Segmenting Multi-Source images using hidden Markov fields with copula-based multivariate statistical distributions.

Journal:	*Transactions on Image Processing*
Manuscript ID	TIP-16184-2016.R2
Manuscript Type:	Regular Paper
Date Submitted by the Author:	17-Feb-2017
Complete List of Authors:	Lapuyade-Lahorgue, Jérôme; Universite de Rouen, LITIS, Quantif; LITIS Xue, Jing-Hao; University College London, Statistical Science Ruan, Su; Universite de Rouen, LITIS, Quantif
EDICS:	2. SMR-SMD Statistical-Model Based Methods < Image & Video Sensing, Modeling, and Representation, 27. COI-TOM Tomographic Imaging < Computational Imaging, 28. COI-MRI Magnetic Resonance Imaging < Computational Imaging, 12. TEC-BIP Biomedical and Biological Image Processing < Image & Video Processing Techniques, 9. TEC-RST Restoration and Enhancement < Image & Video Processing Techniques

SEGMENTING MULTI-SOURCE IMAGES USING HIDDEN MARKOV FIELDS WITH COPULA-BASED MULTIVARIATE STATISTICAL DISTRIBUTIONS

J. Lapuyade-Lahorgue⋆ *J.-H. Xue*⋆⋆ *S. Ruan*⋆

⋆ LITIS, Eq. Quantif, University of Rouen, Rouen, France
⋆⋆ Department of Statistical Science, University College London, London, UK

ABSTRACT

Nowadays, multi-source image acquisition attracts an increasing interest in many fields such as multi-modal medical image segmentation. Such acquisition aims at considering complementary information to perform image segmentation since the same scene has been observed by various types of images. However, strong dependency often exists between multi-source images. This dependency should be taken into account when we try to extract joint information for precisely making a decision. In order to statistically model this dependency between multiple sources, we propose a novel multi-source fusion method based on the Gaussian copula. The proposed fusion model is integrated in a statistical framework with the hidden Markov field inference in order to delineate a target volume from multi-source images. Estimation of parameters of the models and segmentation of the images are jointly performed by an iterative algorithm based on Gibbs sampling. Experiments are performed on multi-sequence MRI to segment tumors. The results show that the proposed method based on the Gaussian copula is effective to accomplish multi-source image segmentation.

Index Terms— Data fusion, copulas, multi-source images, tumor segmentation, Bayesian inference, hidden Markov fields.

1. INTRODUCTION

Segmentation of multi-source images is being widely used for many applications with the advances in imaging technology. Using multi-source data can improve segmentation because they allow to reduce data ambiguity and explore dependency between the different sensors. The main challenge of image segmentation from multi-sources is how to fuse different sources to get the best compromise for a precise segmentation. Data fusion can be considered as handling the automatic detection, association, estimation and combination of data and information from several sources [1]. It depends usually on the context of applications. Hence, it relates to numerous mathematical and technical domains. The most known methods as reviewed in [2] include the methods based on belief functions theory, fuzzy sets theory and probabilistic models.

In the methods based on the belief function theory, each source is first modeled by an evidential mass. The Dempster-Shafer rule is then applied to fuse all sources [3, 4, 5]. The main applications of the belief function theory can be found in medical imaging [5], scene understanding [6] and geo-science [7].

In the methods based on the fuzzy set theory, the fuzzy measure quantifies the degree of membership relative to a decision for each source. The fusion of several sources is achieved by applying the fuzzy operators to the fuzzy sets [8, 9, 10, 11, 12, 13]. Fuzzy fusion is widely used in many fields, such as in medical imaging [14].

The main difficulty to use the belief function theory and the fuzzy set theory relates to the choice of the evidential mass, the fuzzy measure and the fuzzy conjunction function. Some hybrid fusion methods have been proposed to overcome the difficulties. In [15], it is proposed to estimate the evidential mass in a Bayesian context; [16] combines both the Dempster-Shafer theory and the fuzzy set theory. Other types of fusion methods have also been proposed, such as machine learning techniques using "Support Vector Machine" (SVM) [17], non-linear classifier [18] and neural networks [19, 20].

Probabilistic methods have achieved great success in image and signal processing such as speech recognition [21], radar signal processing [22], frequency and pitch recognition [23, 24], bioscience [25, 26], climatology [27], gesture recognition [28] and medical imaging [29]. In probabilistic methods, the uncertainty of data is modeled by a probability distribution. Different models have been proposed, such as hidden Markov chains [30], hidden Markov fields, Conditional Random Fields [31] and Bayesian networks [32] to deal with image segmentation. Conditional Random Fields [31] consider the dependence of image points based on the conditional distribution of the hidden states given the observations, leading to global spatial dependency between voxels. Consequently, it results in an increased complexity for inference than the hidden Markov fields. In Bayesian networks, a topology should be defined. A natural choice of this topology for dealing with images is a hidden Markov field. Generative probabilistic model and discriminative extensions as discussed in [33] are particular cases of Bayesian networks,

which have been used for the brain lesion segmentation using a probability atlas as prior information. For the estimation of these models, Bayes estimators are usually used to make decision from different sources. The computation of the Bayes estimator can be achieved in different ways such as the Baum-Welsh and Viterbi algorithms [34, 35], the Kalman filtering [30] and its extensions [36], "Markov Chain Monte Carlo" (MCMC) [37], particle filtering [38] and Gibbs sampling [39].

As the contrast is low for many voxels, if a segmentation with naïve Bayes method or an HMM is separately applied in each type of image, these voxels will be misclassified. A final combination of all individual segmentations would be too late to correct the misclassification. For this reason, we propose to model the dependency between modalities by a copula. From the point of theory view, Markov fields can easily take into account the correlation between multi-variables by using copula; and from the point of application view, the tumor and non-tumor regions verify well the properties of Markov fields. Therefore, hidden Markov fields are chosen as the base of our unsupervised segmentation. Amongst the previous use of copula, [40] gives an overview of the different uses of copula in finance. In [41, 42], the copula is used to model dependency between different scales of a multiscale image to segment mono-modal images. In [22], the copula measures the dependency inside complex sampled signals to model the frequency distribution. In [43], the copula is used in the pairwise Markov chain to model the dependency between voxel intensities in a mono-modal image, while in [44], the copula is used for multiband image segmentation inside a hidden Markov tree. It also has been used for texture analysis [45] and image retrieval [46]. As far as we know it has not been used in fusion of multi-source images.

In this paper, we propose a new probabilistic fusion method which aims to consider the dependency between different modalities of images. The method uses the copula to model dependency between the intensity values of the different modalities, because copula can express the joint cumulative density function from the marginal cumulative density functions. Hence, it can be considered as a good way to fuse the different modalities. We focus on the Gaussian copula in this work. A non-standard Gamma distribution is chosen to model the marginal density of each image source. A probabilistic Bayesian framework based on the Gibbs sampling algorithm is proposed to jointly segment images and estimate model parameters. An alternative algorithm is the Hastings-Metropolis algorithm [47]. However, the Hastings-Metropolis algorithm needs to choose correctly the instrumental distribution to obtain good results. The proposed method is applied on MRI medical images in order to segment tumors in 3D. The MRI (Magnetic Resonance Imaging) images are got with three modalities: FLAIR, T1 and T2, which are used clinically to observe brain tumor.

