Robust Estimation of Unbalanced Mixture Models on Samples with Outliers

Alfiia Galimzianova, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin

Abstract—Mixture models are often used to compactly represent samples from heterogeneous sources. However, in real world, the samples generally contain an unknown fraction of outliers and the sources generate different or unbalanced numbers of observations. Such unbalanced and contaminated samples may, for instance, be obtained by high density data sensors such as imaging devices. Estimation of unbalanced mixture models from samples with outliers requires robust estimation methods. In this paper, we propose a novel robust mixture estimator incorporating trimming of the outliers based on component-wise confidence level ordering of observations. The proposed method is validated and compared to the state-of-the-art FAST-TLE method on two data sets, one consisting of synthetic samples with a varying fraction of outliers and a varying balance between mixture weights, while the other data set contained structural magnetic resonance images of the brain with tumors of varying volumes. The results on both data sets clearly indicate that the proposed method is capable to robustly estimate unbalanced mixtures over a broad range of outlier fractions. As such, it is applicable to real-world samples, in which the outlier fraction cannot be estimated in advance.

Index Terms—Mixture model, robust estimation, trimmed likelihood estimation, outlier detection, expectation-maximization, magnetic resonance imaging (MRI), brain structure segmentation

1 Introduction

Mixture models are widely used in pattern recognition, computer vision, medical image analysis, etc. [1], because they compactly and efficiently model samples from heterogeneous sources. The components of the mixture model carry information about the sources of observations. To extract this information from the sample, the mixture model component parameters and mixture weights have to be estimated. Most often the mixture model parameters are estimated by maximum likelihood estimators (MLEs), because these generally have several desirable properties such as consistency and efficiency [2]. For fitting mixture models by MLEs, the expectation-maximization (EM) algorithm [3] is usually the preferred technique. Although MLE-EM methods are known for their high convergence rate, the correct convergence cannot be guaranteed especially when sample sources generate different or unbalanced numbers of observations, which, besides, are mixed to some extent [4]. The problem of estimating unbalanced mixtures on samples without outliers was addressed by deterministic annealing MLE-EM [5].

However, real-world samples not only contain unbalanced and overlapping observations, but are generally contaminated by outlying observations or *outliers*, i.e., noisy and erroneous observations that do not conform the theoretical mixture model of heterogeneous sources. The outliers may adversely affect the estimation of mixture parameters and are very critical in the MLE, since even a single outlier can lead to a degenerate mixture estimate, which does not provide information about the distribution of inlying observations [6], [7]. The fraction of outliers can vary significantly between different samples and it is often difficult, if at all possible, to determine the fraction of outliers in advance. In samples drawn from unbalanced mixtures with outliers, robust estimation of parameters of components represented by a small number of observations thus requires a careful selection of corresponding inlying observations.

1.1 Previous Work

The degree of contamination by outliers in a sample is usually expressed by outlier fraction h, i.e., the number of outliers versus sample size N, while the robustness of an estimator is measured by its breakdown point (BDP), defined as the smallest fraction of outliers that can cause degenerate mixture estimates. Neykov and Müller [6] showed that MLE has BDP of zero. To overcome this deficiency of MLE, several estimators based on MLE were developed that trade some of its efficiency for robustness to outliers, while still providing consistent estimates of mixture model parameters [2], [8], [9], [10]. There are three main approaches to robust mixture parameter estimation: 1) capturing outliers in a separate mixture component [11], [12], 2) capturing outliers by using heavy-tailed component models (e.g., t-distributions and skew-symmetric distributions) [8], [13], [14], [15] or 3) trimming to discard the outliers [2], [6], [9], [16], [17]. Modeling the outliers in the first two approaches is difficult on samples, for which the amount and distribution of the outliers are highly unpredictable. Moreover, as the outliers lying arbitrarily far from the mixture model may cause degenerate estimates, the BPD of methods incorporating heavy-tailed distributions is the same as when using the normal distribution, thus these

• *The authors are with the Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia. E-mail: {alfiia.galimzianova, franjo.pernus, bostjan.likar, ziga.spiclin}@fe.uni-lj.si.*

Manuscript received 25 Apr. 2014; revised 19 Dec. 2014; accepted 5 Feb. 2015. Date of publication 18 Feb. 2015; date of current version 7 Oct. 2015. Recommended for acceptance by A.J. Storkey. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TPAMI.2015.2404835

methods may also benefit from trimming [15]. We focus on the trimming approach, which can achieve a BDP of up to 50 percent [9].

