Noname manuscript No.
(will be inserted by the editor)

Non-local Spatially Varying Finite Mixture Models for Image Segmentation

Javier Juan-Albarracín · Elies Fuster-Garcia · Alfons Juan · Juan M. García-Gómez

Received: date / Accepted: date

Abstract In this work we propose a new Bayesian model for unsupervised image segmentation based on a combination of the Spatially Varying Finite Mixture Models (SVFMMs) and the Non Local Means (NLM) framework. The probabilistic NLM weighting function is successfully integrated into a gauss-markov random field, yielding a prior density that adaptively imposes a local regularization to simultaneously preserve edges and enforce smooth constraints in homogeneous regions of the image. Two versions of our model are proposed: a pixel-based model and a patch-based model, depending on the design of the probabilistic NLM weighting function. Contrary to previous methods proposed in the literature, our approximation does not introduce new parameters to be estimated into the model, because the NLM weighting function is completely known once the neighborhood of a pixel is fixed. The proposed model can be estimated in closed-form solution via a Maximum A Posteriori (MAP) estimation in an expectation-maximization scheme. We have compared our model with previously proposed SVFMMs using two public datasets: the Berkeley Segmentation dataset and the BRATS 2013 dataset. The proposed model performs favorably to previous approaches in the literature, achieving better results in terms of Rand Index and Dice metrics in our experiments.

Keywords Spatially Varying Finite Mixture Models · Non-local Means · Unsupervised learning

Javier Juan-Albarracín
Biomedical Data Science Lab (BDSLab), Instituto Universitario de Tecnologías de la Información y Comunicaciones (ITACA), Universitat Politècnica de València, Valencia, Spain
Tel.: +34 963877000 Ext. 75278
E-mail: jajuaal1@ibime.upv.es

Elies Fuster-Garcia
Biomedical Data Science Lab (BDSLab), Instituto Universitario de Tecnologías de la Información y Comunicaciones (ITACA), Universitat Politècnica de València, Valencia, Spain
Department of Diagnostic Physics, Oslo University Hospital, Oslo, Norway

Alfons Juan
Machine Learning and Language Processing (MLLP) research group, Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València, València, Spain

Juan M. García-Gómez
Biomedical Data Science Lab (BDSLab), Instituto Universitario de Tecnologías de la Información y Comunicaciones (ITACA), Universitat Politècnica de València, Valencia, Spain

1 Introduction

Image segmentation is one of the most important core problems in computer vision, with an extended research history. Unsupervised learning has historically played a key role in the image segmentation task, constituting one of the first paradigms to automatically identify objects and structures in an image (Zhang et al., 2008). Specifically, clustering gathered most of the efforts in unsupervised image segmentation research. Clustering is the task of finding natural groups of data within a population, sharing a similar set of features (Rokach and Maimon, 2005). Many clustering techniques have been proposed in the literature during the past decades (Saxena et al., 2017), ranging from distance based techniques such as partitional clustering or hierarchical clustering; density-based techniques such as DBSCAN (Ester et al., 1996) or Mean Shift (Cheng, 1995); graph based algorithms such as graph-cuts (Boykov et al., 2001); or probabilistic models such as Finite Mixture Models (FMMs)(Pal and Pal, 1993).

Specifically, probabilistic models intend to learn the probability density function (pdf) of an image by means of fitting a multi-parametric statistical model to the data. In particular, FMMs fit of a weighted sum of statistical distributions, each one representing a component of the image, to capture

the heterogeneity nature of the image information. Gaussian Mixture Models (GMMs) are the most extended FMMs, being widely employed for image segmentation, as they have proven to successfully capture the complexity of an image (Juan-Albarracín et al., 2015). Moreover, GMMs can be efficiently estimated by means of maximum likelihood estimation via the Expectation Maximization (EM) algorithm (Dempster et al., 1977).

However, images are structured arrangements of data in which, in addition to the pixel intensities, the location of the pixels provides important information to properly understand its content. Images show patterns of local regularity and spatial redundancy that enclose the idea that adjacent pixels tends to belong to the same component. Conventional FMMs, by the opposite, do not inherently take into account this information. FMMs make a heavy assumption that data in an image is independent and identically distributed (i.i.d.), ignoring local information that may be useful to generate more accurate and realistic results.

To overcome this limitation, several solutions have been proposed in the literature (Blake and Rother, 2011). Most of them rely on the inclusion of a Markov Random Field (MRF) to model the local dependencies between pixels in an image. Specifically, a variant to the FMM called SVFMM was proposed in (Sanjay-Gopal and Hebert, 1998), which replaces the classics mixing coefficients of the FMM by *contextual mixing coefficients* for each pixel of the image. This approximation allows to introduce a MRF over these contextual mixing coefficients to incorporate the idea that neighboring pixels tends to share the same component.

Many variants of MRFs have been proposed in the literature to capture the local information contained in an image. In (Nikou et al., 2007) a family of Gauss-Markov Random Fields (GMRFs) was proposed, successfully producing better results than the classic FMMs. However, such approximation introduces a local isotropic smoothing over the contextual mixing coefficients, that ignores the presence of edges in the image. Therefore, the contextual mixing coefficients estimated under the GMRF approximation are iteratively smoothed, yielding prior probability maps that lose the information of image edges. Sfikas et al (Sfikas et al., 2008) proposed a t-Student MRF that allowed to regulate the smoothing between pixels in an edge. However, this approximation introduces new parameters to be estimated in the model, yielding a non closed-form analytic solution for it.

In this work, we present a fully Bayesian SVFMM model, called Non-Local Spatially Varying Finite Mixture Model (NLSVFMM), that combines the SVFMM framework with the NLM filtering schema. Our proposed model has 2 variants: the pixel-wise version (NLv-SVFMM) and the patch-wise version (NLp-SVFMM). The model introduces a GMRF weighted by the probabilistic NLM function proposed in (Wu et al., 2013) to regulate the smoothing depending on the structure of the image. Such approximation avoids the introduction of new parameters, reducing the degrees of freedom of the model and the number of samples required for a reliable estimation of the parameters. Results obtained show that our method performs favorably to other SVFMM approaches in terms of Rand Index (RI) and Dice metrics.