The paper is organized as follows. In section 2, we detail the Bayesian inference of the hidden Markov field with copulas. In section 3, we present some experiments on medical images. The experiments are conducted on simulated MRI images with different contrasts and real MRI images.

2. METHODOLOGY

The choice of the observation models is crucial for Bayesian inference. It should fit as accurately as possible to the data. The choice concerns two parts: the marginal statistical model for each image source and the dependency between them. We introduce a formulation of multivariate statistical distribution which is defined from marginal distributions and an aggregation term (e.g. copula) to model multi-source data.

2.1. Definition of copula

The notion of copula was born from the works of Sklar [48]; its first real application was in finance [40] and has progressively met success in other application domains such as image and radar processings [22]. The copula is a function allowing to describe the joint distribution of a multivariate random vector from its marginal distributions. In general, it is impossible to deduce the joint distribution only from the marginal distributions except in independent case. However, with a copula to fuse these marginal distributions, it is possible.

Let $\vec{Y} = (Y^{(1)}, \ldots, Y^{(M)})$ be a random vector whose values are taken from a subset of $\mathbb{R}^M$. We recall that the joint cumulative density function F of $\vec{Y} = (Y^{(1)}, \ldots, Y^{(M)})$ is a function defined on $\mathbb{R}^M$ and taking its values in $[0, 1]$ such that $F(y^{(1)}, \ldots, y^{(M)})$ represents the probability of the event $(Y^{(1)} \leq y^{(1)}, \ldots, Y^{(M)} \leq y^{(M)})$. Let $F^{(1)}, \ldots, F^{(M)}$ be the marginal cumulative density functions. The copula is used to represent the cumulative function F from the marginal cumulative density functions $F^{(1)}, \ldots, F^{(M)}$. More precisely, the relation is given by the Sklar theorem [48] demonstrating that there exists a function $C : [0, 1]^M \to [0, 1]$ called copula (of $\vec{Y}$) such that for any $(y^{(1)}, \ldots, y^{(M)}) \in \mathbb{R}^M$,

$$F(y^{(1)}, \ldots, y^{(M)}) = C(F^{(1)}(y^{(1)}), \ldots, F^{(M)}(y^{(M)})). \quad (1)$$

We can interpret the function C as a way to fuse the sources $(F^{(1)}(y^{(1)}), \ldots, F^{(M)}(y^{(M)}))$ to construct an aggregated source $F(y^{(1)}, \ldots, y^{(M)})$. Moreover, the copula can represent dependency between random variables. Indeed, if there exists $(w^{(1)}, \ldots, w^{(M)}) \in [0, 1]^M$ such that $C(w^{(1)}, \ldots, w^{(M)}) \neq w^{(1)} \times \ldots \times w^{(M)}$, then the marginals are dependent random variables. Consequently, copulas can be used to statistically fuse different information sources by taking into account their dependency.

The function C is the joint cumulative density function of the random vector $(F^{(1)}(Y^{(1)}), \ldots, F^{(M)}(Y^{(M)}))$ whose marginals follow uniform distributions on $[0, 1]$. In the case where C is derivable, the respective densities of the marginals

$Y^{(1)}, \ldots, Y^{(M)}$ and the random vector $\vec{Y}$ are linked through the relation:

$$p(y^{(1)}, \ldots, y^{(M)}) = p^{(1)}(y^{(1)}) \times \ldots \times p^{(M)}(y^{(M)})$$
$$\times c\left(F^{(1)}(y^{(1)}), \ldots, F^{(M)}(y^{(M)})\right) \quad (2)$$

where $c = \frac{\partial^M C}{\partial w^{(1)} \partial w^{(2)} \ldots \partial w^{(M)}}$ is the derivative of C called "copula density" of the random vector $(Y^{(1)}, \ldots, Y^{(M)})$. The dependency modeled by copula does not depend on the choice of marginals. For example, if the marginals are Gaussian, it is possible to model the dependency from the covariance matrix. If the marginals are not Gaussian, the joint distribution may not be deduced only from the marginals. However, the dependency can still be modeled by a Gaussian copula.

2.2. Marginal density model

The expression $p^{(m)}(y^{(m)})$ in (2) represents the m-th marginal density of the image source. Its choice depends on applications. We are interested in tumor segmentation in our application. Fig. 1 shows three types of 3D MRI images of brain: FLAIR, T1 and T2 reflecting different signals to observe the same tumor to be segmented (SimBRATS [49]). Green contours represent the ground truth of the tumor area. Correspondingly, three histograms of the tumor volumes are shown in Fig. 2. Asymmetry of the histograms is clearly observed. Hence, Gaussian distribution is not appropriate for such images. A non-standard Gamma distribution is proposed to model the marginal density of each image source, given by:

$$\frac{1}{\Gamma(a)b^a}(y-\mu)^{a-1} \exp\left(-\frac{y-\mu}{b}\right) \quad (3)$$

where $y \in [\mu, +\infty[$, $\mu \in \mathbb{R}$ is the lower bound, $a > 0$ is the shape parameter, $b > 0$ is the dispersion parameter and Γ is the Eulerian Gamma function.
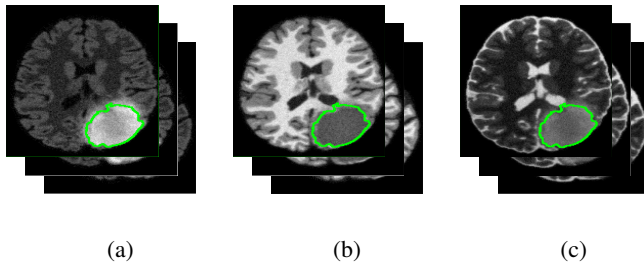
Fig. 1. Some slices of three MRI image volumes corresponding to three types of acquisition coming from SimBRATS [49]: (a) FLAIR, (b) T1 and (c) T2. The brain tumor contour is represented in green.

Fig. 2 shows comparisons between the histograms of the tumor volume and the estimated densities with the Gamma distribution. The parameters of the density of the non-standard

(a)   (b)   (c)
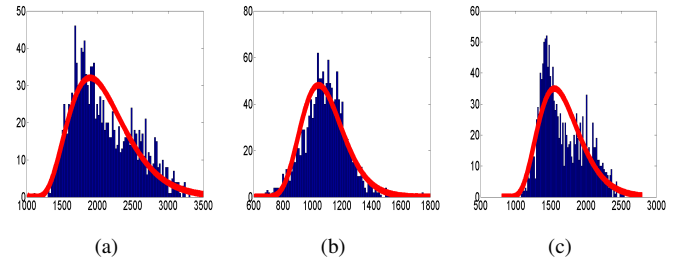
Fig. 2. Comparison between histograms of MRI intensities in the tumor volume and those of the estimated non-standard Gamma distributions by using moment estimators for (a) FLAIR, (b) T1 and (c) T2.

