



Expectation–maximization algorithm with total variation regularization for vector-valued image segmentation

Jun Liu^a, Yin-Bon Ku^b, Shingyu Leung^{b,*}

^a School of Mathematical Sciences, Laboratory of Mathematics and Complex Systems, Beijing Normal University, Beijing 100875, PR China

^b Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

ARTICLE INFO

Article history:

Received 20 February 2012

Accepted 3 September 2012

Available online 10 September 2012

Keywords:

Gaussian mixture model

Expectation–maximization

Total variation

Unified cost functional

Image segmentation

Vector-valued images

Fast algorithm

Alternative minimization

ABSTRACT

We integrate the total variation (TV) minimization into the expectation–maximization (EM) algorithm to perform the task of image segmentation for general vector-valued images. We first propose a unified variational method to bring together the EM and the TV regularization and to take advantages from both approaches. The idea is based on operator interchange and constraint optimization. In the second part of the paper we propose a simple two-phase approach by splitting the above functional into two steps. In the first phase, a typical EM method can classify pixels into different classes based on the similarity in their measurements. However, since no local geometric information of the image has yet been incorporated into the process, such classification in practice gives unsatisfactory segmentation results. In the second phase, the TV-step obtains the segmentation of the image by applying a TV regularization directly to the clustering result from EM.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Image segmentation is a basic yet very important task in image processing and computer vision. A large body of algorithms has been developed for these applications. One popular group of methods has been proposed based on partial differential equations (PDE) and variational principles. For example, to name just a few, Rudin et al. [47] have proposed regularizing a gray scale image by the total variation (TV) norm. With the level set method [42], such norm has also been used to regularize the boundary of different segments in the geodesic active contour [13] or the Chan–Vese model [16] in image segmentation. Another group of image processing methods consists of statistical-based approaches, such as wavelet, Markov random fields (MRF) and so on. Each approach has its own advantage: statistical methods are good at dealing with the randomness of noise while the PDE and variational methods can incorporate geometric information more naturally. However, techniques we mentioned above have been developed mostly for gray-level intensity images. The aim of this paper is to introduce new and efficient algorithms for image segmentation of general vector-valued images, which combine both variational and statistical approaches. A vector-valued image can be represented by an image function $\mathbf{f} : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R}^d$, where Ω is the set of all pixels of

the image and each vector component (channel) of \mathbf{f} corresponds to a measurement of the image. For color images, $d = 3$ and the three vector components are red, green and blue color intensities in the RGB model. For texture images, vector components can be some local geometric features extracted by a bank of Gabor filters. The dimension of such images might be extremely high and the scales between different channels might not be identical. Numerous methods have been proposed to generalize or to extend those techniques from gray-level images ($d = 1$) to various vector-valued image ($d > 1$) processing. Among them, PDE and variational formulation approaches have been very popular. For example, Sapiro [48] has proposed the color snake model, while Chan et al. [15] have also generalized the original active contour without edges to deal with color image segmentation. For color image denoising, Blomgren and Chan [6] have generalized the original ROF model by defining the so-called color TV-norm. Based on this color TV-norm, various approaches have been developed in [4,8,22]. Extending the color TV-norm, Christiansen et al. [17] have recently developed a vector TV-norm for matrix-valued images. In [29], image is regarded as an embedded surface in \mathbb{R}^{2+d} and an area-minimizing flow is used to perform denoising. Vese and Chan [56] have applied level set method to the famous Mumford–Shah model [38]. Generalizing the Mumford–Shah energy, Kay and Tomasi [27] have proposed a color image segmentation functional based on a vector-valued Allen–Cahn phase-field method. A general framework for vector-valued regularization with PDE has been proposed in [55]. A recent extension of this approach has been developed in [46].

* Corresponding author.

E-mail addresses: jliu@bnu.edu.cn (J. Liu), maybku@ust.hk (Y.-B. Ku), masyleung@ust.hk (S. Leung).

Many other PDE-based methods have also been developed [18,30,58,49,28,57,53], etc. Recently, some other methods have been proposed to combine EM algorithm and TV regularization [10,59], mainly for reconstructing images with Poisson noise.

Methods mentioned above work well for color images or vector-valued images with more or less a pre-determined scale for each channel. In particular, one can argue that the weight from each channel are the same in a color image. However, there is no natural way to define the TV-norm or active contour models for a general vector-valued image whose channels have different scales that are not known a priori. In fact, for general high-dimensional data, statistical-based approaches are more natural and their main advantage is that the randomness information originated from the noise, features in natural images, texture and so on is explicitly taken into consideration. But unlike the PDE and variational formulation approaches where geometric features such as edges and boundaries can be dealt with ease, it is not straightforward to incorporate local geometric information of an image into a statistical model.

One major contribution of this paper is to combine both statistical and variational approaches by a unified segmentation functional. Theoretically, the fundamental principles of the EM algorithm (which is statistical) and the TVL¹ algorithm (which is variational) are very different and it is not easy to bring these two concepts together in a natural way. One way to overcome the difficulty is to reinterpret the EM algorithm using a constraint optimization framework [32,33]. In this paper, we are going to show that the EM-TVL¹ process corresponds to a saddle point problem. Moreover, searching for such a saddle point with an alternating minimization algorithm would lead to both the EM step and the regularized TV step. The proposed EM-TV segmentation algorithm integrates the advantages of both the EM and the TV-based level set method such as CV model [16]. Compared with EM, this proposed method has a geometrical constraint, which makes this algorithm robust for noise. On the other hand, compared with the level set method, the statistical information is taken into the model and it is more suitable for natural images. In addition, our method can conveniently handle multi-clusters and does not need any extra method to reinitialize the level set function.

In the second part of the paper, we propose a reduced algorithm by proposing a two-phase algorithm to decouple the EM phase and the TV regularization step. The main idea is as follows. In this first phase, regarding \mathbf{f} as a set of d -dimensional data, we use statistical method (EM) to classify them into different classes. Suppose there are K classes. Define $u : \Omega \mapsto \{1, 2, \dots, K\}$ by $\mathbf{f}(x)$ belongs to the k th class, where k is the labeling of the pixel at x . Next, in step two, we apply the TVL¹ model to regularize u into \tilde{u} . Moreover, it can be shown that \tilde{u} is again a piecewise constant function like u , i.e., $\tilde{u} : \Omega \mapsto \{1, 2, \dots, K\}$ and hence this regularized classification \tilde{u} represents the final image segmentation of \mathbf{f} . There are various nice characteristics of the proposed algorithm. Similar to our earlier work in [31], the proposed method combines both the local geometry information and the global statistics of vector-valued image. Based on recent developed fast algorithms for TV-regularization, we can efficiently incorporate local geometry into the regularization.

The outline of the paper is as follows. In Section 2, we first introduce the Gaussian mixture model (GMM) and briefly explain how the classification can be done using the EM algorithm. Then in Section 3, we consider the TVL¹ regularization and describe various important properties for our algorithms. In Section 4, we first propose a unified functional to naturally combine two approaches into one single energy with four variables. By applying the convex relaxation technique and the dual method, we solve the resulting saddle point problem by an alternating minimization algorithm. This gives an efficient EM-TV algorithm for image segmentation. To simplify this algorithm, we propose in Section 5 a two-phase method to decouple the GMM and the TV regularization. In

Section 6, extensive numerical experiments are conducted to demonstrate the effectiveness of our algorithms.

2. Gaussian mixture model and EM algorithm

Gaussian mixture model (GMM) and EM algorithm have been widely used in image processing community [12,64,43,25,19,62,63]. For image segmentation, the main idea is to classify the measurement at pixels by modeling them using a linear mixture of Gaussian distributions. Therefore, the algorithm extracts global statistics from the whole data set. Another advantage of such segmentation method is that the measurements at each pixel can be from high dimensions, while the implementation is very simple.

In statistics, a Gaussian mixture model [5,44,36] uses a linear mixture of Gaussian distributions to model the phenomenon of interest. It is particularly well-suited for image segmentation due to its simplicity and power of representation. Let $\mathbf{y} \in \mathbb{R}^d$ be a d -dimensional measurement vector at a pixel. A GMM models its distribution using the following form

$$p(\mathbf{y}) = \sum_{k=1}^K \alpha_k p_k(\mathbf{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad \text{with} \quad \sum_{k=1}^K \alpha_k = 1, \quad (1)$$

where K is the total number of mixtures, α_k 's are the weights of each Gaussian component, and $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ are the mean vector and covariance matrix of the k th Gaussian distribution probability density function p_k , respectively. The number of components, K , is usually assumed to be known. Otherwise, it is possible to use a model selection procedure to determine a good K along with estimations of other model parameters [44].

The Gaussian assumption is flexible in capturing the image dynamics due to the mixing properties. In practice, Gaussian or near-Gaussian noises are the most prominent noise types. Furthermore, it is possible to use nonlinear transformation to make the transformed signal be of more Gaussian mixture-like [7]. We will show in the experiment section that our algorithms can still get good result even if K is over-estimated.