2 Background on Spatially Variant Finite Mixture Models

The SVFMM is a modification of the classic FMM, mainly oriented to image data, in which the coefficients of the mixture are extended for each pixel of the image.

Let $X = \left(\mathbf{x}^1, \ldots, \mathbf{x}^N\right)$ a set of observations corresponding to the pixels of an image, where $\mathbf{x}^i \in \mathbb{R}^D$ and represents a vector of D features for the i^{th} pixel. The SVFMM is defined as:

$$p\left(X|\Theta\right) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_j^i \phi\left(\mathbf{x}^i|\theta_j\right)$$

where $\phi\left(\mathbf{x}^i|\theta_j\right)$ is a pdf used to model the data, $\Pi = \left\{\pi^1, \ldots, \pi^N\right\}$ are called the *contextual* mixing coefficients, with

$$\forall \pi^i,\ \ 0 \leq \pi_j^i \leq 1,\ \ \sum_{j=1}^{K} \pi_j^i = 1$$

and $\Theta = \left\{\theta_1, \ldots, \theta_K, \pi^1, \ldots, \pi^N\right\}$ is the set of parameters of the model.

A Maximum A Posteriori (MAP) estimate of Θ is employed to impose a proper prior over Π to introduce the idea that neighboring pixels in an image tend to belong to the same component.

$$\hat{\Theta}_{MAP} = \arg\max_{\Theta} \log p\left(X|\Theta\right) + \log p\left(\Pi\right)$$

Several variants of $p\left(\Pi\right)$ has been previously proposed in the literature. Specifically, the directional class-adaptive GMRF, which we have used as the basis of our approach, takes the form:

$$p\left(\Pi\right) = \prod_{i=1}^{N} \prod_{j=1}^{K} \prod_{d=1}^{D} \prod_{m \in \mathscr{M}_d^i} \frac{1}{\sqrt{2\pi\beta_{j,d}^2}} \exp\left(-\frac{\left(\pi_j^i - \pi_j^m\right)^2}{2\beta_{j,d}^2}\right)$$

where $\beta_{j,d}^2$ is the variance of the corresponding gaussian instance and $\mathscr{M}_d^i$ is the set of neighbors of the i^{th} observation that lies in the d direction. Typically, 4 directions can be considered in 2-D images, which correspond to the horizontal, vertical and diagonals.

In short, the directional class-adaptive GMRF encodes the idea that differences between adjacent mixing coefficients are Gaussian distributed in the form

$$\pi_j^i - \pi_j^m \sim \mathcal{N}\left(0, \beta_{j,d}^2\right)$$

Inference on this model is not analytically tractable so numerical optimization methods must be employed. Thus, a MAP-EM algorithm is typically used to iteratively estimate the parameters of the model. The corresponding Q function for the EM schema is:

$$Q\left(\Theta|\Theta^{(t)}\right) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_j^i \left(\log \pi_j^i + \log \phi\left(\mathbf{x}^i|\theta_j\right)\right) + \log p\left(\Pi\right)$$

The posterior density of the hidden variables $Z = \left(\mathbf{z}^1, \ldots, \mathbf{z}^N\right)$, associated to each component, is computed at the E-step as follows

$$z_j^{i\,(t)} = \frac{\pi_j^{i\,(t)} \phi\left(\mathbf{x}^i|\theta_j{}^{(t)}\right)}{\sum\limits_{k=1}^{K} \pi_k^{i\,(t)} \phi\left(\mathbf{x}^i|\theta_k{}^{(t)}\right)}$$

while at the M-step the parameters of the model Θ, when $\phi\left(\mathbf{x}^i|\theta_j\right) \sim \mathcal{N}\left(\mathbf{x}^i|\mu_j, \Sigma_j\right)$, are estimated by

$$\mu_j^{(t+1)} = \frac{1}{\sum\limits_{i=1}^{N} z_j^{i\,(t)}} \sum_{i=1}^{N} z_j^{i\,(t)} \mathbf{x}^i$$

$$\Sigma_j^{(t+1)} = \frac{1}{\sum\limits_{i=1}^{N} z_j^{i\,(t)}} \sum_{i=1}^{N} z_j^{i\,(t)} \left(\mathbf{x}^i - \mu_j^{(t+1)}\right)\left(\mathbf{x}^i - \mu_j^{(t+1)}\right)^T$$

The updates for contextual mixing coefficients $\pi_j^{i(t+1)}$ are obtained as the roots of the following second degree equation

$$\left(\pi_j^{i(t+1)}\right)^2 \sum_{d=1}^{D} \frac{\left|\mathscr{M}_d^i\right|}{\beta_{j,d}^2} - \left(\pi_j^{i(t+1)}\right) \sum_{d=1}^{D} \frac{\sum\limits_{m\in\mathscr{M}_d^i} \pi_j^m}{\beta_{j,d}^2} - \frac{z_j^i}{2} = 0$$

which always have a real non-negative solution. However, a limitation of this estimation is that it does not take into account the constraint that $\sum_j \pi_j^i = 1, \ \forall i$. Instead, *reparatory* techniques such as the one proposed in (Blekas et al., 2005) must be employed to ensure the probabilities to sum 1.

An interesting alternative to avoid reparatory projections is to consider that Π is governed by a Dirichlet Compound Multinomial (DCM) distribution. This means that the hidden random variable $\mathbf{z}^i$ is governed by a multinomial distribution with parameters π^i, which in turn is governed by a Dirichlet distribution with parameters α^i. In (Nikou et al., 2010) Nikou *et al* demonstrated that, following such hierarchical model, π_j^i can be computed as

$$\pi_j^i = \frac{\alpha_j^i}{\sum\limits_{k=1}^{K} \alpha_k^i}$$

yielding a fully Bayesian model that always guarantees that $\sum_j \pi_j^i = 1 \ \ \forall i$.