Gamma distribution were estimated from the intensities of the voxels belonging to the tumor area in the ground-truth by using the moment estimators as to be detailed in section 2.5.1. According to a Kolmogorov test, we obtained p-value equal to 0.99, which allows us to validate the choice of the Gamma distribution.

2.3. Modeling the dependency of multi-source

The definition of a multivariate distribution from its marginals and the corresponding copula open a door to fuse multi sources when each marginal is known. The choice of the copula is made independently from the choice of the marginals. In this paper, we focus on the "Gaussian copula" whose density is given by [50]:

$$c\left(w^{(1)}, \ldots, w^{(M)}\right) =$$
$$\frac{1}{\sqrt{\det \Lambda}} \exp\left[-\frac{1}{2}\phi^{-1}(\vec{w})^T \left(\Lambda^{-1} - I\right) \phi^{-1}(\vec{w})\right] \quad (4)$$

for any $(w^{(1)}, \ldots, w^{(M)}) \in [0, 1]^M$, where Λ is a correlation matrix (ie. a symmetric and positive definite matrix whose diagonal elements are 1), I is the identity matrix and $\phi^{-1}(\vec{w}) = \left(\varphi^{-1}(w^{(1)}); \ldots; \varphi^{-1}(w^{(M)})\right)$ the column vector where φ^{-1} is the inverse of the cumulative density function of a standard Gaussian variable. The non-diagonal coefficients of the correlation matrix are called "correlation coefficients" of the Gaussian copula. In the case where Λ is the identity matrix I, then $c \equiv 1$ and the marginals of any random vector having this copula are independent. Simulations are carried out to show Gaussian copula behaviors in case of $M = 3$. Fig. 3 presents 1000 realizations of a three-dimensional Gaussian copula for $\Lambda = \begin{pmatrix} 1 & 0.9 & -0.6 \\ 0.9 & 1 & -0.8 \\ -0.6 & -0.8 & 1 \end{pmatrix}$. Each realization (W_1, W_2, W_3) produces a point in the three-dimensional volume $[0, 1]^3$. In order to easily visualize the realizations of the three-dimensional copula, we project them in 2D corresponding to (W_1, W_2) where $\rho = 0.9$, (W_1, W_3) where $\rho = -0.6$

and (W_2, W_3) where $\rho = -0.8$, with ρ being the correlation coefficient of the Gaussian copula.


(a)   (b)   (c)
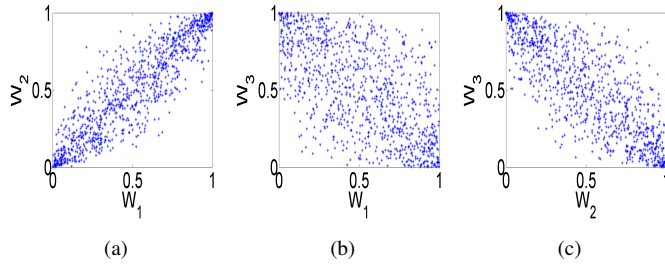
Fig. 3. Simulation of a three-dimensional Gaussian copula. For a easy visualization, projected points in pairs of modalities are shown with coefficient correlations equal to (a) 0.9, (b) -0.6 and (c) -0.8, respectively.

One can see that more nearly the absolute value of the correlation coefficient in Λ closes to 1, more number of points close to the diagonal. Fig. 4 presents joint intensities of a pair of MRI images corresponding respectively to FLAIR-T1, FLAIR-T2 and T1-T2. A strong dependency between the modalities FLAIR and T2 is observed, which is in concordance with the fact that the FLAIR modality has a similar acquisition way as the T2 modality only with a contrast enhancement in addition. To model the correlation between the different types of MRI images, we have carried out some simulations. Fig. 5 presents the simulations of joint distributions with Gaussian copula and non-standard Gamma margins. The parameters of the model were estimated from the MRI images.

(a)   (b)   (c)
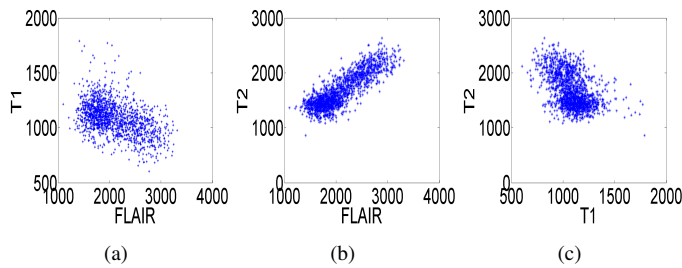
Fig. 4. Joint intensity distribution of MRI images [49]: (a) FLAIR-T1, (b) FLAIR-T2 and (c) T1-T2. The intensity of the first modality is read on abscissa axis and that of the second modality on the ordinate axis.

We can see that, by using the Gaussian copula and non-standard Gamma distributions, we can well reproduce the real joint distributions.

(a)   (b)   (c)
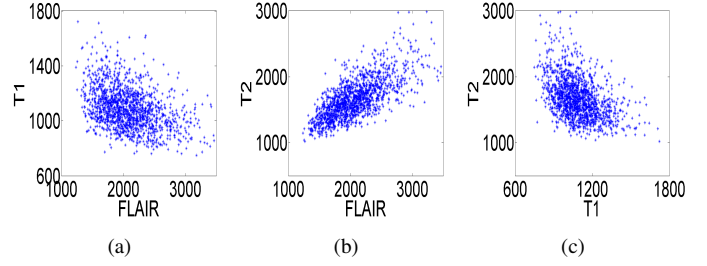
Fig. 5. Joint intensity distribution of our simulated MRI images: (a) FLAIR-T1, (b) FLAIR-T2 and (c) T1-T2. The intensity of the first modality is read on abscissa axis and that of the second modality on the ordinate axis.