The central idea in trimming is to detect the outliers and to discard them from the likelihood function. If the current estimate of mixture model is close to optimal, the outliers are likely to correspond to observations with low likelihood, therefore, Hadi and Luceño [2] proposed to trim the observations with low contribution to the likelihood function. Their maximum trimmed likelihood estimator (MTLE) proceeds by drawing all possible subsamples of size M from the given sample of size N, i.e., a total of $\binom{N}{M}$ samples, and fitting to each sample the mixture model by MLE. Final mixture parameters are given by the sample with highest trimmed likelihood. Even though MTLE can achieve a BDP of up to 50 percent, the combinatorial sampling renders it is less efficient for high BDP and, especially, for applications involving large samples. For large samples, Neykov et al. [6], [9] developed fast trimmed likelihood estimator (FAST-TLE), an efficient approximation to MTLE, based on iterative resampling techniques proposed in [18]. The basic idea is to take a predefined, finite number of random samples of size $M^* < N$, fit to each the initial mixture model by MLE and then for each sample 1) find new samples based on ordering all N observations according to their contributions to the likelihood function and trimming to sample size $M > M^*$, and 2) improve the fit by MLE on new samples. Steps 1 and 2 are iterated until convergence and the final mixture parameters are given by the sample corresponding to highest trimmed likelihood. For normal mixture models, similar ideas were adopted in robust model based clustering methods such as EMT [17] and TCLUST [19], however, FAST-TLE has the advantage as it can be used with various models of component distributions and with any MLE.

A crucial parameter of trimming approach is the trimming fraction, defined as $\alpha = (N - M)/N$. Trimming fraction should be set equal or, to increase the BDP margin, even higher than the expected amount of outliers in the contaminated sample ($\alpha \geq h$). In some applications, trimming fraction can be determined by tuning the value of α to achieve best performance on given training samples [20]. However, this is often not possible and the trimming fraction should be set to a high, marginal value of expected outlier fraction [21].

Our goal is to obtain robust estimates on large samples of highly unbalanced mixtures, which contain a high and unpredictable amount of outliers. The main challenge of parameter estimation on samples of unbalanced mixtures is to prevent trimming of those observations belonging to the component(s) with a low number of observations, since in such situation the estimation method will not be able to recover the optimal parameters of the particular component. Besides, if a high fraction of outliers is expected in a sample, then also a high fraction of observations must be trimmed, therefore, a careful selection of observations to be trimmed is very critical.

The existing state-of-the-art robust methods, however, have not proved adequate for robust estimation of unbalanced mixtures. For example, the FAST-TLE method uses the likelihood based ordering of the observations that assigns a lower rank to observations belonging to components represented by smaller number of observations and components with larger scales, which is especially critical when the trimming fraction is set much higher than the actual outlier fraction. Therefore, the estimation of unbalanced mixtures and mixtures with unbalanced scales in the presence of outliers may not be robust by the FAST-TLE method.

1.2 Contribution

Based on our previous work [22], we propose a novel method for estimation of mixtures on samples contaminated with outliers that is robust to the unknown fraction of outliers. The proposed method is robust even on highly unbalanced mixtures and efficient on large samples. The robustness is achieved by selecting the outliers based on component-wise confidence level ordering of observations, which also enables the estimation of highly unbalanced mixtures. The computational efficiency is achieved by two computational approximations of the ordering of observations. Both approximations provide a compact representation of the sample space and enable the application of confidence level ordering to any model of mixture components. The proposed method was extensively validated and compared to the state-of-the-art FAST-TLE method [9] in two experiments. For the first experiment, synthetic samples were generated with known fractions of outliers and with known unbalanced mixture weights so as to evaluate the robustness of the methods. In the second experiment, realistic samples from multisequence structural MR brain images that contained tumors of varying volumes were used to test the two methods for robust estimation of structural intensity model and segmentation of the normal brain structures. The results indicate that the proposed method is robust over a broad range of trimming fractions regardless of the actual outlier fraction. This characteristic is very important as it makes the method applicable to a wide range of mixture modeling applications.

The paper is organized as follows. Section 2 describes the proposed method. Extensive evaluation of the proposed and FAST-TLE methods using synthetic samples is presented Section 3, while Section 4 presents a practical application of the methods for segmenting normal structures in the brain MR images, which also contain large pathological structures. Analysis of performance is reported in Section 5, while Discussion is given in Section 6.

2 Methods

2.1 Maximum Likelihood Estimators

Let $X = \{x_1, x_2, \ldots x_N\}, x_j \in \mathbb{R}^d$ represent a random sample of N i.i.d. observations drawn from an unobservable d-variate multimodal distribution $\tilde{\psi}$. By assuming $\tilde{\psi}$ has K distinct modes, each belonging to a known family of parametric unimodal distribution models, the distribution $\tilde{\psi}$ can be represented as a K-component mixture [1]:

$$\tilde{\psi} \approx \psi(x|\Theta) = \sum_{k=1}^{K} \pi_k \, p(x|\theta_k), \tag{1}$$

where $\Theta = \{\pi_k, \theta_k\}_{k=1}^{K}$ denotes a set of unknown mixture parameters, with mixing weights π_k that obey $\pi_k > 0, \forall k$