The parameters of the Dirichlet distribution only require to satisfy that $\alpha_j^i > 0 \ \ \forall i, j$, making easier its optimization. Thus, a directional class-adaptive GMRF density can be imposed over $A = \left\{\alpha^1, \ldots, \alpha^N\right\}$ to enforce local regularity

$$p\left(A\right) = \prod_{i=1}^{N}\prod_{j=1}^{K}\prod_{d=1}^{D}\prod_{m\in\mathscr{M}_d^i} \frac{1}{\sqrt{2\pi\beta_{j,d}^2}} \exp\left(-\frac{\left(\alpha_j^i - \alpha_j^m\right)^2}{2\beta_{j,d}^2}\right)$$

Hence, the Q function associated to the DCM-SVFMM becomes

$$Q\left(\Theta|\Theta^{(t)}\right) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_j^i \left(\log \frac{\alpha_j^i}{\sum\limits_{k=1}^{K}\alpha_k^i} + \log \phi\left(\mathbf{x}^i|\theta_j\right)\right) + \log p\left(A\right)$$

Optimizing the corresponding Q function for this model yields identically updates for $\mu_j^{(t+1)}$ and $\Sigma_j^{(t+1)}$, but setting $\partial Q/\partial \alpha_j^i$ yields a third degree equation of the form

$$\left(\alpha_j^{i(t+1)}\right)^3 + \left(\alpha_j^{i(t+1)}\right)^2 \left(A_{-j}^i - \frac{C_j^i}{B_j^i}\right) - \left(\alpha_j^{i(t+1)}\right)\left(\frac{A_{-j}^i C_j^i}{B_j^i}\right) - \frac{z_j^i A_{-j}^i}{2B_j^i} = 0$$

where

$$A_{-j}^i = \sum_{\substack{k=1 \\ k\neq j}}^{K} \alpha_k^i$$

$$B_j^i = \sum_{d=1}^{D} \frac{\left|\mathscr{M}_d^i\right|}{\beta_{j,d}^2}$$

$$C_j^i = \sum_{d=1}^{D} \frac{\sum\limits_{m\in\mathscr{M}_d^i} \alpha_j^m}{\beta_{j,d}^2}$$

Finally, $\beta_{j,d}^2$ is estimated by

$$\beta_{j,d}^2 = \frac{1}{N}\sum_{i=1}^{N} \sum_{m\in\mathscr{M}_d^i} \frac{\left(\alpha_j^i - \alpha_j^m\right)^2}{\left|\mathscr{M}_d^i\right|}$$

3 Background on Probabilistic Non Local Means

The NLM filter (Buades et al., 2005) proposes a schema for image filtering where pixels are restored by a weighted sum of similar neighbor patches.

The core of NLM schema is the computation of the weight function between patches, which has taken a lot of variants in the literature. Specially, Wu et al. (Wu et al., 2013) derived the probabilistic version of the NLM algorithm and its associated probabilistic weighting function.

Linking the description of the probabilistic NLM with the SVFMM background, let's consider $d_{j,d}^{i,m}$ the distance between a pair of adjacent Dirichlet parameters

$$d_{j,d}^{i,m} = \frac{\left(\alpha_j^i - \alpha_j^m\right)^2}{2\beta_{j,d}^2}$$

Assuming that local differences are i.i.d., we have $d_{j,d}^{i,m} \sim \chi_1^2$. For a patch-based version of the algorithm, the distance between two patches centered at i^{th} and m^{th} locations is defined as

$$D_{j,d}^{i,m} = \sum_{k\in\mathbb{P}} d_{j,d}^{i+k,m+k}$$

where $\mathbb{P}$ is the set of offsets that define a local patch around a given pixel. If patches are completely disjoint, then $D_{j,d}^{i,m} \sim \chi_{|\mathbb{P}|}^2$, however, in most cases, overlapping occurs between patches, so the i.i.d. assumption does not hold. In such cases, an approximation to the sum of a set of correlated χ^2 distributions can be computed as

$$D_{j,d}^{i,m} \sim \gamma_m \chi_{\eta_m}^2$$

where

$$\gamma_m = \mathrm{var}\left[D_{j,d}^{i,m}\right] / 2\mathrm{E}\left[D_{j,d}^{i,m}\right]$$
$$\eta_m = \mathrm{E}\left[D_{j,d}^{i,m}\right] / \gamma_m$$

and

$$\mathrm{E}\left[D_{j,d}^{i,m}\right] = |\mathbb{P}|$$
$$\mathrm{var}\left[D_{j,d}^{i,m}\right] = 2|\mathbb{P}| + \left|\mathbb{O}^{i,m}\right|$$

with $\mathbb{O}^{i,m}$ the set of overlapping pixels between the patches centered at i^{th} and m^{th} pixels.

Hence, the weight function $u_{j,d}^{i,m}$ proposed in the probabilistic NLM approach is defined as

$$u_{j,d}^{i,m} = \chi_{\eta_m}^2\left(D_{j,d}^{i,m}/\gamma_m\right) = \frac{\left(D_{j,d}^{i,m}/\gamma_m\right)^{(\eta_m/2)-1} \exp\left(-D_{j,d}^{i,m}/2\gamma_m\right)}{2^{\eta_m/2}\Gamma\left(\eta_m/2\right)}$$

4 The Non Local Spatially Variant Finite Mixture Model

One of the main drawbacks of the previous aforementioned SVFMM is that this model enforces local smoothness on the contextual mixing coefficients, without taking into account the structure of the image. In other words, the SVFMM iteratively applies an isotropic local Gaussian smoothing to the contextual mixing coefficients, which finally yields into a over-smoothed prior probability map that losses the information of edges and structure in the image.