2.4. Bayesian image segmentation using hidden Markov fields

As the joint distribution model of all image sources and the marginal density models are known, Bayesian segmentation can be performed. From the Bayesian point of view, image segmentation is represented as a realization $x_V = (x_u)_{u \in V}$ of a random field $X_V = (X_u)_{u \in V}$, with V being a finite subset representing the coordinates of each voxel, $x_u \in \{1, \ldots, K\}$ the class label for a voxel of coordinates u and K the number of classes. The vectorial observations are represented as a realization $\vec{y}_V = (\vec{y}_u)_{u \in V}$, where $\vec{y}_u = \left(y_u^{(1)}, \ldots, y_u^{(M)} \right)$ and $y_u^{(m)}$ is intensity of the voxel of coordinates u in the mth modality. The segmentation consists in recovering x_V from $\vec{y}_V$ by maximizing the posterior conditional probability $p(x_V | \vec{y}_V) \propto p(\vec{y}_V | x_V) \times p(x_V)$. In image segmentation, we can suppose a spatial dependency between the neighbored voxels. The easiest model for $p(x_V)$ taking into account the dependency is the Markov field in which

$$p(x_u | (x_t)_{t \neq u}) = p(x_u | (x_v)_{v \in V_u}), \qquad (5)$$

where V_u is a set of neighbors of the voxel of coordinates u. In this paper, we have chosen the Ising model given by

$$p(x_u | (x_v)_{v \in V_u}) \propto \exp \left[-\frac{1}{T} \sum_{v \in V_u, v \neq u} |x_u - x_v| \right], \quad (6)$$

where T is the "temperature" parameter to be estimated. In the model of hidden Markov field, the relation between the hidden states x_V and the observations y_V is given by:

$$p(\vec{y}_V | x_V) = \prod_{u \in V} p(\vec{y}_u | x_u) \qquad (7)$$

with:

$$
\begin{aligned}
p(\vec{y}_u | x_u) &= \left[\prod_{m=1}^{M} p^{(m)}(y_u^{(m)} | x_u) \right] \\
&\times c_{x_u} \left(F_{x_u}^{(1)}(y_u^{(1)}), \ldots, F_{x_u}^{(M)}(y_u^{(M)}) \right)
\end{aligned} \qquad (8)
$$

where $p^{(m)}(y_u^{(m)}|x_u)$ is the marginal distribution for the class x_u, $F_{x_u}^{(m)}$ the corresponding cumulative density function and c_{x_u} is the Gaussian copula for the class x_u. The distribution $p(\vec{y}_u|x_u)$ is a conditional multivariate distribution defined from its marginal distributions and its copula. The parameters of the marginal distributions and the copula depend on x_u, represented by $\Theta = \{\mu_i, a_i, b_i, \Lambda_i : i \in \{1, \ldots, K\}\}$. Therefore, we propose to integrate the estimation of these parameters to the segmentation process.

2.5. Parameter estimation

Gibbs sampling algorithm [51] is a good parameter estimation method as the dependence on the initialization is weak. Based on it, we propose to simultaneously estimate the hidden states x_u and parameters of $p(x_u)$ and $p(\vec{y}_u|x_u)$ with the following algorithm:

1. Initialization: Simulate for each $u \in V$, the hidden state $x_u^{(0)}$ according to the uniform distribution on $\{1, \ldots, K\}$.

2. Iteration: from the hidden state $(x_u^{(k)})_{u \in V}$:

    (a) Estimate the parameters of each marginal $p^{(m)}(y_u^{(m)}|x_u)$ from the complete data $(x_u^{(k)}, \vec{y}_u)_{u \in V}$.

    (b) Estimate the parameters of the copula c_{x_u} from the complete data $(x_u^{(k)}, \vec{y}_u)_{u \in V}$.

    (c) Estimate the temperature of the hidden distribution $p(x_u|(x_v)_{v \in V_u})$ from $(x_u^{(k)})_{u \in V}$.

    (d) Simulate $(x_u^{(k+1)})_{u \in V}$ from the posterior distribution:

    $$p(x_u|(x_v^{(k)})_{v \in V_u}, \vec{y}_u) \quad \propto \quad p(x_u|(x_v^{(k)})_{v \in V_u})$$
    $$\times \quad p(\vec{y}_u|x_u).$$

3. Repeat the step 2. until convergence. The convergence is reached when the Euclidean distance between the previously estimated values of the parameters and the updated values does not exceed 10^{-7} during 10 iteration steps.

2.5.1. Estimation of parameters of marginal distribution (step a)

Let us look at how to estimate the parameters of the marginal distributions from moment estimators for non-standard Gamma distributions. The first step consists in estimating the bounds. The observation is transformed into a sample of standard Gamma distribution in order to estimate the shape and dispersion parameters using a moment estimator. Let us consider $(x_u^{(k)}, \vec{y}_u)_{u \in V}$ the complete data. Let $(\vec{y}_u)_{u \in V(i)}$ be the set

of observations where $x_u^{(k)} = i \in \{1, \ldots, K\}$ being the class for which we estimate the parameters. The minimal parameter $\mu_{i,m}$ for the class i in the modality m is easily estimated by $\hat{\mu}_{i,m} = \min(y_u^m : u \in V(i))$. After estimating the minimum, we transform the sample into $z_u^{(m)} = y_u^{(m)} - \hat{\mu}_{i,m}$ for the Gamma distribution when $x_u^{(k)} = i$. The shape parameter $a_{i,m}$ and the dispersion parameters $b_{i,m}$ are then estimated from the sequence $(z_u^{(m)})_{u \in V(i)}$ using the classical first and second order moment estimators:

$$\hat{b}_i = \frac{\text{Var}((z_u^{(m)})_{u \in V(i)})}{\text{Mean}((z_u^{(m)})_{u \in V(i)})}$$

$$\hat{a}_i = \frac{\text{Mean}((z_u^{(m)})_{u \in V(i)})}{\hat{b}_i} \quad (9)$$

where Mean and Var are the empirical estimators for mean and variance respectively.

2.5.2. Estimation of the parameters of the copula (step b)

The first step consists in transforming the observations $(\vec{y}_u)_{u \in V(i)}$ into a sample of random vector $\vec{w}_u = (w_u^{(1)}, \ldots, w_u^{(M)})$, realization of the copula. Let $\hat{F}_i^{(m)}$ be the cumulative density function for the mth modality in regard of the class i with the parameters estimated in the previous step.

The transformation by $\vec{w}_u = \left(\hat{F}_i^{(1)}(y_u^{(1)}), \ldots, \hat{F}_i^{(M)}(y_u^{(M)})\right)$ can be considered as a realization of the random vector whose cumulative density function is the copula to be estimated. For Gaussian copula, we use the fact that $(\varphi^{-1}(w_u^{(1)}), \ldots, \varphi^{-1}(w_u^{(M)}))$ is a realization of a Gaussian vector with an appropriate correlation matrix (i.e. the parameter of the Gaussian copula) when φ is the cumulative density function of a standard Gaussian real random variable. Consequently, the correlation matrix for the class i is estimated using the empirical Pearson correlation estimator from $(\varphi^{-1}(w_u^{(1)}), \ldots, \varphi^{-1}(w_u^{(M)}))$ when $x_u = i$. This estimation is given by

$$\hat{\Lambda}_i = \hat{D}_i^{-\frac{1}{2}} \hat{\Sigma}_i \hat{D}_i^{-\frac{1}{2}} \quad (10)$$

where $\hat{D}_i$ is the diagonal matrix whose elements are the diagonal elements of the matrix $\hat{\Sigma}_i$ given by:

$$\hat{\Sigma}_i = \frac{1}{N_i} \sum_{u \in V(i)} (\varphi^{-1}(w_u^{(1)}); \ldots; \varphi^{-1}(w_u^{(M)}))$$
$$\times \quad (\varphi^{-1}(w_u^{(1)}); \ldots; \varphi^{-1}(w_u^{(M)}))^T \quad (11)$$

where N_i is the number of voxels u such that $x_u^{(k)} = i$ and $V(i)$ the subset of voxels u such that $x_u^{(k)} = i$. All parameters used in our framework are automatically estimated.

