-0.0005 -0.0022 0.0004 -0.2544</td>
</tr>
</tbody>
</table>

Fig. 7. The optimal number of classes for the true image. We observe that the graph is consistent with the *a priori* knowledge of the object.

results. Several analysis tools are available in the literature. We cite among others the coherency matrix decomposition approach [3] and the polar decomposition [7]. For this image we observe that the five obtained matrices depolarize a large amount of any incoming polarized state which is in agreement with the physical situation. We can also compute the product of the two matrices $M_2$ and $M_3$ corresponding to the second and third classes (which are weak polarizers oriented orthogonally to each other) to obtain

$$M_2M_3 = \begin{bmatrix} 1 & 0.03 & 0.01 & 0.00 \\ -0.05 & 0.00 & 0.00 & 0.00 \\ -0.02 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix},$$ (22)
which is very close to a pure depolarizer in accordance with the physics of the object (transmit-
ing axes nearly orthogonal).

Finally, let us mention that running the algorithm takes 2 hours for $K = 4$ classes ($256 \times 256$ image, 2.8 GHz standard desktop computer, programmation in C langage).

5. Conclusion and perspectives

In this paper, we present a polarimetric data reduction method addressing piecewise constant images in the Mueller framework imagery. The proposed Bayesian method enforces the physical admissibility conditions while providing a robust clustering of the estimated images. The results obtained are significantly better and more robust than those given by classical pseudo-inverse or K-means estimation. This framework will enable the researcher to enhance and deepen the understanding of the object constituents. Extension of the approach to smoothly varying polarization signatures is under investigation.

The software that processes Stokes and Mueller images is available on request for interested readers wishing to process their own images.

Acknowledgement

The authors are indebted to two anonymous reviewers for helpful comments and suggestions and acknowledge the financial support of the Agence Nationale de la Recherche (ANR Polarimetry).