To overcome this limitation, Nikou et al (Nikou et al., 2007) proposed a variant of the SVFMM where local differences between Dirichlet parameters follow a t-Student distribution. Such an approach was intended to exploit the heavy-tailed nature of the t-Student distribution, to perform a robust estimation of the Dirichlet coefficients when edges and structures are present in their local neighborhoods.

$$\alpha_j^i - \alpha_j^m \sim \mathscr{S}t\left(0, \beta_{j,d}^2, \nu_j\right)$$

Following the Bishop's development in (Bishop, 2006), a $\mathscr{S}t$ distribution can be expressed as

$$\alpha_j^i - \alpha_j^m \sim \mathscr{N}\left(0, \beta_{j,d}^2/g_{j,d}^{i,m}\right)$$
$$g_{j,d}^{i,m} \sim \mathscr{G}\left(\nu_{j,d}/2, \nu_{j,d}/2\right)$$

This model introduces a new set of latent variables $g_{j,d}^{i,m}$, whose posterior density should be estimated at the E-step, and an additionally new set of parameters $\nu_{j,d}$, which yield a non closed-form analytic solution of the model. Numerical optimization methods should be employed to estimate $\nu_{j,d}$.

In this sense, we propose the NLSVFMM as a modification of this model by replacing the $g_{j,d}^{i,m}$ random variable by the probabilistic NLM $u_{j,d}^{i,m}$ weight. Therefore, we propose to reformulate the local differences between contextual Dirichlet parameters in the form

$$\alpha_j^i - \alpha_j^m \sim \mathscr{N}\left(0, \beta_{j,d}^2/\chi_{\eta_m}^2\left(D_{j,d}^{i,m}/\gamma_m\right)\right)$$
$$D_{j,d}^{i,m} \sim \chi^2\left(\eta_m\right)$$

with $D_{j,d}^{i,m}$ being latent variables of the model.

Following the conventional EM scheme, the posterior densities of $D_{j,d}^{i,m}$ should be calculated at the E-step. However, this leads to a different calculation of $D_{j,d}^{i,m}$ than the proposed by Wu et. al (see Section 3). Therefore, in order to preserve the use of the original NLM weights, we will follow a *Variational EM* approach (Neal and Hinton, 1999; Bishop, 2006). The Variational EM framework introduces the concept of partial E-step, in which a functor of the latent variables can be used when the posterior densities of

these variables cannot be calculated, or when it is desirable to calculate them differently for reasons of efficiency or performance. As demonstrated by Neal and Hinton (1999), such functor can take any form as long as the log-likelihood function is increased at each iteration, effectively driving the model to a local optimum of the function, and hence to an optimum of the parameters of the model. Therefore, following this framework, the $D_{j,d}^{i,m}$ latent variables are estimated at the E-step as the standard quantitative Chi-squared test proposed by Wu et. al:

$$D_{j,d}^{i,m} = \sum_{k\in\mathbb{P}} \frac{\left(\alpha_j^{i+k} - \alpha_j^{m+k}\right)^2}{2\beta_{j,d}^2}$$

Once these latent variables are estimated, the $u_{j,d}^{i,m}$ weights are calculated at the M-step following $u_{j,d}^{i,m} = \chi^2_{\eta_m}\left(D_{j,d}^{i,m}/\gamma_m\right)$. Since $u_{j,d}^{i,m}$ depends on both the i^{th} and m^{th} observations, this model specifies a different instance of a Gaussian distribution for each $\left(\alpha_j^i - \alpha_j^m\right)$ pair of contextual Dirichlet coefficients in the MRF. This allows $u_{j,d}^{i,m}$ to regulate the variance of the corresponding Gaussian between the i^{th} and m^{th} observations, if an edge or an homogeneous area is detected at this location. Thus, as $u_{j,d}^{i,m}$ increases, the Gaussian distribution for the corresponding pair shrinks around zero imposing a hard smoothing between the observations. On the contrary, as $u_{j,d}^{i,m}$ decreases, the variance of the Gaussian distributions increases producing a lower pdf value that prevents the smooth.

This approximation avoids the introduction of new parameters since η_m and γ_m are completely known once i and m are fixed. Therefore, no numerical approximate methods are required, simplifying the model and reducing its degrees of freedom and the number of samples required for its statistically reliable estimation.

The graphical model of the NL-SVFMM is shown in Figure 1.

Imposing the directional class-adaptive GMRF to this model, the new density for $p(A)$ becomes

$$p(A) = \prod_{i=1}^{N}\prod_{j=1}^{K}\prod_{d=1}^{D}\prod_{m\in\mathscr{M}_d^i} \frac{1}{\sqrt{2\pi\beta_{j,d}^2/u_{j,d}^{i,m}}} \exp\left(-\frac{\left(\alpha^i-\alpha_j^m\right)^2 u_{j,d}^{i,m}}{2\beta_{j,d}^2}\right)$$

which setting $\partial Q/\partial\alpha_j^i$ yields a new third degree equation of the form

$$\left(\alpha_j^{i(t+1)}\right)^3 + \left(\alpha_j^{i(t+1)}\right)^2\left(A^i_{-j} - \frac{\widehat{C}_j^i}{\widehat{B}_j^i}\right) - \left(\alpha_j^{i(t+1)}\right)\left(\frac{A^i_{-j}\widehat{C}_j^i}{\widehat{B}_j^i}\right) - \frac{z_j^i A^i_{-j}}{2\widehat{B}_j^i} = 0$$

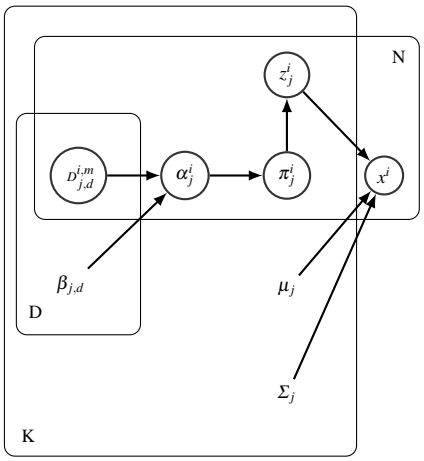

Fig. 1 Graphical model for the non-local spatially varying finite mixture model. Superscripts $i,m \in [1,N]$ denote pixel indexes, subscript $j \in [1,K]$ denotes mixture component and subscript $d \in [1,D]$ denotes neighborhood direction.