The expectation-maximization (EM) constitutes an efficient algorithm to obtain parameter estimates in maximizing the likelihood function. Given the image vectors $\mathbf{f}(x_1), \dots, \mathbf{f}(x_N)$ at the set of pixels $\{x_1, \dots, x_N\}$. We assume that all the vectors are independent of each other. Suppose the initial guess for the Gaussian mixture model is $\alpha_k^{(0)}, \boldsymbol{\mu}_k^{(0)}, \boldsymbol{\Sigma}_k^{(0)}$, then the EM algorithm iterates between the following two steps till the parameter estimates converge ($t = 0, 1, \dots$) [37]:

- E-step. For $i = 1, \dots, N$

$$\phi_k^{(t+1)}(x_i) = \frac{\alpha_k p_k(\mathbf{f}(x_i); \boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Sigma}_k^{(t)})}{\sum_{l=1}^K \alpha_l p_l(\mathbf{f}(x_i); \boldsymbol{\mu}_l^{(t)}, \boldsymbol{\Sigma}_l^{(t)})}, \quad (2)$$

- M-step

$$\begin{aligned} \alpha_k^{(t+1)} &= \frac{\sum_{i=1}^N \phi_k^{(t+1)}(x_i)}{N}, & \boldsymbol{\mu}_k^{(t+1)} &= \frac{\sum_{i=1}^N \phi_k^{(t+1)}(x_i) \mathbf{f}(x_i)}{\sum_{i=1}^N \phi_k^{(t+1)}(x_i)}, \\ \boldsymbol{\Sigma}_k^{(t+1)} &= \frac{\sum_{i=1}^N \phi_k^{(t+1)}(x_i) (\mathbf{f}(x_i) - \boldsymbol{\mu}_k^{(t)})' (\mathbf{f}(x_i) - \boldsymbol{\mu}_k^{(t)})}{\sum_{i=1}^N \phi_k^{(t+1)}(x_i)}. \end{aligned} \quad (3)$$

After obtaining estimated parameters given the image data, the usual way to segment an image is to classify pixels into the class (or segment) with the largest conditional probability of each pixel x_i :

$$\begin{aligned} \text{label}(\mathbf{f}(x_i)) &= \arg \max_{1 \leq k \leq K} \phi_k(x_i) \\ &= \arg \max_{1 \leq k \leq K} \alpha_k p_k(\mathbf{f}(x_i); \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \end{aligned} \quad (4)$$

The GMM can be viewed as a global method in the sense that all data contribute evenly to the final parameter estimate. The EM algorithm is computationally fast; hence, it is suitable for many image or data analysis. However, one shortcoming of this model is that the statistical model is solely based on the distribution of the data and the classification does not utilize any neighborhood coherence or geometric information. Moreover this model does not take into account of any adaptivity ability according to the local statistic variation. These information might be crucial for image segmentation and other data analysis. In [31], we have proposed a simple statistical method that directly incorporate adaptivity and spatial and geometric information, and furthermore, we have combined the EM algorithm and a neighborhood based penalty mechanism with a variational method in [34]. Using neighborhood and geometric information, such as local orientation and anisotropy, we have proposed an optimal adaptive local filtering strategy which can both improve the robustness of the method and also keep fine features in the image. In this paper, however, we explore some other possibilities in regularizing the classification.

3. TVL¹ model

In this section, we summarize a popular image regularization model for gray-level intensity images which is based on the TV-norm and an L¹ fidelity.

Variational models for image regularization have become very influential since the work by Rudin, Osher and Fatemi (ROF) [47]. The idea behind the ROF model is to consider the following minimization problem:

$$\inf_{v \in BV(\Omega)} \mathcal{E}_2(v, u, \lambda) \quad \text{where } \mathcal{E}_2(v, u, \lambda) = \int_{\Omega} |\nabla v| dx + \lambda \int_{\Omega} |u - v|^2 dx \quad (5)$$

where λ is a positive real parameter, $u : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ is the original noisy gray scaled image in L^2 and $BV(\Omega)$ is the space of all functions with bounded variations on Ω .

However, the standard ROF model (5) has one limitation: it will cause the loss of contrast even for a noise-free original image – suppose the image is a disk, then the regularized image is the same disk but with a lower intensity (see [51] for more details). Indeed, significant improvements have recently been made based on the Bregman iterations [40,24,60,41]. In this paper, however, we will concentrate on the following related minimization problem. A small but very significant change is to replace the L^2 fidelity by a L^1 fidelity,

$$\inf_{v \in BV(\Omega)} \mathcal{E}_1(v, u, \lambda) \quad \text{where } \mathcal{E}_1(v, u, \lambda) = \int_{\Omega} |\nabla v| dx + \lambda \int_{\Omega} |u - v| dx \quad (6)$$

for any given image $u \in L^1(\Omega)$. This model usually refers to the TVL¹ model and it has been seriously studied in tremendous literatures [2,3,39,14] to name just a few.

This model has the following interesting properties which turn out to be desirable for the current application:

1. (Affine invariance) Suppose v is a minimizer of $\mathcal{E}_1(\cdot, u, \lambda)$. Then for any real numbers a and b , $av + b$ is a minimizer of $\mathcal{E}_1(\cdot, au + b, \lambda)$.
2. (Contrast invariance) Let g be an increasing C^1 -diffeomorphism with $\sup_{\mathbb{R}} |g'| < +\infty$. Then v is a minimizer of $\mathcal{E}_1(\cdot, u, \lambda)$ if and only if $g \circ v$ is a minimizer of $\mathcal{E}_1(\cdot, g \circ u, \lambda)$.

We assume that u is piecewise constant with K levels, i.e.

$$u = \sum_{k=1}^K c_k \cdot \mathbf{1}_{D_k} \quad (7)$$

where D_1, \dots, D_K are K bounded and disjoint domains that partition Ω and c_1, \dots, c_K are K distinct positive real numbers. The following theorem says that there exists a minimizer of $\mathcal{E}_1(\cdot, u, \lambda)$ that has the same form as u .

Theorem 3.1. Suppose $u \in BV(\Omega)$ is of the form (7). Then there exists a minimizer of $\mathcal{E}_1(\cdot, u, \lambda)$ such that it is of the form $\sum_{k=1}^K c_k \cdot \mathbf{1}_{D'_k}$, where D'_1, \dots, D'_K are K bounded (possibly empty) and disjoint domains that partition Ω .

To prove this theorem, we first need to understand $\mathcal{E}_1(v, u, \lambda)$ through the following layer-cake formula [14]:

$$\mathcal{E}_1(v, u, \lambda) = \int_{-\infty}^{+\infty} \text{Per}(V_t) + \lambda |V_t \Delta U_t| dt \quad (8)$$

where $V_t = \{x \in \Omega | v(x) > t\}$ and $U_t = \{x \in \Omega | u(x) > t\}$ are the upper level sets of v and u , respectively. $\text{Per}(V_t)$ is the perimeter of V_t , which is well-defined since $v \in BV(\Omega)$. $V_t \Delta U_t$ is the symmetric difference of V_t and U_t . Using this formula, we can see that the TVL¹ minimization is closely related to the following family of geometric problems: For any $t \in \mathbb{R}$, we denote the following minimization problem $\mathcal{P}_t : \inf_{V \subset \mathbb{R}^2} (\text{Per}(V) + \lambda |V \Delta U_t|)$.

In [21], the following equivalence between \mathcal{P}_t and the minimization of $\mathcal{E}_1(\cdot, u, \lambda)$ was proved:

Theorem 3.2. Suppose $u \in L^1$, the following assertions are equivalent: (i) v is a solution to $\inf_{v' \in BV} \mathcal{E}_1(v', u, \lambda)$. (ii) The upper level set V_t is a solution to \mathcal{P}_t for almost every t .

Having this equivalence, we are ready to prove Theorem 3.1:

Proof 3.1 (Proof of Theorem 3.1). By the contrast invariance property, we can assume $u = \sum_{k=1}^K k \cdot \mathbf{1}_{D_k}$ without loss of generality. Then let V_k be a solution to \mathcal{P}_{k-1} for $k = 1, 2, \dots, K$. That is to say, $V_k = \arg \min_{V \subset \mathbb{R}^2} (\text{Per}(V) + \lambda |U \Delta C_k|)$ where $C_k = \bigcup_{i=k}^K D_i = U_{k-1}$. Let $V'_k = \bigcup_{i=k}^K V_i$ for $k = 1, 2, \dots, K$. By the monotonicity results established in [61], V'_k is also a solution to \mathcal{P}_{k-1} . Define $D'_K = V'_K$ and $D'_k = V'_k \setminus V'_{k+1}$ for $k = 1, 2, \dots, K - 1$. Then let $v = \sum_{k=1}^K k \cdot \mathbf{1}_{D'_k}$. Its level set V_t is V'_k , where $k - 1 \leq t < k$ for any $t \in [0, K]$. By definition, V_t is a solution to \mathcal{P}_t . Hence, by Theorem 3.2, v is a minimizer of $\mathcal{E}_1(\cdot, u, \lambda)$. \square

Unlike the ROF model, the above TVL¹ model was not originally designed for image denoising but the minimization of $\mathcal{E}_1(\cdot, u, \lambda)$ can be regarded as a shape regularization process that depends only on λ and the geometry of D_k , $k = 1, \dots, K$, but not on the intensity levels associated to them. Moreover, the parameter λ corresponds to the scale of the geometry in the following way: as λ decreases, the D'_k will be coarser as small scale objects will merge to form large scale structures.

4. A unified cost functional for EM–TVL¹

In this section, we propose a unified cost functional which naturally combines the GMM-EM algorithm and the TVL¹ regularization. Since the TV regularization is formulated in a variational form, we first rephrase the EM algorithm with a continuous method. This can be achieved following our previous work [32] on denoising. The log-likelihood functional derived from the usually GMM based model would contain a log-sum functional, which makes the related maximum-likelihood estimation (MLE) problem difficult to be directly optimized. In [32], we note that the difficulty of optimizing a log-sum type functional essentially comes from the fact that the logarithm and summation operations are in general non-commutative. To overcome this difficulty, we have introduced a

constraint optimization framework with a convex relaxation method to optimize the log-sum type functional. We have also shown that this constraint optimization method is intrinsically equivalent to the EM algorithm. In a series of studies, we have generalized and have extended the idea into various applications including denoising and illumination correction [32,33].