2.5.3. Estimation of the hidden field distribution (step c)

The temperature T of the hidden field distribution $p(x_u|(x_v)_{v \in V_u})$ is estimated by using the maximum likelihood. The likelihood for the hidden states is given by

$$\prod_{u \in V} \frac{1}{Z_{V_u}(T)} \exp \left[-\frac{1}{T} \sum_{v \in V_u : v \neq u} |x_u - x_v| \right], \qquad (12)$$

where

$$Z_{V_u}(T) = \sum_{i=1}^{K} \exp \left[-\frac{1}{T} \sum_{v \in V_u : v \neq u} |i - x_v| \right] \qquad (13)$$

in which K is the number of classes.
The log-likelihood is then given by:

$$\sum_{u \in V} \left[-\log(Z_{V_u}(T)) - \frac{1}{T} \sum_{v \in V_u : v \neq u} |x_u - x_v| \right]. \qquad (14)$$

By differentiating the log-likelihood, one can show that the estimation $\hat{T}$ is given by solving the equation:

$$\sum_{u \in V} \left(\sum_{v \in V_u} |x_u - x_v| - \frac{\hat{T}^2 Z'_{V_u}(\hat{T})}{Z_{V_u}(\hat{T})} \right) = 0, \qquad (15)$$

where $Z'_{V_u}(\hat{T})$ is the derivative of $Z_{V_u}(\hat{T})$ according to $\hat{T}$.
This non-linear equation is solved using the Newton-Raphson algorithm.

3. EXPERIMENTATION

3.1. Data

The goal of our method is to delineate tumor volumes from multi-source images. In our experimental studies, we focus on MRI 3D images. The six neighboring voxels include the four closest points in the current image and the two closest points in the previous and the next image slices. The datasets composed of 3D MRI images including 50 simulated images and 10 real patient images which come from BRATS (Multi-modal Brain Tumor Image Segmentation Benchmarks) [49]. The MRI data have three types of images, T1, T2 and FLAIR. Six classes are considered for segmentation: tumor, white substance, gray substance, edema, meninges and background. For evaluating the results, ground truths which are manual delineations of the tumor volume achieved by medical specialists and the following measures are used: the Dice-Sørensen coefficient D, the false positive ratio P, the false negative ratio N and the false detection ratio F. These coefficients are respectively defined as $D = 2 \times \frac{|T \cap E|}{|T| + |E|}$, $P = \frac{|E \setminus T|}{|T| + |E \setminus T|}$, $N = \frac{|T \setminus E|}{|T|}$ and $F = \frac{P + N}{2}$, where E and T are respectively the estimated and true core tumor volumes.

The false positive ratio that we apply is a variant of the usual definition. In the usual definition, the number of false positives is divided by the sum of the number of the false positives and the number of the true negatives. However, this definition is not an effective measure in the cases of extreme imbalance between classes. For example in an application like ours, the number of non-tumor (true negative) pixels is much larger than the number of the tumor (true positive) pixels and the number of pixels wrongly classified as tumor (false positive), resulting in a very small false positive ratio which cannot reflect the real concern of the medical practitioners. Hence, in the variant definition that we apply, the relative size of the false positives to the true tumor is instead considered.

3.2. Evaluation on simulated data in function of contrast

In this section, we first aim at the evaluation of the robustness of our fusion method in terms of contrast-to-noise ratio. The performances are evaluated with differently contrasted images. A comparison between the results obtained by segmentations with and without fusion is performed to demonstrate better performance achieved by fusion.
In order to generate differently contrasted images, we simulate three images respectively for FLAIR, T1 and T2 with fixed parameters of the Gamma marginals and the Gaussian copula. The contrast between the tumor and its surrounding is changed by making the following transform for each modality:

$$y_u := y_u - (1 - \alpha) \times (\mu(\text{tumor}) - \mu(\text{around})) \qquad (16)$$

where $\mu(\text{tumor})$ is the mean of the voxel intensities in the tumor volume, $\mu(\text{around})$ is the mean of the voxel intensities in the region surrounding the tumor volume and $\alpha \in [0,1]$ controls the contrast. Fig. 6 presents the different simulated MRI modalities with the four different contrasts $\alpha = 1$, $\alpha = 0.75$, $\alpha = 0.5$ and $\alpha = 0.25$. Fig. 7 presents the segmentation results on the simulated images for the three contrasts with $\alpha = 1$, $\alpha = 0.5$ and $\alpha = 0.25$. The results are obtained by using three methods: mono-modal segmentation from only T1 (Fig. 7a), the fusion with $c \equiv 1$ (independence sources) (Fig. 7b) and the fusion with the Gaussian copula from the three modalities (Fig. 7c). The segment contours (in blue) and the ground truth contours (in green) are superposed on the simulated T1 image. We can observe that the best results were obtained by the fusion with the copula to model the dependency between the modalities.
Fig. 8 and Fig. 9 present the trends of the Dice coefficient and of the false detection ratio in function of the contrast for the mono-modal segmentations, the multi-modal fusion based segmentation with and without copula. The segmentation without the copula means that we suppose the independence of the three types of images.
From Fig. 8 and Fig. 9, it can be observed that the fusion method with copula is better than the other two methods.
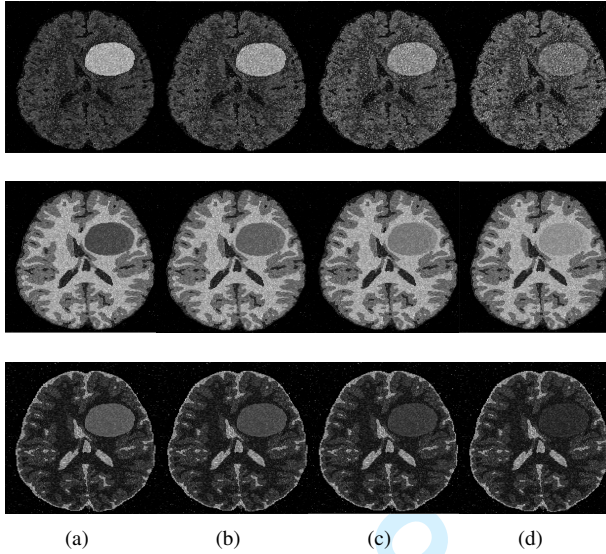
Fig. 6. Simulations of MRI images with different contrasts. Rows from top to bottom: FLAIR, T1 and T2 respectively. Columns: contrasts with (a) $\alpha = 1$, (b) $\alpha = 0.75$, (c) $\alpha = 0.5$ and (d) $\alpha = 0.25$.

Moreover, the performance of the proposed multi-source fusion method decayed much slower than other methods when the contrast in images worsened.