where

$$\widehat{B}_j^i = \sum_{d=1}^{D} \frac{\sum_{m\in\mathscr{M}_d^i} u_{j,d}^{i,m}}{\beta_{j,d}^2}$$

$$\widehat{C}_j^i = \sum_{d=1}^{D} \frac{\sum_{m\in\mathscr{M}_d^i} \alpha_j^m u_{j,d}^{i,m}}{\beta_{j,d}^2}$$

Finally, $\beta_{j,d}^2$ is now estimated as

$$\beta_{j,d}^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{m\in\mathscr{M}_d^i} \frac{\left(\alpha_j^i - \alpha_j^m\right)^2 u_{j,d}^{i,m}}{\left|\mathscr{M}_d^i\right|}$$

Hereafter, the pixel-wise χ_1^2 version of the proposed NLSVFMM will be referred as NLv-SVFMM, while the patch-wise $\chi^2_{\eta_m}$ will be referred as NLp-SVFMM.

5 Experimental results

First, we have performed an evaluation of our algorithm with the Brain Tumor Segmentation (BRATS) 2013 high grade glioma synthetic dataset, which includes 25 cases segmented into 7 classes: 1) white matter (WM), 2) grey matter (GM), 3) cerebro-spinal fluid (CSF), 4) peripheral edema (ED), 5) tumor core (split into enhancing tumor (5.1) and necrotic core (5.2)) (TC) and 6) vessels (VS). For each 3-D voxel, intensities on pre- and post-gadolinium T1-weighted MRI, T2-weighted MRI and FLAIR sequence were employed for the segmentation.

Later, we have evaluated our proposed model with the 300 real-world images of the Berkeley Segmentation Dataset.

In our experimentation, we employed a 3-dimensional feature vector for each pixel, comprising the 3 channels of the L*a*b color space. We also applied a local median smoothing to each channel using a 5×5 window centered at each pixel. We have evaluated the performance of each algorithm for different values of $K = \{3,5,7,10,15,20\}$.

We have compared our proposed NLv- and NLp-SVFMM model with the conventional FMM, the SVFMM and the $\mathscr{S}$t-SVFMM. For the spatially varying algorithms we have employed the DCM Bayesian approximation and the directional class-adaptive GMRF prior specified in 4. All algorithms in all experiments were initialized with a deterministic version of K-means++ (Arthur and Vassilvitskii, 2007) to ensure a fair comparative.

Figure 2 compares the behavior of the weighting functions $G = \left\{ g_{j,d}^{i,m} \right\}$ for the $\mathscr{S}$t-SVFMM model and $U = \left\{ u_{j,d}^{i,m} \right\}$ for the NLv-SVFMM model (NLp-SVFMM weighting function is not depicted because is not comparable to the $\mathscr{S}$t and NLv-SVFMM models). As figure shows, U function behaves more aggressive for differences between observations than the $\mathscr{S}$t-SVFMM, hence yielding more dichotomous weighting maps (see Figure 3). For the shake of simplicity, each pixel of each picture of Figure 3 represent $\sum_{d=1}^{D} \sum_{m \in \mathscr{M}^i} \lambda_{j,d}^{i,m}$, with $\lambda = g$ for $\mathscr{S}$t-SVFMM and $\lambda = u$ for NLv- or NLp-SVFMM models respectively.

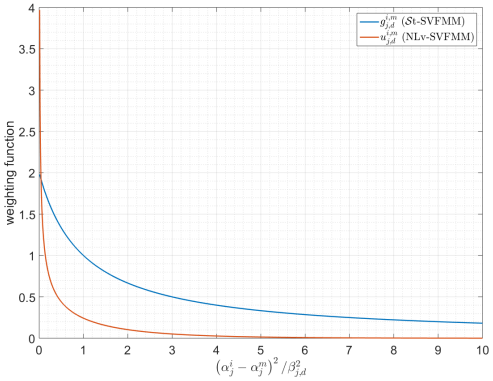

Fig. 2 Comparison between the behavior of the weighting functions $G = \left\{ g_{j,d}^{i,m} \right\}$ for the $\mathscr{S}$t-SVFMM model and $U = \left\{ u_{j,d}^{i,m} \right\}$ for the NLv-SVFMM (NLp-SVFMM weighting function is not depicted because is not numerically comparable to the $\mathscr{S}$t and NLv-SVFMM functions).

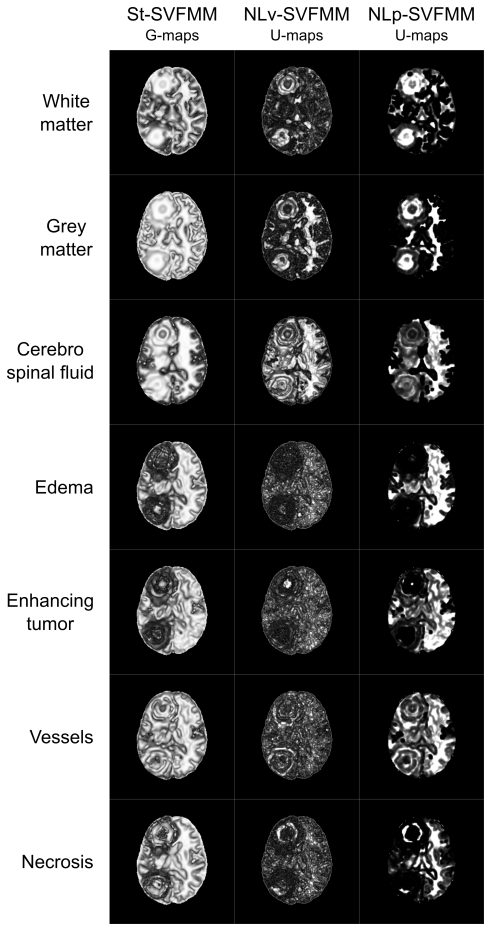

Fig. 3 Comparison between G maps of the $\mathscr{S}$t-SVFMM and U maps for the NLv- and NLp-SVFMM models for a case of the BRATS 2013 dataset. Each pixel i of the images represent $\sum_{d=1}^{D} \sum_{m \in \mathscr{M}^i} \lambda_{j,d}^{i,m}$, with $\lambda = g$ or $\lambda = u$ for $\mathscr{S}$t-SVFMM and NLv- or NLp-SVFMM models respectively.