In this part of the paper, we propose a generalization to incorporate a local geometry constraint for image segmentation. This particular continuous analysis method can be shown to be well-suited for such extending of the GMM. We here first recall some key steps of this constraint optimization method and refer interested readers to [32,33] for a complete discussion.

In continuous case, the negative log-likelihood functional of the GMM model is given by

$$\mathcal{L}(\Theta) = - \int_{\Omega} \log \sum_{k=1}^K \alpha_k p_k(x; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) dx, \quad (9)$$

where $\Theta = (\alpha_1, \dots, \alpha_K, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K)$ is a parameter vector. The existence of the log-sum function makes \mathcal{L} non-quadratic even if each $p_k(x)$ is a Gaussian function. We notice that the log-sum operations are commutative if we apply the Lagrangian multiplier method.

Lemma 4.1 (Commutativity of Log-sum operations, [54,33]). *Given two functions $\alpha_k(x) > 0$, $p_k(x) > 0$, we have*

$$\begin{aligned} & - \log \sum_{k=1}^K \alpha_k(x) p_k(x) \\ &= \min_{\phi(x) \in \Delta_+} \left\{ - \sum_{k=1}^K \log[\alpha_k(x) p_k(x)] \phi_k(x) + \sum_{k=1}^K \phi_k(x) \log \phi_k(x) \right\}, \end{aligned}$$

where $\phi(x) = (\phi_1(x), \phi_2(x), \dots, \phi_K(x))$ is a vector-valued function, and

$$\Delta_+ = \left\{ \phi(x) : 0 < \phi_k(x) < 1, \text{ and } \sum_{k=1}^K \phi_k(x) = 1 \right\}$$

is a convex relaxation of a characteristic function decomposition.

Now, since the logarithm and summation operations can be interchanged, \mathcal{L} becomes a quadratic function and can now be optimized easily. Applying this lemma, we obtain a new functional $\mathcal{H}(\phi, \Theta)$ with an additional variable ϕ

$$\begin{aligned} \mathcal{H}(\phi, \Theta) &= - \int_{\Omega} \sum_{k=1}^K [\log p_k(x) + \log \alpha_k(x)] \phi_k(x) dx \\ &+ \int_{\Omega} \sum_{k=1}^K \phi_k(x) \log \phi_k(x) dx, \end{aligned}$$

and we find a minimizer of \mathcal{L} by minimizing \mathcal{H} instead. The following proposition from [33] shows that both \mathcal{L} and \mathcal{H} have the same global minimizer:

Proposition 4.1. *Both $\mathcal{H}(\phi, \Theta)$ and $\mathcal{L}(\Theta)$ have the same global minimizer Θ^* if $\phi \in \Delta_+$.*

Moreover, if we minimize this new functional using the following alternative minimization scheme

$$\begin{cases} \phi^{(t+1)} &= \arg \min_{\phi \in \Delta_+} \mathcal{H}(\phi, \Theta^{(t)}), \\ \Theta^{(t+1)} &= \arg \min_{\Theta} \mathcal{H}(\phi^{(t+1)}, \Theta), \end{cases} \quad (10)$$

we can also show that minimizing \mathcal{H} will lead to the descent of \mathcal{L} and, therefore, this guarantees the local convergence of the iteration scheme (10) [33].

Proposition 4.2 (Energy Descent). *The sequence $\Theta^{(t)}$ produced by iteration scheme (10) satisfies*

$$\mathcal{L}(\Theta^{(t+1)}) \leq \mathcal{L}(\Theta^{(t)}).$$

Let us point out that the first step in scheme (10) is in fact the E-step in the EM algorithm, and the second step corresponds to the M-step: substituting p_k with a Gaussian function into \mathcal{H} , we have

$$\begin{cases} \phi_k^{(t+1)}(x) &= \frac{\alpha_k^{(t)} p_k(\mathbf{f}(x); \boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Sigma}_k^{(t)})}{\sum_{l=1}^K \alpha_l^{(t)} p_l(\mathbf{f}(x); \boldsymbol{\mu}_l^{(t)}, \boldsymbol{\Sigma}_l^{(t)})}, \\ \alpha_k^{(t+1)} &= \frac{\int_{\Omega} \phi_k^{(t+1)}(x) dx}{\int_{\Omega} 1 dx}, \\ \boldsymbol{\mu}_k^{(t+1)} &= \frac{\int_{\Omega} \phi_k^{(t+1)}(x) \mathbf{f}(x) dx}{\int_{\Omega} \phi_k^{(t+1)}(x) dx}, \\ \boldsymbol{\Sigma}_k^{(t+1)} &= \frac{\int_{\Omega} \phi_k^{(t+1)}(x) (\mathbf{f}(x) - \boldsymbol{\mu}_k^{(t)})^T (\mathbf{f}(x) - \boldsymbol{\mu}_k^{(t)}) dx}{\int_{\Omega} \phi_k^{(t+1)}(x) dx}, \end{cases} \quad (11)$$

which is the same as the iteration scheme of the EM algorithm in Section 2.

In this paper, we propose to extend this approach in [33] by utilizing geometric information such as TV regularization into the EM algorithm. Results from the GMM-EM model can be easily regularized by constraining the function ϕ , and we get the following regularized functional based on the GMM model \mathcal{H} we discussed above

$$\mathcal{E}(\phi, \Theta) = \mathcal{H}(\phi, \Theta) + \lambda \text{TV}(\phi). \quad (12)$$

Note that each ϕ_k in Eq. (11) is actually a smoothed characteristic function which is related to the classification results. To be precise, if each Gaussian function p_k is a delta function (corresponds to $\boldsymbol{\Sigma}_k$ goes to $\mathbf{0}$), then following (11), every ϕ_k would be a characteristic function given by

$$\phi_k(x) = \begin{cases} 1, & \mathbf{f}(x) = \boldsymbol{\mu}_k, \\ 0, & \text{else.} \end{cases}$$

Here $\boldsymbol{\Sigma}_k$ actually controls the width of the interface between 0 and 1. In particular, if \mathbf{f} is a piecewise constant image, then each $\boldsymbol{\Sigma}_k = \mathbf{0}$ and the clustering results ϕ_k would be binary, and it also coincides with k -means segmentation method. Like the well-known level set method, the total variation of ϕ in (11) approximates the length of the boundaries of clusters. Therefore, the geometrical constraints such as the smoothness of clusters boundaries can be realized by imposing a TV regularization on ϕ .

Unfortunately, this energy is difficult to optimize since the linear structures of ϕ is destroyed by the newly added TV term. As a result, we cannot follow the same procedure as in [32] and directly employ the iteration scheme (10) since there is no close-formed solution for each $\phi^{(t+1)}$. To obtain a simple iterative scheme for $\phi^{(t+1)}$, we propose the following splitting technique and the dual method.

We first introduce an auxiliary variable \mathbf{v} which gives the following constraint minimizing problem

$$\min_{\phi, \Theta, \mathbf{v}} \{ \mathcal{H}(\phi, \Theta) + \lambda \text{TV}(\mathbf{v}) \} \text{ s.t. } \mathbf{v} = \phi.$$

Using the penalty technique, we obtain the unconstrained problem

$$\min_{\phi, \Theta, \mathbf{v}} \{ \mathcal{H}(\phi, \Theta) + \lambda \text{TV}(\mathbf{v}) + \eta \|\phi - \mathbf{v}\|_1 \},$$

where $\eta > 0$ is a penalty parameter and $\|\cdot\|_1$ is the L^1 norm. Indeed, one natural choice of the norm in the penalty term is the quadratic penalty $\|\phi - \mathbf{v}\|_2^2$. However, when we minimize the functional with respect to ϕ , such choice will still destroy the close-form solution. Even though the L^1 norm is non-differentiable at 0 so minimizing it directly would not be easy, the dual representation of L^p norm (see e.g. [35]) can help us to overcome the problem.

Lemma 4.2. (Dual representation of L^p norm). Let $\varphi : \mathbb{R}^K \mapsto \mathbb{R}$ is a convex function, $C \subset \mathbb{R}^K$ is a convex set and φ^* is the Fenchel transform of φ , i.e., $\varphi^*(\mathbf{y}) = \sup_{\mathbf{x} \in C} \langle \mathbf{x}, \mathbf{y} \rangle - \varphi(\mathbf{x})$, then $\varphi(\mathbf{x}) = \sup_{\mathbf{y} \in C^*} \langle \mathbf{x}, \mathbf{y} \rangle - \varphi^*(\mathbf{y})$, where C^* is the conjugate set defined by $C^* = \{\mathbf{y} \in \mathbb{R}^K \mid \sup_{\mathbf{x} \in C} \langle \mathbf{x}, \mathbf{y} \rangle - \varphi(\mathbf{x}) < \infty\}$. Specially, if $\varphi(\mathbf{x}) = \|\mathbf{x}\|_p$, then $\|\mathbf{x}\|_p = \sup_{\mathbf{y} \in C^*} \langle \mathbf{x}, \mathbf{y} \rangle$, in which $C^* = \{\mathbf{y} \in \mathbb{R}^K \mid \|\mathbf{y}\|_q \leq 1, \frac{1}{p} + \frac{1}{q} = 1\}$. Here $p = 1$ or 2 and $q = \infty$ if $p = 1$.