3.3. Segmentation of MRI images

The dataset is composed of 25 simulated high-grade images (tumor with high grade, noted as HG), 25 simulated low-grade images (tumor with low grade, noted as LG) of the BRATS challenge [49] and ten real patient images are used to test our method, in order to demonstrate that the fusion with a Gaussian copula can really improve segmentation performance. We choose ten real HG images to show the improvements obtained with the Gaussian copula. The average size of tumors is 68.926 mL (68926 voxels) with a minimum of 25.234 mL (25234 voxels) and a maximum of 110.256 mL (110256 voxels). It is worth noting that the size of tumor should not be too small to guarantee the reliability of the estimated statistical models. As we focus on tumor segmentation, the Dice coefficient is calculated only for the tumor volume. Fig. 10 presents segmentation results of a simulated image and three real images obtained from mono-modal segmentations and the fusion using the Gaussian copula. In simulated images, the results on T2 and FLAIR are far from the ground truth. Although the results on T1 are better, they are still worse than fusion based segmentation. In real patient images, T1 images are very low-contrasted so that the results are also often far from the ground truth. Our fusion
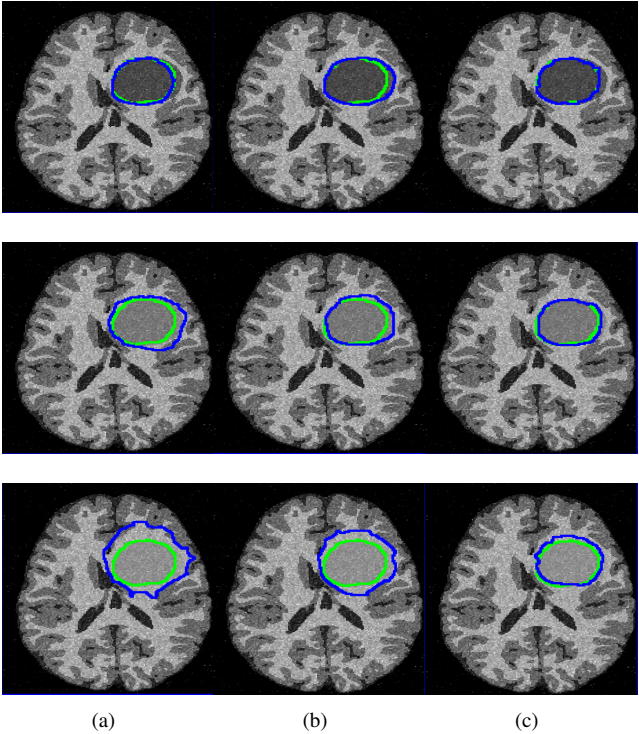

Fig. 7. Segmentation results for the simulated MRI images with brain tumors. From the first row to the last one: contrasts are $\alpha = 1$, $\alpha = 0.5$ and $\alpha = 0.25$ respectively. Columns: results obtained by using: (a) T1 image only (b) fusion with $c \equiv 1$ (independent sources) and (c) fusion with the Gaussian copula. True contour is in green, estimated contour is in blue.
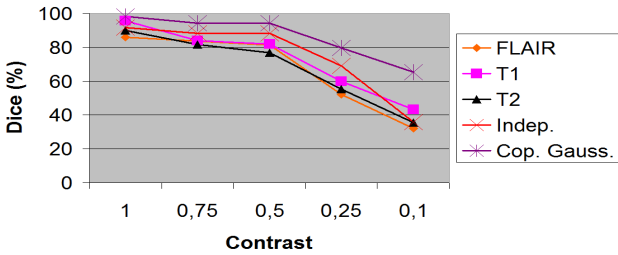
Fig. 8. The Dice coefficient in function of the contrast. Five segmentation methods are compared: the proposed method with the copula (Cop.Gauss), fusion with $c \equiv 1$ (Indep.) and mono-modal segmentation in T1, T2 and FLAIR images.

method provides good results, moreover they are stable regardless of the patients. Table 1 presents respectively the mean and standard deviation of the Dice coefficients and false detection ratios using the mono-modal segmentations and the multi-modal fusion-based segmentations with or without copula for the 25 LG and 25 HG simulated images in the chal-
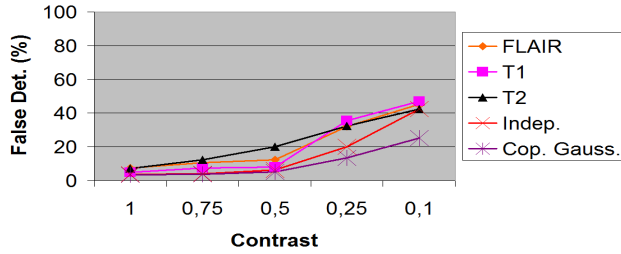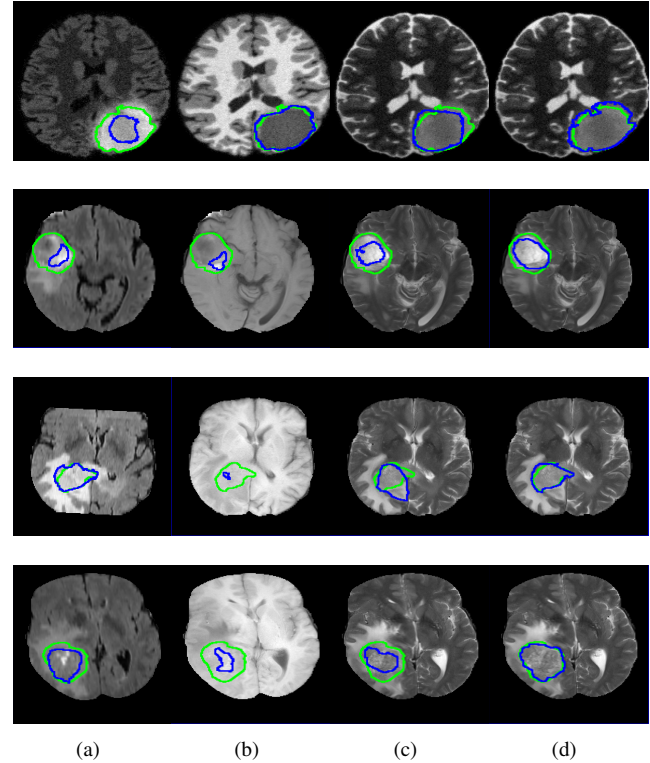
Fig. 9. The false detection ratio in function of the contrast. Other caption is same as for Fig. 8.

lenge. The average Dice coefficient and the average false detection ratio show that the best performance is obtained when the copula is considered for multi-source segmentation. Table 2 presents the comparison of our method with the others of the BRATS challenge [49] in terms of the Dice coefficient. Most of the methods are based on supervised segmentation (or machine learning), such as hierarchical random forest with energy minimization [52, 53], hybrid clustering-logistic regression [54] and decision tree [55]. Our method is placed as the second best method for both HG and LG. The best method of the challenge is based on decision tree providing an average Dice coefficient (DC) of 85.5%, but it needs a learning step, while our method is an unsupervised one with a DC of 83%. Compared with the Markov random fields based method [56] which uses also a Bayes framework, our method provides much better results. That is because our method takes into account the dependency of the three types of images, unlike that method supposing independence of each type of images.