Table 1 Contextual mixing coefficients for different voxels of the HG0014 case of the BRATS 2013 challenge. Voxels correspond to coordinates $a = (151,127,85)$, $b = (167,75,85)$, $c = (151,152,85)$, $d = (97,89,85)$, $e = (117,62,85)$, $f = (110,71,85)$ and $f = (128,99,85)$

	SVFMM	$\mathscr{S}$t-SVFMM	NLv-SVFMM	NLp-SVFMM
π_{WM}^a	0.457	0.486	0.481	**0.499**
π_{GM}^b	0.296	0.429	0.497	**0.606**
π_{CSF}^c	0.253	0.426	0.446	**0.447**
π_{ED}^d	0.294	0.374	0.394	**0.411**
π_{ET}^e	0.248	0.442	**0.453**	0.381
π_{NC}^f	0.267	0.291	0.285	**0.361**
π_{VS}^g	0.118	0.305	**0.355**	0.265

Table 1 and Figure 4 show the superiority of the proposed NL-SVFMM (in both variants) to generate higher confidence prior probability maps for each component. An example of the contextual mixing coefficient maps for the HG0014 case of the BRATS 2013 dataset and its associated mixing coefficient values for different pixels obtained by each method is shown. In almost all evaluations, the NLp-SVFMM version achieves the best results, indicating that the patch-based probabilistic NLM weighting function better captures the local similarities in the images.

Table 2 shows the Dice coefficients obtained for the evaluation based on the BRATS 2013 dataset. Consistently with previous results, the NLp-SVFMM variant achieves the best results in terms of segmentations based on the maximization of the posterior probabilities (Bayes minimum classification error). An improvement of about 3 points in Dice is

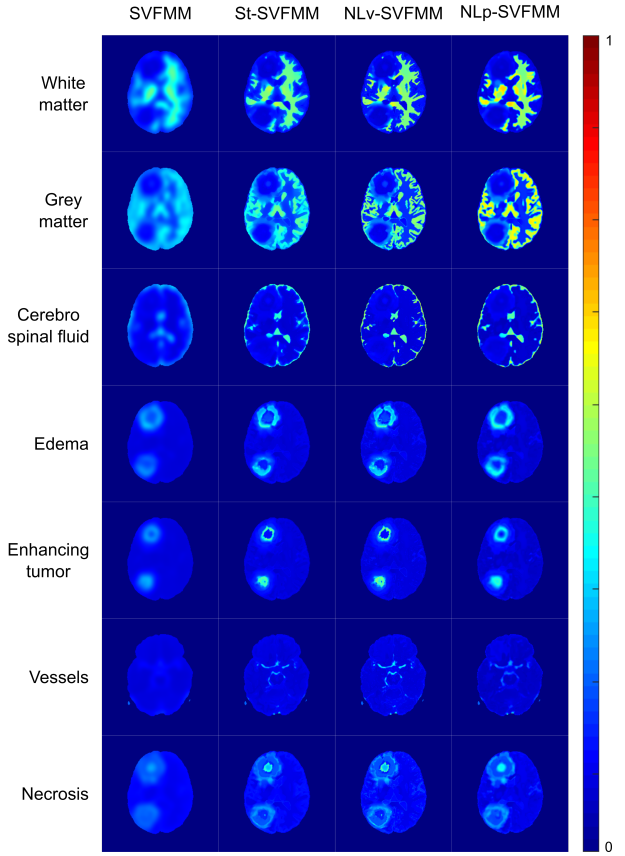

Fig. 4 Example of contextual mixing coefficient maps for a case of BRATS 2013, for each class of the segmentation.

obtained when comparing the NLp-SVFMM with the standard SVFMM and more than 1 point in Dice with respect to the $\mathscr{S}$t-SVFMM, thanks to the proposed prior density. Moreover, in order to explore the capabilities of the proposed prior densities to yield accurate segmentations, we have also computed the Dice coefficients for the segmentations based only on the maximization of the prior probability maps generated by each method. As Table 2 shows, the NLv-SVFMM method, followed by the NLp-SVFMM, achieves the best results. Of course the Dice coefficients are significantly low because the segmentations do not take into account the pixel intensities. However, the aim is to evaluate which method better captures the local similarities in the images, hence producing more accurate prior information about the image.

Table 3 shows the results for the evaluation of the 300 images of the Berkeley dataset. RI is employed to measure the degree of concordance between the automated segmentation and the manual segmentations. Our experiments show that the proposed methods are superior in terms of RI to the other approaches in almost all situations. The NLv-SVFMM performs comparable to the $\mathscr{S}$t variant in most of cases, achieving very similar results. However, the NLv-SVFMM requires less parameters, hence alleviating the computational cost to obtain the segmentations. Nevertheless, the NLp-SVFMM method achieves, both in average and median cases, the best results. Only in the $K = 3$ case (the simplest segmentation), the SVFMM method outperforms the rest of the models. However, as segmentation complexity increases, the models including edge preserving priors performs better in all cases.