By applying Lemma 4.2, we get

$$\begin{aligned} & \min_{\phi, \Theta, \mathbf{v}} \{ \mathcal{H}(\phi, \Theta) + \lambda \text{TV}(\mathbf{v}) + \eta \int_{\Omega} \|\phi(x) - \mathbf{v}(x)\|_p dx \} \\ &= \min_{\phi, \Theta, \mathbf{v}} \{ \mathcal{H}(\phi, \Theta) + \lambda \text{TV}(\mathbf{v}) + \eta \int_{\Omega} \sup_{\|\mathbf{d}(x)\|_q \leq 1} \langle \phi(x) - \mathbf{v}(x), \mathbf{d}(x) \rangle dx \} \\ &= \min_{\phi, \Theta, \mathbf{v}} \sup_{\|\mathbf{d}\|_{\infty} \leq 1} \{ \mathcal{H}(\phi, \Theta) + \lambda \text{TV}(\mathbf{v}) + \eta \int_{\Omega} \langle \phi(x) - \mathbf{v}(x), \mathbf{d}(x) \rangle dx \}, \end{aligned}$$

where $\|\mathbf{d}\|_{\infty} \triangleq \max_x \|\mathbf{d}(x)\|_q$.

Let us denote

$$\mathcal{E}(\phi, \Theta, \mathbf{v}, \mathbf{d}) = \mathcal{H}(\phi, \Theta) + \lambda \text{TV}(\mathbf{v}) + \eta \int_{\Omega} \langle \phi(x) - \mathbf{v}(x), \mathbf{d}(x) \rangle dx.$$

Then the primal problem is to find a saddle point of \mathcal{E} with all constraint conditions $\phi, \mathbf{v} \in \Delta_+$, and $\|\mathbf{d}\|_{\infty} \leq 1$, which can be easily solved by applying the following alternating method.

$$\phi^{(t+1)} = \arg \min_{\phi \in \Delta_+} \mathcal{E}(\phi, \Theta^{(t)}, \mathbf{v}^{(t)}, \mathbf{d}^{(t)}), \tag{13}$$

$$\Theta^{(t+1)} = \arg \min_{\Theta} \mathcal{E}(\phi^{(t+1)}, \Theta, \mathbf{v}^{(t)}, \mathbf{d}^{(t)}), \tag{14}$$

$$\mathbf{v}^{(t+1)} = \arg \min_{\mathbf{v} \in \Delta_+} \mathcal{E}(\phi^{(t+1)}, \Theta^{(t+1)}, \mathbf{v}, \mathbf{d}^{(t)}), \tag{15}$$

$$\mathbf{d}^{(t+1)} = \arg \min_{\|\mathbf{d}\|_{\infty} \leq 1} \mathcal{E}(\phi^{(t+1)}, \Theta^{(t+1)}, \mathbf{v}^{(t+1)}, \mathbf{d}). \tag{16}$$

The subproblems (13)–(16) can be efficiently implemented by some existing algorithms: the solutions of subproblems (13) and (14) are the same as the EM iteration (11) except for

$$\phi_k^{(t+1)}(x) = \frac{\alpha_k^{(t)} p_k(\mathbf{f}(x); \boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Sigma}_k^{(t)}) \exp\{-\eta \mathbf{d}_k^{(t)}(x)\}}{\sum_{l=1}^K \alpha_l^{(t)} p_l(\mathbf{f}(x); \boldsymbol{\mu}_l^{(t)}, \boldsymbol{\Sigma}_l^{(t)}) \exp\{-\eta \mathbf{d}_l^{(t)}(x)\}}. \tag{17}$$

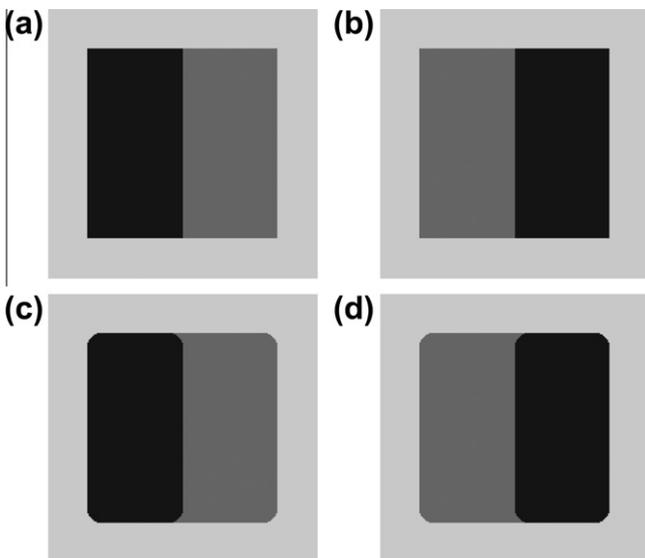


Fig. 1. (a) The image of u , (b) the image of $P(u)$, (c) TVL¹ minimization of u , and (d) TVL¹ minimization of $P(u)$. The TVL¹ regularization is invariant under a permutation of labeling except in a small set of pixels.

As to subproblem (15), it is very similar to the TVL¹ model except that the L^1 norm is now replaced by a linear data term together with a constraint $\mathbf{v} \in \Delta_+$. Such constraint condition is very important since the minimizer of subproblem (15) would be infinity if we remove the condition $0 < v_k(x) < 1$. In practice, we relax $0 < v_k(x) < 1$ to $0 \leq v_k(x) \leq 1$ in order to obtain a close-formed scheme. Note that if $K = 2$, Eq. (15) becomes a two-phase partition and it would be reduced to the global convex segmentation as in [9]. In recent years, many fast and efficient algorithms have developed for TV regularization, such as split Bregman method [24,23], operator splitting [50], augmented Lagrangian method [52], etc. In this part of the paper, we use the split Bregman method to solve for $\mathbf{v}^{(t+1)}$. Let us emphasize that the constraint $\mathbf{v} \in \Delta_+$ can be kept by the Lagrangian multiplier method in the split Bregman iteration scheme. Concerning the last subproblem, since \mathcal{E} is just a linear functional with respect to \mathbf{d} , we can obtain a closed-form expression for solving $\mathbf{d}^{(t+1)}$. By using Holder's inequality, we have

$$\begin{aligned} \int_{\Omega} \langle \phi^{(t+1)}(x) - \mathbf{v}^{(t+1)}(x), \mathbf{d}(x) \rangle dx &\leq \int_{\Omega} \|\phi^{(t+1)}(x) - \mathbf{v}^{(t+1)}(x)\|_p \|\mathbf{d}(x)\|_q dx \\ &\leq \|\mathbf{d}\|_{\infty} \int_{\Omega} \|\phi^{(t+1)}(x) - \mathbf{v}^{(t+1)}(x)\|_p dx \\ &\leq \int_{\Omega} \|\phi^{(t+1)}(x) - \mathbf{v}^{(t+1)}(x)\|_p dx. \end{aligned}$$

Therefore, we conclude that

$$\begin{cases} \mathbf{d}_k^{t+1}(x) = \frac{\phi_k^{(t+1)}(x) - v_k^{(t+1)}(x)}{\mathcal{N}}, & \mathcal{N} \neq 0, \\ \mathbf{d}_k^{t+1}(x) = 0, & \mathcal{N} = 0 \end{cases} \tag{18}$$

is a maximizer of subproblem (16). Here $\mathcal{N} = |\phi_k^{(t+1)}(x) - v_k^{(t+1)}(x)|$ for $p = 1$ and $\mathcal{N} = \|\phi^{(t+1)}(x) - \mathbf{v}^{(t+1)}(x)\|_2$ when $p = 2$.

To end this section, we summarize our algorithm as follows:

Algorithm 4.1. EM-TV algorithm for segmentation

Choose an initial Θ^0 , set $\mathbf{d}^0 = 0$, for $t = 1, 2, \dots$. Repeat the following steps until $\frac{\|\mathbf{v}^{(t+1)} - \mathbf{v}^{(t)}\|_2}{\|\mathbf{v}^{(t)}\|_2} < 10^{-3}$.

1. E-step: find $\phi^{(t+1)}$ using Eq. (17).
2. M-step: update parameter $\Theta^{(t+1)}$ according to the last three equations in (11).
3. TV-step: find $\mathbf{v}^{(t+1)}$ in (15) using split Bregman iteration.
4. Dual-step: calculate the dual variable $\mathbf{d}^{(t+1)}$ by Eq. (18).

5. A splitting algorithm for segmentation

As described in Section 2, since the GMM model is based solely on the distribution of the measurements, the classification u from the typical GMM does not utilize any local geometric information and hence the function u might contain many unwanted fine features. Applying the algorithm we proposed in the previous section, this labeling process is regularized geometrically by imposing the TV term for the function ϕ such that $label(\mathbf{f}(x_i)) = \{k : \phi_k(x_i) = 1\}$. In other words, we can introduce $u : \Omega \rightarrow \{1, \dots, K\}$ such that $u = \sum_{k=1}^K k \cdot \mathbf{1}_{D_k}$ where $D_k = \{x_i \in \Omega : label(\mathbf{f}(x_i)) = k\}$. Note however that the function value of u itself has no other meaning than providing a labeling for each pixel. For instance, we might introduce a permutation $P : \{1, \dots, K\} \rightarrow \{1, \dots, K\}$ and write $u' = P(u) = \sum_{k=1}^K P(k) \cdot \mathbf{1}_{D_k}$. For a given pixel $x_i \in \Omega$, $u(x_i) \neq u'(x_i)$ in general. However, any two pixels x_i and x_j with $u(x_i) = u(x_j)$ implies $u'(x_i) = u'(x_j)$.