Fig. 10. Segmentation results of brain tumor. From the first row to the third row: one simulated patient and three real patients. Columns: (a)-(c) results of mono-modal segmentations on FLAIR, T1 and T2 images, respectively, and (d) shows the results of multi-modal segmentation with the copula. True contours are indicated in green, while the estimated contour is in blue.

	Mono-M. LG/HG			**Multi-M. LG/HG**	
DC %	FL.	T1	T2	Ind.	G. Cop.
Mean	68/56	58/59	45/48	77/68	85/81
Std.	16/23	7/23	22/30	6/20	3/11
FD %	FL.	T1	T2	Ind.	G. Cop.
Mean	26/30	35/34	45/31	17/20	13/15
Std.	12/32	5/35	17/32	5/20	2/10

Table 1. Average Dice coefficients (DC) and average false detection (FD) ratio of tumor segmentation over 50 synthetic MRI-HG and LG images obtained by using mono-modal (Mono-M.) segmentations respectively in FLAIR (FL.), T1 and T2, and multi-modal (Multi-M.) segmentation without copula (Ind.) and with copula (G. Cop.)

BRATS 2012	**DC (%)**	
Method	**LG/HG/Mean**	**Unsup.**
Gaussian copula [ours]	85/81/83	Yes
Hierarchical MRF [56]	42/40/41	Yes
Tumorcut [57]	46/80/63	Yes
Logistic regression [54]	2/4/3	Yes
Decision tree [55]	84/87/85	No
Random forest [52]	86/78/82	No
Decision forests [53]	54/66/60	No

Table 2. Comparison of the different methods for the multimodal segmentation of the core brain tumors from BRATS 2012.

We have estimated the mean correlation coefficients with the Gaussian copula over the 25 simulated images: they are equal to -0.52, 0.89 and -0.63 respectively for the pairs FLAIR-T1, FLAIR-T2 and T1-T2. Therefore, a real correla-

tion exists in data. Table 3 lists the average Dice coefficients of the ten real patient MRI images for the mono-modal segmentations and the fusion with and without Gaussian copula. As shown by all these results, the fusion method with copula can improve considerably the segmentation compared to the mono-modal segmentations.

DC %	Mono-M.			Multi-M.	
	FL.	**T1**	**T2**	**Ind.**	**G. Cop.**
Mean	77	31	47	74	81
Std.	19	25	27	14	10
FD %	**FL.**	**T1**	**T2**	**Ind.**	**G. Cop.**
Mean	21	58	42	23	17
Std.	21	31	34	13	11

Table 3. Average Dice coefficients (DC) and average false detection (FD) ratio of tumor segmentation over 10 real MRI images obtained by using mono-modal (Mono-M.) segmentations respectively in FLAIR, T1 and T2, and multi-modal (Multi-M.) segmentation without copula (Ind.) and with copula (G. Cop.)

4. CONCLUSION

In this paper, we have presented a Bayesian method to segment a target from multi-source images. Gaussian copula can capture the dependency between the different sources in the target region, and improve considerably the results comparing to the fusion supposing independence of each source or mono-modal segmentation. We have tested our method on three types of MRI images: T1, T2 and FLAIR. Our method provided better results compared with the most of the segmentation methods in [49]. Moreover, this work opens a door to develop more general techniques, such as using other copulas or selecting automatically a convenient copula, and other distributions of marginals according to applications. We have validated our method on the BRATS 2012 challenge data for illustration. Our method will be applied to other datasets including the BRATS challenge data from more recent years for further validation.

5. REFERENCES

[1] L. A. Klein, "Sensor and Data Fusion. Concepts and Applications," *Society of Photo-optical Instrumentation Engineers*, 1999.

[2] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, pp. 28–44, 2013.

[3] L. A. Zadeh, "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination," *AI Magazine*, vol. 7, pp. 85–90, 1986.

[4] P. Smets, "The transferable belief model for quantified belief representation," *Handbook of defeasible reasoning and uncertainty management systems*, vol. 1, pp. 267–301, 1998.

[5] B. Lelandais, S. Ruan, T. Denoeux, P. Vera, and I. Gardin, "Fusion of multitracer PET images for dose painting," *Medical Image Analysis*, vol. 18, pp. 1247–1259, 2014.

[6] P. Xu, F. Davoine, J. B. Bordes, H. Zhao, and T. Denoeux, "Information fusion on oversegmented images: An application for urban scene understanding.," International Conference on Machine Vision and Application, 2013, pp. 189–193.

[7] O. Althuwaynee, B. Pradhan, and S. Lee, "Application of an evidential belief function model in landslide susceptibility mappling," *Computers and Geosciences*, vol. 44, pp. 120–135, 2012.

[8] H.-J. Zimmermann, "Fuzzy set theory," *Advanced Review*, vol. 2, pp. 317–332, 2010.

[9] B. Quost, M-H. Masson, and T. Denoeux, "Classifier fusion in the Dempster-Shafer framework using optimized t-norm based combination rules," *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 353–374, 2011.

[10] M. Grabisch, "Fuzzy pattern recognition by fuzzy integrals and fuzzy rules," *Pattern recognition - From classical to modern approaches*, pp. 357–280, 2002.

[11] C. Labreuche and M. Grabisch, "Generalized Choquet-like aggregation functions for handling bipolar scales," *European Journal of Operational research*, vol. 172, no. 3, pp. 931–955, 2006.

[12] W. Dou, S. Ruan, Y. Chen, D. Bloyet, and J.-M. Constans, "A framework of fuzzy information fusion for the segmentation of brain tumor," *Image and Vision Computing*, vol. 25, no. 2, pp. 164–171, 2007.

[13] S. Ruan, B. Moretti, J. Fadili, and D. Bloyet, "Fuzzy Markovian Segmentation in Application of Magnetic Resonance Images," *Computer Vision and Image Understanding*, vol. 85, no. 1, pp. 54–69, 2002.

[14] J. Lapuyade-Lahorgue, D. Visvikis, and M. Hatt, "Multi-class Multimodal SPEQTACLE: a new fuzzy clustering algorithm for fully automatic delineation of tumors in multimodality imaging," IEEE Nuclear Science Symposium and Medical Imaging Conference, 2014.

[15] P. Lanchantin and W. Pieczynski, "Unsupervised Restoration of hidden non-stationary Markov chains using Evidential priors," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3091–3098, 2005.

[16] J. Yen, "Generalizing the Dempster-Shafer theory to fuzzy sets," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 3, pp. 559–570, 1990.

[17] B. Waske and J. A. Benediktsson, "Fusion of Support Vector Machines for classification of multisensor data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3858–3866, 2007.