It is worth noting that differences in RI are not significantly large between methods. This may be due because prior probabilities become weaker when the number of observations increase, which is the case of pixel classification in an image. As expected, small differences between methods demonstrate that prior densities have a limited effect on the final segmentation when following a Bayesian decision rule.

Additionally, a comparison in terms of the computational time required by each method has been performed. Table 4 shows the average times (in seconds) and the std. deviation of each method evaluated in the Berkeley 300 dataset for different number of segments calculated in the images.

As expected, the SVFMM is the fastest method since it doesn't carry the extra computation of the weights for constrain the $\beta_{j,d}^2$ variances. It should be noted that only the NLv-SVFMM and the $\mathscr{S}$t-SVFMM are directly comparable since both perform the calculation of the u and g weights respectively, and those weights are computed pixelwise. It can be seen that both methods perform very similar, with no significant difference between them. Although the $\mathscr{S}$t-SVFMM model requires a numerical iterative approximation of the $\nu_{j,d}$ parameters, which is often a slow procedure, the complexity in the computation of the $g_{j,d}^{i,m}$ weights is lighter than the $u_{j,d}^{i,m}$ weights. That is the reason why the NLv-SVFMM is a bit slower than the $\mathscr{S}$t-SVFMM. The calculation of $u_{j,d}^{i,m}$ weights requires the computation of $NKD|\mathscr{M}_d|$ $\chi_{\eta_m}^2$ pdf values, which finally equals or even slightly increases the computational time with respect the g weights. The NLp-SVFMM performs the best in terms of Dice and RI scores, but also requires more time to compute the segmentation since it carries the extra computation of the patch-based similarity.

Finally, Figure 5 shows several examples of segmentations of images of the Berkeley dataset obtained with the NLp-SVFMM method.

6 Conclusion

In this study we have proposed a new unsupervised image clustering algorithm that successfully merges the SVFMM framework with the well-known NLM filtering scheme. The main advantage of this algorithm is the proposed new MRF density over the contextual mixing proportions, which enforces local smoothness while preserving edges and the struc-

Table 2 Results on the DICE coefficient over the 25 synthetic high-grade gliomas of the BRATS 2013 dataset for each algorithm evaluated. Segmentations based on the maximization of the posterior and prior probabilities are shown.

Segmentation	SVFMM			$\mathscr{S}$t-SVFMM			NLv-SVFMM			NLp-SVFMM		
	Mean	Median	St. dev.	Mean	Median	St. dev.	Mean	Median	St. dev.	Mean	Median	St. dev.
Posterior	0.7766	0.7680	0.0368	0.7912	0.7817	0.0384	0.7988	0.7833	0.0397	**0.8044**	**0.7936**	0.0351
Prior	0.2231	0.2247	0.0092	0.2467	0.2461	0.0103	**0.2576**	**0.2578**	0.0113	0.2494	0.2495	0.0106

Table 3 Results on the RI over the 300 images of the Berkeley dataset for each algorithm evaluated.

SVFMM				$\mathscr{S}$t-SVFMM				NLv-SVFMM				NLp-SVFMM			
K	Mean	Median	St. dev.	K	Mean	Median	St. dev.	K	Mean	Median	St. dev.	K	Mean	Median	St. dev.
3	**0.6952**	**0.6915**	0.0986	3	0.6941	0.6891	0.0988	3	0.6940	0.6891	0.0988	3	0.6944	0.6897	0.0987
5	0.7274	0.7478	0.1086	5	0.7284	**0.7482**	0.1086	5	0.7284	**0.7482**	0.1085	5	**0.7288**	0.7480	0.1086
7	0.7283	0.7585	0.1208	7	0.7312	0.7596	0.1207	7	0.7313	0.7597	0.1206	7	**0.7316**	**0.7599**	0.1207
10	0.7250	0.7618	0.1335	10	0.7281	0.7632	0.1333	10	0.7283	0.7634	0.1334	10	**0.7288**	**0.7639**	0.1334
15	0.7184	0.7585	0.1431	15	0.7215	0.7594	0.1428	15	0.7214	0.7595	0.1428	15	**0.7221**	**0.7612**	0.1429
20	0.7136	0.7495	0.1479	20	0.7161	0.7538	0.1478	20	0.7162	0.7538	0.1478	20	**0.7166**	**0.7545**	0.1479