Unlike regularizing the measurement vectors in \mathbb{R}^d where different components may have different scales, this generalization

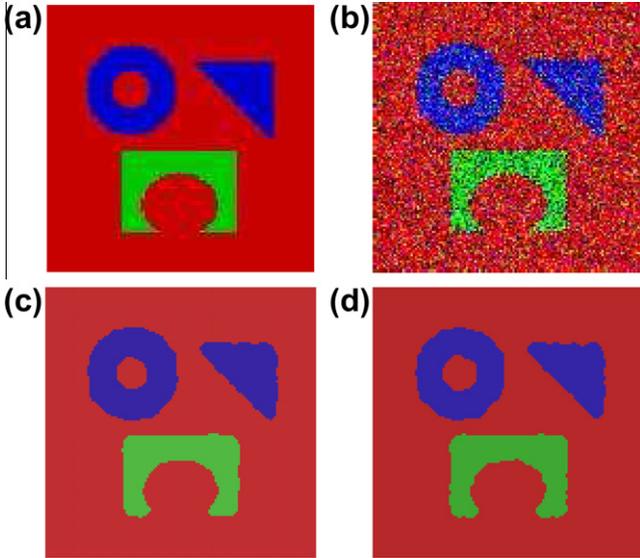


Fig. 2. (a) The clean image, (b) the noisy image (additive Gaussian noise: $\sigma = 77$), (c) the segmentation using Algorithm 4.1, (d) the segmentation using Algorithm 5.1.

of TV from a scale function to a vector ϕ is natural since each component of this \mathbb{R}^K vector is a (smoothed) characteristic function and, therefore, these channels are of the same scale.

In this part, we propose to simplify the above unified function by decoupling the GMM and the TV regularization. We first interpret the minimizer from the function $\mathcal{H}(\phi, \Theta)$ in Eq. (12) as a preliminary segmentation of the image. For the geometry constraint, we propose not to impose the regularity penalty on ϕ (using $\text{TV}(\phi)$) which is a vector in \mathbb{R}^K , but directly on the scalar labeling function $u(x_i) = \text{label}(\mathbf{f}(x_i)) = \{k : \phi_k(x_i) = 1\}$. Since u corresponds to the preliminary classification, this labeling function is simply a scalar function independent of the dimension of the measurement d or the number of classifications K . A similar two-phase approach has been proposed recently in [31] which combines the global statistics from the GMM with neighborhood coherence or local geometric information such as local probability distribution, orientation and anisotropy. However, such algorithm requires computations directly on all measurements and, therefore, the complexity of the resulting algorithm depends on the dimension of the vector-valued image. This could cause computational difficulties when dealing with vector-valued images in extremely high dimensions. There are many well-developed fast algorithms that can be utilized to speed up the TV regularization process for a scalar function, and this simplification gives a computationally more efficient algorithm. In particular, because of these realtime algorithms for TVL¹ algorithms, we can easily regularize the classification from the EM for a wide range of λ s and the user can then pick a suitable solution for any particular application.

Mathematically, we solve the following minimization problem: $\inf_{v \in BV(\Omega)} \mathcal{E}_1(v, u, \lambda)$. By Theorem 3.1, we already see that there exists a minimizer \tilde{u} which is of the same form as u , $\tilde{u} = \sum_{k=1}^K k \cdot \mathbf{1}_{D'_k}$ where D'_1, \dots, D'_K are bounded and disjoint domains in Ω . Therefore, \tilde{u} represents the required image segmentation of the vector-valued image u . Moreover, it can be shown that D'_k is generally smoother than D_k . For example, the following result stems from Remark 1.5.2 in [1]:

Theorem 5.1. Suppose $u = \sum_{k=1}^K k \cdot \mathbf{1}_{D_k} \in BV(\Omega)$ and $\tilde{u} = \sum_{k=1}^K k \cdot \mathbf{1}_{D'_k}$ is a minimizer of $\mathcal{E}_1(\cdot, u, \lambda)$. Then for any positive real number y , the boundary of $U = \{\tilde{u} > y\}$ coincides with the reduced boundary $\partial^* U$ of U in the sense that $H^1(\partial U \setminus \partial^* U) = 0$, and the latter is a $C^{1,1}$ curve.

Applying this theorem, it can be deduced that if D_1, \dots, D_K are sets of finite perimeter, each D'_k has a $C^{1,1}$ boundary outside a set of singular points which has 1-dimensional Hausdorff measure zero. Moreover, it is known that by decreasing the value of the parameter λ , tiny objects in the preliminary segmentation of the image will be destroyed progressively [21,61]. Unwanted details of D_k can be eliminated by choosing a suitable value of λ . Therefore, the choice of λ depends on how much fine scale details one wants to preserve in the segmentation. In practice, such parameter would depend on the type and the variance of noise in the image, and therefore it will be chosen for different applications in a case-by-case manner. These are desirable properties for a good image segmentation.

Here we consider the following question: is TVL¹ minimization of u invariant under a permutation of the labeling of the pixels? That is to say, given any permutation $P : \{1, \dots, K\} \rightarrow \{1, \dots, K\}$, if $\tilde{u} = \sum_{k=1}^K k \cdot \mathbf{1}_{D'_k}$ is a minimizer of $\mathcal{E}_1(\cdot, u, \lambda)$, is $P(\tilde{u})$ a minimizer of $\mathcal{E}_1(\cdot, P(u), \lambda)$? This is true for almost all $x \in \Omega$ except for a small set of pixels depending on the regularization parameter λ . We consider the following simple example as shown in Fig. 1: Let $u = 3 \cdot \mathbf{1}_{\Omega \setminus (D_2 \cup D_3)} + 2 \cdot \mathbf{1}_{D_2} + \mathbf{1}_{D_3}$, where D_2 and D_3 are rectangles that share a common side and P be the transposition of 1 and 2. Then we can see from the figures that the discrepancy between $P(\tilde{u})$ and a minimizer of $\mathcal{E}_1(\cdot, P(u), \lambda)$ occurs only near the common corners of two rectangles. More precisely, $P(\tilde{u})$ and the minimizer of $\mathcal{E}_1(\cdot, P(u), \lambda)$ are equal on the set of all pixels x such that the open ball $B(x, r)$ (centered at x with radius r) contains pixels from at most two classes for some $r > 4/\lambda$ (see [21]). Segmentation result in this particular example might be improved by replacing the TV norm by the anisotropic TV $\int |u_x| + |u_y| d\Omega$. However such norm is not rotational invariant and similar problem will occur for a T-junction making a non-zero angle with the x - or y -axis.

This splitting algorithm is different from the unified algorithm in various ways. The first is that the energy $\text{TV}(\phi)$ in general has a different value than $\text{TV}(u)$ where $u(x_i) = \{k : \phi_k(x_i) = 1\}$. In particular, $\|\phi(x_i) - \phi(x_j)\|_2 = \sqrt{2}$ if two adjacent pixels x_i and x_j are in different classes, while the quantity $|u(x_i) - u(x_j)|$ would depend on the initial labels. If each ϕ_k is a characteristic function of D_k for any partition $\Omega = \bigcup_{k=1}^K D_k$, then L^1 -norm $\text{TV}(\phi) = \int_{\Omega} \sum_{k=1}^K |\nabla \phi_k| dx = \sum_{k=1}^K \text{length}(\partial D_k)$, which is two times of the length of the clusters boundaries. But $\text{TV}(u)$ do not have this property in general case. Another major influence of the approximation is that the global statistics will not be updated according to the local geometry anymore in this algorithm. Unlike the unified algorithm where we minimize both information through one single functional, the decoupling in the reduced algorithm does not allow the geometry to improve the statistical estimates from the GMM. One interesting improvement of this simplified approach is to feedback the local geometry information to the EM classification by following a similar idea as in [31] to iterate between the GMM-EM and the TVL¹ regularization.

Here we end the section by summarizing the algorithm.

Algorithm 5.1. A simplified EM-TV algorithm for segmentation

Choose an initial $\alpha_k^{(0)}, \mu_k^{(0)}, \Sigma_k^{(0)}$.

1. EM-Phase: Iterate (2–3) until convergence.
2. TV-Phase: Let $u = \text{label}(\mathbf{f})$ according to (4). The regularized classification is given by

$$\tilde{u} = \arg \min_v \left\{ \int |\nabla v| dx + \lambda |u - v| dx \right\}.$$