[18] A. Smolinska, L. Blanchet, and L. Coulier et al., "Interpretation and visualisation of non-linear data fusion in Kernel space: study on metabolomic characterization of progression of multiple sclerosis," 10th Forum of Neuroscience, 2012.

[19] C. X. Cheng, N. Y. Dan, and W. Ying, "The application of data fusion technology based on neural network in the dynamic risk assessment," *Physics Procedia*, vol. 25, pp. 1696–1700, 2012.

[20] Y. Meng, S. Yan, Z. Tang, Y. Chen, and J. Liu, "Data fusion based on neural networks and particle swarm algorithm and its application in sugar boiling," *Advances in Neural Networks*, vol. 5263, pp. 176–185, 2008.

[21] D. Norris, J. M. McQueen, and A. Cutler, "Prediction, Bayesian inference and feedback in speech recognition," *Language, Cognition and Neuroscience*, vol. 31, no. 1, pp. 4–18, 2016.

[22] N. J.-B. Brunel, J. Lapuyade-Lahorgue, and W. Pieczynski, "Modeling and unsupervised classification of multivariate hidden Markov chains with copulas," *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 338–349, 2010.

[23] E. Vincent and X. Rodet, "Music transcription with ISA and HMM," 5th International Symposium on ICA and BSS, 2004.

[24] C. Raphael, "Automatic segmentation of acoustic musical signals using hidden Markov models," *IEEE Transactions on Pattern Analysis*, vol. 4, pp. 360–370, 1999.

[25] T. Koski, *Hidden Markov models for bioinformatics*, Kluwer Academic Publishers, 2001.

[26] P. Nicolas, L. Bize, F. Muri-Majoube, M. Hoebeke, F. Rodolphe, S. Dusko Ehrlich, B. Prum, and P. Bessire, "Mining Bacillus subtilis chromosome heterogeneities using hidden Markov models," *Nucleic Acid Research*, vol. 30, pp. 1418–1426, 2002.

[27] E. Bellone, J.P. Hughes, and P. Guttorp, "A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts," *Climate Research*, vol. 15, no. 1, pp. 1–12, 2000.

[28] D. Wu, L. Pigou, P.-J. Kindermans, N. D.-H. Le, and L. Shao, "Deep Dynamic Neural Networks for Multi-modal Gesture Segmentation and Recognition," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, 2016.

[29] Y. Wang, S. M. Resnick, and C. Davatzikos, "Spatio-temporal analysis of brain MRI images using hidden Markov models," *MICCAI 2010*, vol. 13, pp. 160–168, 2010.

[30] G. Welch and G. Bishop, "An introduction to the Kalman filter," Tech. Rep., Dpt of Computer Science, University of North Carolina, 1995.

[31] C. S. Magnano, A. Soni, S. Natarajan, and G. Kunapuli, "Conditional Random Fields for brain tissue segmentation," SDM 2014, 2014.

[32] C. Solares and A. M. Sanz, "Bayesian Network Classifiers. An application to Remote Sensing image classification," Int. Conf. on Neural Networks, 2005.

[33] B. H. Menze, K. V. Leemput, D. Lashkari, and et. al, "A generative probabilistic model and discriminative extensions for brain lesion segmentation – with application to tumor and stroke," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 933–946, 2015.

[34] P. A. Devijver, "Baum's forward-backward algorithm revisited," *Pattern recognition*, vol. 3, no. 6, pp. 369–373, 1985.

[35] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, 1989.

[36] S. J. Julier, "A new extension of the Kalman filter to nonlinear systems," International Symposium on Aerospace/Defense Sensing, Simulation and Controls, 1997.

[37] B. A. Berg, "Markov Chain Monte Carlo simulations and their statistical analysis," World Scientific, 2004.

[38] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," IEE Proceedings, 1993.

[39] G. Casella and E. I. George, "Explaining the Gibbs sampler," *The American Statistician*, vol. 46, no. 3, pp. 167–174, 1992.

[40] C. Kharoubi-Rakotomalala and F. Maurer, "Copulas in Finance: Ten years later," *The Journal of Applied Business Research*, vol. 29, no. 5, pp. 1555–1566, 2013.

[41] A. Voisin, V. Krylov, G. Moser, S. Serpico, and J. Zerubia, "Multichannel hierarchical image classification using multivariate copulas," IS&T/SPIE Electronic Imaging, 2012.

[42] F. Flitti, C. Collet, and E. Slezak, "Wavelet domain astronomical multiband image fusion and restoration using Markov quadtree and copulas," European Signal Processing Conference, 2005.

[43] S. Derrode and W. Pieczynski, "Unsupervised data classification using Pairwise Markov chains with automatic copulas selection," *Computational Statistics and Data Analysis*, vol. 63, pp. 81–98, 2013.

[44] F. Flitti, C. Collet, and A. Joannic-Chardin, "Unsupervised multiband image segmentation using Hidden Markov quadtree and copulas," ICIP 2005, 2005.

[45] Y. Stitou, N.-E. Lasmar, and Y. Berthoumieu, "Copulas based multivariate Gamma modeling for texture classification," ICASSP 2009, 2009.

[46] N-E. Lasmar and Y. Berthoumieu, "Gaussian copula for multivariate modeling for image texture retrieval using wavelet transforms," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 933–946, 2015.

[47] S. Chib and E. Greenberg, "Understanding the Metropolis-hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[48] B. Schweizer and A. Sklar, "Probabilistic metric spaces," *Elsevier Science Publishing Company*, 1983.

[49] B. H. Menze and et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[50] R. B. Nelsen, *An introduction to copulas*, Lecture notes in Statistics, 1998.

[51] S. Geman and D. Geman, "Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 6, 1984.

[52] S. Bauer, T. Fejes, J. Slotboom, and et al, "Segmentation of brain tumor images based on integrated hierarchical classification and regularization," MICCAI 2012, 2012.

[53] E. Geremia, O. Clatz, and B. H. Menze et al, "Spatial decision forests for MS lesion segmentation in multichannel magnetic resonance images," *Neuroimage*, vol. 57, pp. 378–390, 2011.

[54] H.-C. Shin, "Hybrid clustering and logistic regression for multi-modal brain tumor segmentation," MICCAI 2012, 2012.

[55] D. Zikic, B. Glocker, E. Konukoglu, and et al, "Context-sensitive classification forests for segmentation of brain tumors tissues," MICCAI 2012, 2012.

[56] N. K. Subbanna, D. Precup, D. L. Collins, and T. Arbel, "Hierarchical probabilistic Gabor and MRF segmentation of brain tumors in MRI volumes," *Medical Image Computing and Computer-Assisted Intervention*, vol. 16, no. 1, pp. 751–758, 2013.

[57] A. Hamamci, N. Kucuk, K. Karaman, and et al, "Segmentation of brain tumors on contrast enhanced MR images for radiosurgery applications," *IEEE Transactions on Medical Imaging*, vol. 31, no. 3, pp. 790–804, 2012.