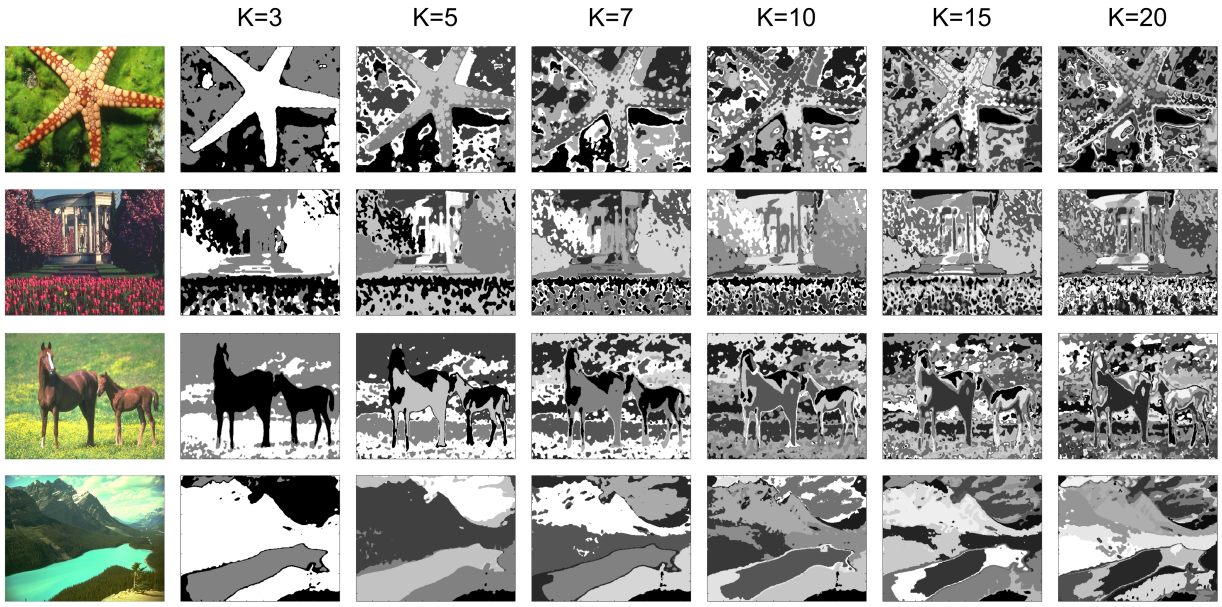

Fig. 5 Example of segmentation maps for $K \in \{3,5,7,10,15,20\}$ obtained with the NLp-SVFMM for 4 images of the Berkeley dataset.

Table 4 Average and std. deviation time comparison (in seconds) for each algorithm evaluated in the study on the Berkeley 300 dataset for each number of segments computed in the images.

K	SVFMM	$\mathscr{S}$t-SVFMM	NLv-SVFMM	NLp-SVFMM
3	0.91 ± 0.04	1.48 ± 0.06	1.73 ± 0.06	2.01 ± 0.09
5	1.39 ± 0.08	2.27 ± 0.12	2.56 ± 0.13	2.99 ± 0.13
7	1.87 ± 0.08	3.05 ± 0.15	3.37 ± 0.14	3.98 ± 0.16
10	2.69 ± 0.14	4.40 ± 0.20	4.92 ± 0.21	5.75 ± 0.17
15	3.91 ± 0.17	6.32 ± 0.22	6.96 ± 0.26	8.23 ± 0.30
20	5.18 ± 0.47	8.39 ± 0.75	9.27 ± 0.82	10.92 ± 0.96

ture of the image. This MRF improves the previously proposed $\mathscr{S}$t-MRF in terms of performance of the model, and also reducing the number of parameters to be estimated. Experimental results demonstrated the superiority of the proposed method with respect to previous state-of-the-art algorithms proposed in the literature when evaluated in a public reference dataset.

Acknowledgements This study is partially supported by Secretaría de Estado de Investigación, Desarrollo e Innovación (DPI2016-80054-R, TIN2013-43457-R) and Agencia Valenciana de la Innovación (INNVAL10/18/048). E.F.G was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (No 844646), and also acknowledges the support of NVIDIA GPU Grant Program.

References

Arthur D, Vassilvitskii S (2007) K-means++: The Advantages of Careful Seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Al-

gorithms, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, SODA '07, pp 1027–1035

Bishop C (2006) Pattern Recognition and Machine Learning. Information Science and Statistics, Springer-Verlag, New York

Blake A, Rother C (2011) Markov random fields for vision and image processing. MIT Press, Cambridge, Mass.

Blekas K, Likas A, Galatsanos NP, Lagaris IE (2005) A spatially constrained mixture model for image segmentation. IEEE Transactions on Neural Networks 16(2):494–498, DOI 10.1109/TNN.2004.841773

Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(11):1222–1239, DOI 10.1109/34.969114

Buades A, Coll B, Morel J (2005) A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol 2, pp 60–65 vol. 2, DOI 10.1109/CVPR.2005.38

Cheng Y (1995) Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(8):790–799, DOI 10.1109/34.400568

Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B (Methodological) 39(1):1–38

Ester M, Kriegel HP, Sander J, Xu X (1996) A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, AAAI Press, Portland, Oregon, KDD'96, pp 226–231

Juan-Albarracín J, Fuster-Garcia E, Manjón JV, Robles M, Aparici F, Martí-Bonmatí L, García-Gómez JM (2015) Automated Glioblastoma Segmentation Based on a Multiparametric Structured Unsupervised Classification. PLOS ONE 10(5):e0125143, DOI 10.1371/journal.pone.0125143

Neal R, Hinton G (1999) A View Of The Em Algorithm That Justifies Incremental, Sparse, And Other Variants. Learning in graphical models 89, DOI 10.1007/978-94-011-5014-9-12

Nikou C, Galatsanos NP, Likas AC (2007) A class-adaptive spatially variant mixture model for image segmentation. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society 16(4):1121–1130

Nikou C, Likas AC, Galatsanos NP (2010) A Bayesian Framework for Image Segmentation With Spatially Varying Mixtures. IEEE Transactions on Image Processing 19(9):2278–2289, DOI 10.1109/TIP.2010.2047903

Pal NR, Pal SK (1993) A review on image segmentation techniques. Pattern Recognition 26(9):1277–1294, DOI 10.1016/0031-3203(93)90135-J

Rokach L, Maimon O (2005) Clustering Methods. In: Maimon O, Rokach L (eds) Data Mining and Knowledge Discovery Handbook, Springer US, Boston, MA, pp 321–352, DOI 10.1007/0-387-25465-X_15

Sanjay-Gopal S, Hebert TJ (1998) Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. IEEE Transactions on Image Processing 7(7):1014–1028, DOI 10.1109/83.701161

Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, Er MJ, Ding W, Lin CT (2017) A review of clustering techniques and developments. Neurocomputing 267:664–681, DOI 10.1016/j.neucom.2017.06.053

Sfikas G, Nikou C, Galatsanos N (2008) Edge preserving spatially varying mixtures for image segmentation. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp 1–7, DOI 10.1109/CVPR.2008.4587416

Wu Y, Tracey B, Natarajan P, Noonan JP (2013) Probabilistic Non-Local Means. IEEE Signal Processing Letters 20(8):763–766, DOI 10.1109/LSP.2013.2263135

Zhang H, Fritts JE, Goldman SA (2008) Image segmentation evaluation: A survey of unsupervised methods. Computer Vision and Image Understanding 110(2):260–280, DOI 10.1016/j.cviu.2007.08.003