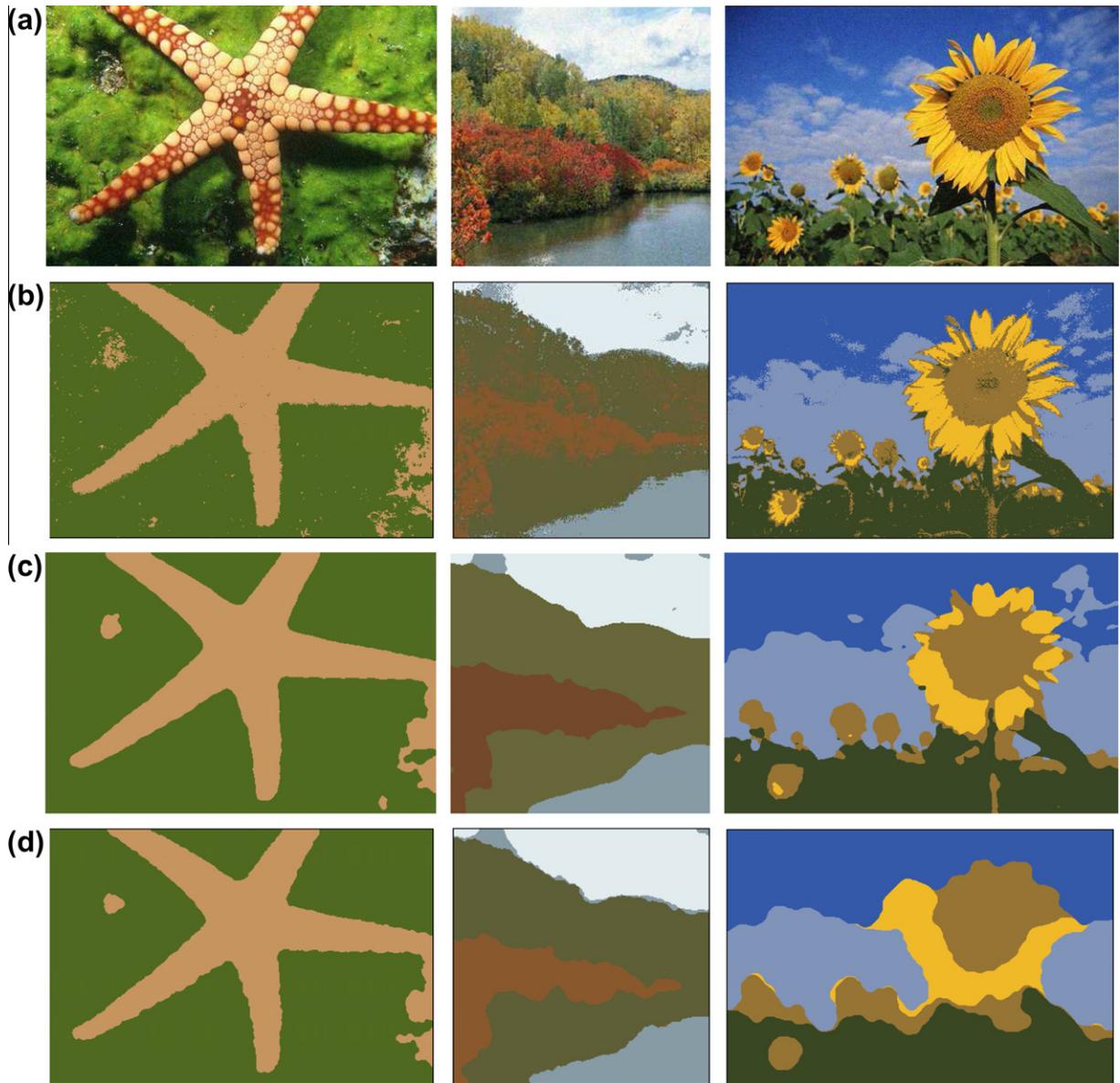


Fig. 3. (a) Original images. (b) Segmentation results by GMM-EM. (c) Segmentation results by Algorithm 4.1. (d) Segmentation results by Algorithm 5.1. In these examples, we choose $K = 2, 4$ and 5 , respectively.

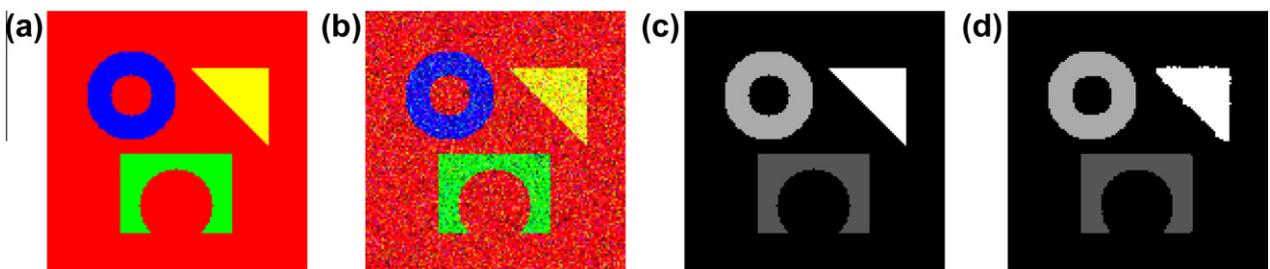


Fig. 4. (a) The clean image, (b) the noisy image (additive Gaussian noise: $\sigma = 70$), (c) the ground truth segmentation, (d) the segmentation results with the proposed method ($K = 4$, Algorithm 4.1).

6. Numerical experiment

The first example is a synthetic color 128-by-128 pixels image consisting of three geometric objects. Gaussian noise of standard deviation $\sigma = 77$ is added to each of the channels, as shown in

Fig. 2(b). In this example, we will demonstrate that our algorithm can eliminate any unnecessary classification class even if we have over-estimated the number of classes K in the EM classification. We use our algorithm to segment the image into four classes (i.e., $K = 4$). Since the segmentation is based on the similarity in

Table 1
Segmentation accuracy index (SAI) of different algorithms under different levels of noise with standard deviation σ . We bold the best SAI number for each testcase.

Noise levels	SAI		
σ	EM	Algorithm 4.1	Algorithm 5.1
50	0.9887	0.9996	0.9989
70	0.8703	0.9967	0.9959
100	0.7330	0.9872	0.9935
150	0.5505	0.9602	0.9606

pixel color, the correct number of segments is actually 3 which implies that $K = 4$ over-estimates the required groups for segmentation. Both the unified functional in Algorithm 4.1 and the reduced approach in Algorithm 5.1 can both eliminate this extra unwanted class in the final classification, as shown in Fig. 2(c) and (d). Fig. 3 shows several more color image examples where $d = 3$. As we can see from these solutions, the unified algorithm and the simplified algorithm do generate different segmentation results. However, both are qualitatively acceptable results.

The second example is to evaluate the segmentation accuracy of the two proposed algorithms. We consider a synthetic image (Fig. 4(a)) corrupted by the Gaussian noise with different standard deviation (σ). To evaluate the performances of different segmentation algorithms, we partition these images using various algorithms and then calculate the following segmentation accuracy index (SAI)

$$SAI = \frac{N_a}{N_t},$$

where N_a is the number of the correctly classified pixels, and N_t stands for the total number of the image. In these experiments, the ground truth segmentation image is given by Fig. 4(c). Since the segmentation results depend on the choice of regularization parameters, thus for each implementation in these experiments, we test on a wide range of parameters values for each algorithm and report only the highest SAI values in Table 1. As can be seen from this table, the segmentation accuracy using the TV regularization is significantly better than that from the simple EM algorithm.

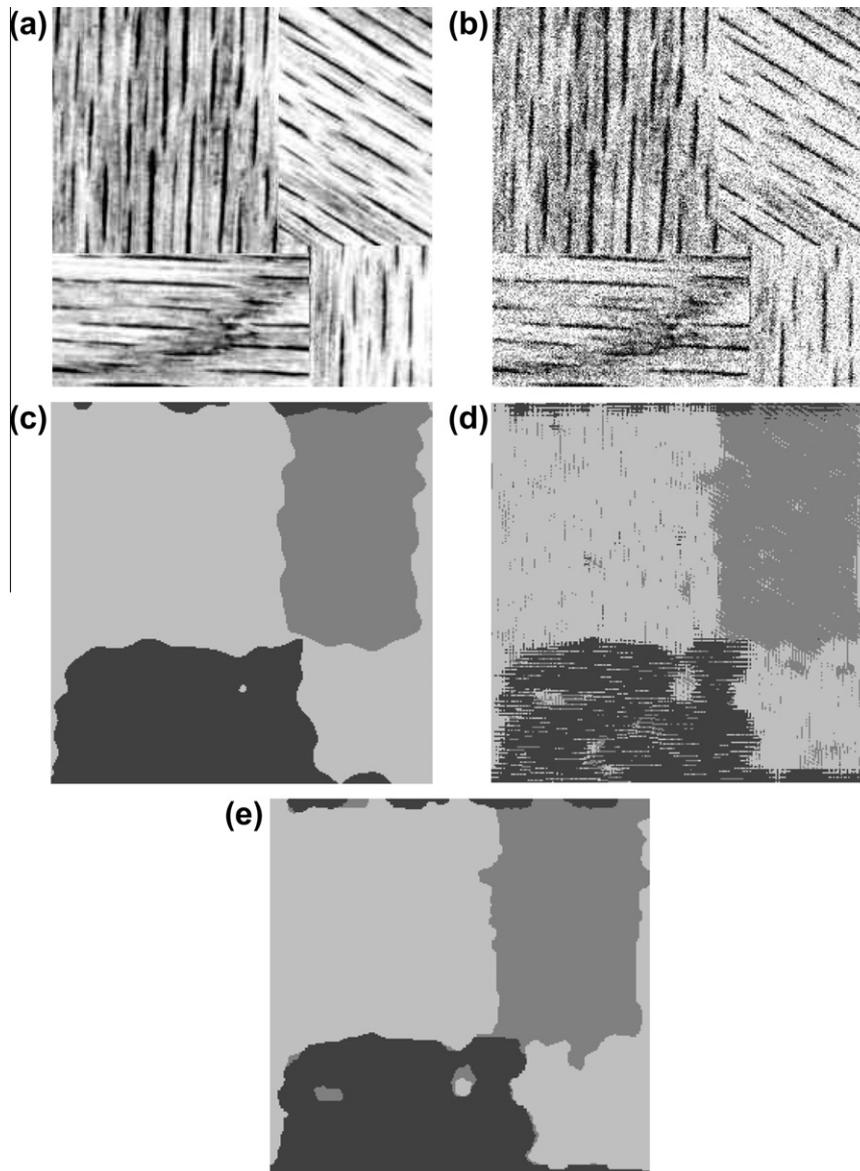


Fig. 5. (a) The clean image, (b) the noisy image (additive Gaussian noise: $\sigma = 51$), (c) the segmentation using Algorithm 4.1, (d) the classification by EM, and (e) the segmentation using Algorithm 5.1.

Compared to the scalar valued function $TV(v)$ in Algorithm 5.1, $TV(\phi)$ in Algorithm 4.1 is a vector-valued TV norm and thus it will be more computational expensive to minimize it. In this experiment, to partition the 128×128 size image into four classes, the average CPU times for Algorithms 4.1 and 5.1 are approximately 0.9 s and 0.7 s, respectively.

The third example is a noisy grayscale texture image of 250-by-250 pixels formed by patching images of three different textures, as shown in Fig. 5(a) and (b). To segment the image according to the texture rather than the intensity level, we first extract the texture features of the image using the Gabor filters. A 2D Gabor filter is a linear filter whose kernel G is a Gaussian function modulated by a sinusoidal plane wave:

$$G(x, y) = e^{-\left(\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)} \cos\left(2\pi \frac{x'}{\xi} + \psi\right) \quad (19)$$

where (x', y') is the clockwise rotation of (x, y) about $(0, 0)$ by angle θ , σ^2 is the variance of the Gaussian function, ξ is the wavelength of

the sinusoidal function, ψ is the phase of the sinusoidal function, and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function. Let $\mathbf{f} = (f^1, \dots, f^d)$. Then for any chosen parameter vector $\mathbf{a} = (\xi, \theta, \gamma, \sigma, \psi)$, we have $f_a^i = f^i * G$ for $i = 1, \dots, d$. Suppose there are m parameter vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$. Then we define $w : \Omega \rightarrow \mathbb{R}^{md}$ by

$$w = (f_{a_1}^1, \dots, f_{a_m}^1, \dots, f_{a_1}^d, \dots, f_{a_m}^d).$$

Therefore, w is a high-dimensional vector-valued image that we will use as the input to our algorithms. Indeed different bank of Gabor filters might lead to different results, we will not concentrate on this issue but will refer interested readers to, for example [26,20,45]. In this example, we choose the parameters $\theta = 0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ$, $\xi = 4, 5, 6$, $\gamma = \sigma = 1$. This implies that we construct an associated vector-valued image $\mathbf{f} : \Omega \rightarrow \mathbb{R}^d$ with $d = 18$. We set $K = 4$ in the first phase of our algorithm and $\lambda = 0.5$ in the second phase of the algorithm. Since the image is formed by combining only three different texture patches,

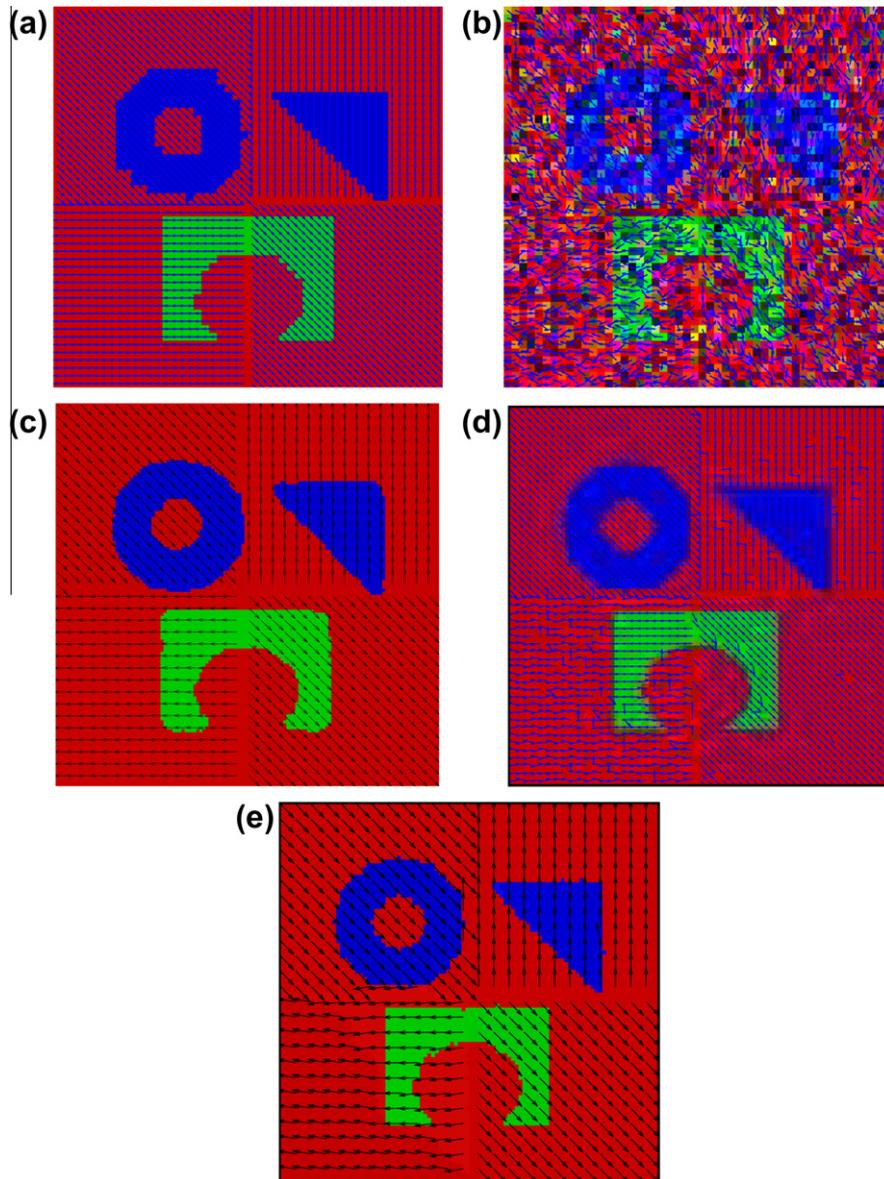


Fig. 6. (a) The clean 4-dimensional image, (b) the noisy image (additive Gaussian noise: $\sigma = 77$ and $3\pi/20$ for color channels and the fourth channel, respectively), (c) the classification by Algorithm 4.1, (d) the classification by GMM-EM, and (e) the segmentation using Algorithm 5.1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$K = 4$ over-estimates the required groups for segmentation. Results from Algorithm 4.1, the usual EM algorithm and also Algorithm 5.1 are shown in Fig. 5(c)–(e), respectively. Pixels corresponding to the same class are plotted in the same intensity color. The image is classified into three classes according to their texture patterns after the TV-L1 regularization and the remaining class contains no pixel.

The last segmentation example is a noisy synthetic four-dimensional image of 128-by-128 pixels $f: \Omega \rightarrow \mathbb{R}^4$ such that $f = (f_1, f_2, f_3, f_4)$ where f_1, f_2 and f_3 are the color channels with intensity level between 0 and 255, and f_4 is the angle of the associated unit vector field such that $0 \leq f_4 \leq 2\pi$. Gaussian noise with the standard deviation of $\sigma = 76.5$ is added to each of those color channels and that with the standard deviation of $3\pi/20$ to the fourth channel. Notice that the noise added to all four channels are of the same scaling. In this example, we show the image by plotting each pixel in color (which takes care of the first three channels) and on each pixel we draw also the unit vector given by $(\cos f_4, \sin f_4)$. To better visualize the data, we sub-sample both the clean and noise images so that only 50-by-50 pixels are shown in Fig. 6(a) and (b), respectively. The solution from Algorithm 4.1 is shown in 6(c). The classification from the EM algorithm, with $K = 16$, is shown in Fig. 6(d). Similar to previous results, since the usual EM algorithm does not have local information in the classification, there are isolated pixels in the segmentation result. The output from the second step of Algorithm 5.1 is shown in Fig. 6(e). Our algorithms classifies the image decently into seven main classes while all other classes are empty.

7. Conclusion

In this paper, we propose two efficient segmentation algorithms for both gray and vector-valued images using the EM algorithm together with the TVL¹ regularization. Experiments show that the combination of the EM algorithm and the TV regularization gives a good segmentation of a variety of vector-valued images. Unlike straight-forward implementation to regularize the measurements in \mathbb{R}^d , we propose in Algorithm 4.1 to impose a TV term on a vector-valued function of length K which acts as a convex relaxation of the characteristic function decomposition. Experimental results have shown that multi-clusters with statistical information and geometrical constraint in segmentation phases can be easily handled. To further lower the dimension of the quantity to be regularized, we propose in Algorithm 5.1 to directly impose the TV constraint on the labeling function. These approaches can be naturally extended to other image processing tasks such as denoising. For example, following the proposed framework, the nonlocal denoising method [11] also can be regarded as an expectation-maximization process and from this view of point, we can derive some new models with better statistical information for a variety of noise. This will be the subject of a forthcoming paper.

Acknowledgments

Liu is partially supported by the National Natural Science Foundation of China (Nos. 11201032 and 11071023). Leung is partially supported by the HKUST Grant RPC11SC06.

References

- [1] W. Allard, Total variation regularization for image denoising: I. Geometric theory, *SIAM J. Math. Anal.* 39 (2007) 1150–1190.
- [2] S. Alliney, Digital filters as absolute norm regularizers, *IEEE Trans. Signal Process.* 40 (1992) 1548–1562.
- [3] S. Alliney, A property of the minimum vectors of a regularization functional defined by means of absolute norm, *IEEE Trans. Signal Process.* 45 (1997) 913–917.
- [4] J.-F. Aujol, S. Kang, Color image decomposition and restoration, *J. Vis. Commun. Image Represent.* 17 (2006) 916–928.
- [5] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
- [6] P. Blomgren, T. Chan, Color TV: total variation methods for restoration of vector valued images, *IEEE Trans. Image Process.* 7 (1998) 304–309.
- [7] G. Box, D. Cox, An analysis of transformations, *J. R. Stat. Soc. B* 26 (1964) 211–246.
- [8] X. Bresson, T. Chan, Fast dual minimization of the vectorial total variation norm and applications to color image processing, *Inverse Prob. Imaging* 2 (2008) 455–484.
- [9] X. Bresson, S. Esedoglu, P. Vanderghyest, J. Thiran, S. Osher, Fast global minimization of the active contour/snake model, *J. Math. Imaging Vis.* 28 (2007) 151–167.
- [10] C. Brune, M. Burger, A. Sawatzky, T. Kösters, F. Wübbeling, Forward-backward EM-TV methods for inverse problems with Poisson noise (Preprint title: An analytical view on EM-TV based methods for inverse problems with Poisson noise), August 2009, Available from: <<http://wwwmath.uni-muenster.de/num/publications/2009/BBSKW09/>>
- [11] A. Buades, B. Coll, J.-M. Morel, On image denoising methods, *SIAM J. Multiscale Model. Simul.* 4 (2) (2005) 490–530.
- [12] C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using expectation-maximization and its application to image querying, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (8) (2002) 1026–1038.
- [13] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, *Int. J. Comput. Vis.* 22 (1) (1997) 61–79.
- [14] T. Chan, S. Esedoglu, Aspects of total variation regularized L^1 function approximation, *SIAM J. Appl. Math.* 65 (2005) 1817–1837.
- [15] T. Chan, B. Sandberg, L. Vese, Active contours without edges for vector-valued images, *JVCIR* 11 (2000) 130–141.
- [16] T. Chan, L. Vese, Active contours without edges, *IEEE Trans. Image Process.* 10 (2001) 266–277.
- [17] O. Christensen, T.-M. Lee, J. Lie, U. Sinha, T. Chan, Total variation regularization of matrix-valued images, *Int. J. Biomed. Imaging* 2007 (2007) 1–11.
- [18] A. Cumani, Edge detection in multispectral images, *CVGIP: Graph. Models Image Process.* 53 (1991) 40–51.
- [19] J.M. Duarte-Carvajalino, G. Yu, L. Carin, G. Sapiro, Online adaptive statistical compressive sensing of Gaussian mixture models, 2011. Available from: <arxiv.org/>.
- [20] D. Dunn, W.E. Higgins, Optimal Gabor filters for texture segmentation, *IEEE Trans. Image Process.* 4 (1995) 947–964.
- [21] V. Duval, J.-F. Aujol, Y. Gousseau, The TVL1 model: a geometric point of view, *Multiscale Model. Simul.* 8 (2009) 154–189.
- [22] V. Duval, J.-F. Aujol, L. Vese, Projected gradient based color image decomposition, in: *Scale Space and Variational Methods in Computer Vision*, vol. 5567, Lecture Notes in Computer Science, Springer, 2009, pp. 295–306.
- [23] T. Goldstein, X. Bresson, S. Osher, Geometric applications of the split Bregman method: segmentation and surface reconstruction, *J. Sci. Comput.* 45 (2009) 272–293.
- [24] T. Goldstein, S. Osher, The split Bregman method for L1 regularized problems, *SIAM J. Imaging Sci.* 2 (2008) 323–343.
- [25] L. Gupta, T. Sortrakul, A Gaussian-mixture-based image segmentation algorithm, *Pattern Recognit.* 31 (3) (1998) 315–325.
- [26] A. Jain, F. Farrokhnia, Unsupervised texture segmentation using Gabor filters, *Pattern Recognit.* 24 (12) (1991) 1167–1186.
- [27] D. Kay, A. Tomasi, Color image segmentation by the vector-valued Allen-Cahn phase-field model: a multigrid solution, *IEEE Trans. Image Process.* 18 (2009) 2330–2339.
- [28] D. Keren, A. Gotlib, Denoising color images using regularization and correlation terms, *J. Vis. Commun. Image Represent.* 9 (1998) 352–365.
- [29] R. Kimmel, R. Malladi, N. Sochen, Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images, *Int. J. Comput. Vis.* 39 (2) (2000) 111–129.
- [30] H.-C. Lee, D. Cok, Detecting boundaries in a vector field, *IEEE Trans. Signal Process.* 39 (1991) 1181–1194.
- [31] S. Leung, G. Liang, K. Solna, H. Zhao, Expectation-maximization algorithm with local adaptivity, *SIAM J. Imaging Sci.* 2 (2008) 834–857.
- [32] J. Liu, X. Tai, H. Huang, Z. Huan, Simultaneous denoising and illumination correction via local data-fidelity and nonlocal regularization, in: *Scale Space and Variational Methods in Computer Vision*, vol. 6667, Lecture Notes in Computer Science, Springer, 2011, pp. 218–230.
- [33] J. Liu, X. Tai, H. Huang, Z. Huan, Weighted norms and sparsity regularization for denoising images corrupted by mixed noise, *IEEE Trans. Image Process.*, 2012, submitted for publication.
- [34] J. Liu, H. Zhang, Image segmentation using a local GMM in a variational framework, *J. Math. Imaging Vis.* (2012), <http://dx.doi.org/10.1007/s10851-012-0376-5>.
- [35] D. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1969.
- [36] B. McLachlan, D. Peel, *Finite Mixture Models*, Wiley, 2000.
- [37] G.J. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, Wiley Series in Probability and Statistics, Wiley-Interscience Publication, 1996.
- [38] D. Mumford, J. Shah, Optimal approximation by piecewise smooth functions and associated variational problems, *Commun. Pure Appl. Math.* 42 (1998) 577–685.
- [39] M. Nikolova, Minimizers of cost-functions involving nonsmooth data-fidelity terms, *SIAM J. Numer. Anal.* 40 (2002) 965–994.
- [40] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, *SIAM J. Multiscale Model. Simul.* 4 (2) (2005) 460–489.

- [41] S. Osher, Y. Mao, B. Dong, W. Yin, Fast linearized Bregman iteration for compressive sensing and sparse denoising, *Commun. Math. Sci.* 8 (2010) 93–111.
- [42] S.J. Osher, J.A. Sethian, Fronts propagating with curvature dependent speed: algorithms based on Hamilton–Jacobi formulations, *J. Comput. Phys.* 79 (1988) 12–49.
- [43] H. Permuter, J. Francos, I. Jermyn, A study of Gaussian mixture models of color and texture features for image classification and segmentation, *Pattern Recognit.* 39 (4) (2006) 695–706.
- [44] C.E. Priebe, Adaptive mixtures, *J. Am. Stat. Assoc.* 89 (1994) 796–806.
- [45] T. Randen, J.H. Husy, Filtering for texture classification: a comparative study, *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (1999) 291–310.
- [46] A. Roussos, P. Maragos, Vector-valued image interpolation by an anisotropic diffusion–projection PDE, in: *Scale Space and Variational Methods in Computer Vision*, Springer, 2007, pp. 104–115.
- [47] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D* 60 (1992) 259–268.
- [48] G. Sapiro, Color snakes, *Comput. Vis. Image Understand.* 68 (1997) 247–253.
- [49] G. Sapiro, D. Ringach, Anisotropic diffusion of multivalued images with applications to color filtering, *IEEE Trans. Image Process.* 5 (1996) 1582–1586.
- [50] S. Setzer, Operator splittings, Bregman methods and frame shrinkage in image processing, *Int. J. Comput. Vis.* 92 (2011) 265–280.
- [51] D. Strong, T. Chan, Edge-preserving and scale-dependent properties of total variation regularization, *Inverse Prob.* 19 (2003) S165–S187.
- [52] X. Tai, C. Wu, Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model, *SSVM* 5567 (2009) 502–513.
- [53] B. Tang, G. Sapiro, V. Caselles, Color image enhancement via chromaticity diffusion, *IEEE Trans. Image Process.* 10 (2001) 701–707.
- [54] M. Teboulle, A unified continuous optimization framework for centerbased clustering methods, *J. Mach. Learn. Res.* 18 (2007) 65–102.
- [55] D. Tschumperle, R. Deriche, Vector-valued image regularization with PDEs: a common framework for different applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (2005) 506–517.
- [56] L. Vese, T. Chan, A multiphase level set framework for image segmentation using the Mumford and Shah model, *Int. J. Comput. Vis.* 50 (3) (2002) 271–293.
- [57] J. Weickert, Coherence-enhancing diffusion of colour images, *Image Vis. Comput.* 17 (1999) 201–212.
- [58] R. Whitaker, G. Gerig, Vector-valued diffusion, in: B. Romeny (Ed.), *Geometry Driven Diffusion in Computer Vision*, Springer, 1994, pp. 93–134.
- [59] M. Yan, General convergent expectation–maximization (EM)-type algorithms for image reconstruction with background emission and Poisson noise. *CAM Report* 11–56, UCLA, 2011.
- [60] W. Yin, Analysis and generalizations of the linearized Bregman method, *SIAM J. Imaging Sci.* 3 (4) (2010) 856–877.
- [61] W. Yin, D. Goldfarb, S. Osher, The total variation regularized L^1 model for multiscale decomposition, *SIAM J. Multiscale Model. Simul.* 6 (2006) 190–211.
- [62] G. Yu, G. Sapiro, Statistical compressive sensing of Gaussian mixture models, in: *Proceedings of IEEE ICASSP*, 2011.
- [63] G. Yu, G. Sapiro, S. Mallat, Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity, *IEEE Trans. Image Process.* 21 (5) (2012) 2481–2499.
- [64] Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden markov random field model and the expectation–maximization algorithm, *IEEE Trans. Med. Imaging* 20 (1) (2001) 45–57.