and $\sum_{k=1}^{K} \pi_k = 1$, and parameters $\theta_k \subset \mathbb{R}^d$ of d-variate distribution model $p(x|\theta_k)$ of the kth component. In normal mixtures, the parameters θ_k represent the location μ_k and scale Σ_k of the kth component. The unknown mixture parameters Θ can be found by maximizing the likelihood function $l(\Theta|X) = \prod_{x \in X} \psi(x|\Theta)$ of sample X or, equivalently, by maximizing the corresponding log-likelihood function

$$L(\Theta|X) = \sum_{x \in X} \log(\psi(x|\Theta)). \tag{2}$$

If the sample X is contaminated by outliers, maximizing the log-likelihood (2) will lead to biased estimates of Θ [6]. An unbiased estimator is obtained by trimming the outliers such that the log-likelihood is computed on a subsample X_H, i.e., $L(\Theta|X_H)$, which does not contain any outliers [2], [9], [23]. Subsample X_H of size $H = [N(1-\alpha)]$ is obtained by trimming a fraction α of all the observations, where α should be set higher or equal to the expected outlier fraction h ($\alpha \geq h$). Selection of these H observations is a critical step that requires a specific ordering of observations, based on which the inlying observations are more likely to be selected into subsample X_H than the outliers.

2.2 Ordering Observations for Likelihood Trimming

If mixture parameters Θ are initialized *close enough* to their optimal values, then the outliers can be selected based on their conformance with the current estimate of the mixture model $\psi(x|\Theta)$. Consider an ordering of sample indices $\overline{v}(X|\Theta) = (\overline{v}_1, \ldots, \overline{v}_N)$ based on monotonically non-increasing log-likelihood

$$\log(\psi(x_{\overline{v}_1}|\Theta)) \geq \cdots \geq \log(\psi(x_{\overline{v}_N}|\Theta)). \tag{3}$$

Based on the log-likelihood ordering $\overline{v}(X|\Theta)$ a trimmed subsample is obtained as $\overline{X}_H = \{x_{\overline{v}_j}\}_{j=1}^{H}$ and the trimmed log-likelihood function $L(\Theta|\overline{X}_H)$ is computed using (2). The log-likelihood ordering was used in FAST-TLE [6], [9] and TCLUST [19] methods. For unbalanced mixtures, however, the log-likelihood ordering is biased towards the components with higher mixture weights; for any pair of observations x_i and x_j belonging to kth and lth components ($x_i \in X_k, x_j \in X_l$), respectively, and the corresponding mixture weights π_k and π_l, where $\pi_k \gg \pi_l$ and $\pi_l \to 0$, the following ordering is expected $\log(\psi(x_i|\Theta)) > \log(\psi(x_j|\Theta))$. Therefore, the use of log-likelihood ordering and a high trimming fraction α on the unbalanced mixtures could easily lead to trimming of all the observations of a component with a small mixture weight. Besides, the log-likelihood ordering is biased towards components with small scales (for the location-scale component models), since the observations belonging to small scale components are assigned a higher rank as compared to the observations belonging to large scale components. The influence of unbalanced mixture weights and component scales on the log-likelihood ordering is shown in Fig. 1.

2.3 Confidence Level Ordering

The log-likelihood ordering cannot be used to obtain robust estimates of the parameters of unbalanced mixtures and

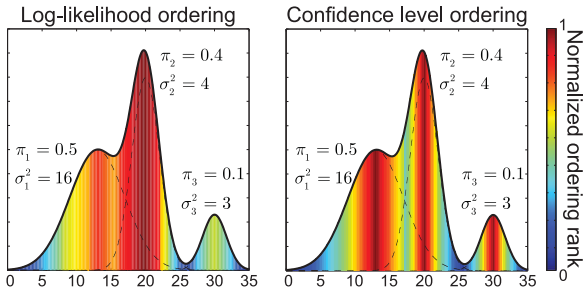


Fig. 1. An unbalanced three-component normal mixture model with heterogeneous scales and the corresponding color-coded normalized ordering ranks obtained by the log-likelihood (*left*) and the proposed confidence level ordering (*right*).

unbalanced scales, therefore, we propose a new ordering of observations $\tilde{v}(X|\Theta) = (\tilde{v}_1, \ldots, \tilde{v}_N)$ based on monotonically non-decreasing component-wise confidence levels of the observations

$$\int_{\Omega(x_{\tilde{v}_1})} p(\omega|\theta_{z_{\tilde{v}_1}})\, d\omega \leq \cdots \leq \int_{\Omega(x_{\tilde{v}_N})} p(\omega|\theta_{z_{\tilde{v}_N}})\, d\omega, \tag{4}$$

where $\Omega(x_j) = \{\omega \in \Omega : p(\omega|\theta_{z_j}) \geq p(x_j|\theta_{z_j})\}$ are corresponding confidence regions, $\Omega \subseteq \mathbb{R}^d$ is the sample space, and $z_j \in \{1, \ldots, K\}$ is a classification of observations to one of the K components. If the EM algorithm [3] is used to estimate the mixture parameters Θ, the observations $x_j \in X$ can be classified based on maximum posterior probability (MAP) as

$$z_j = \underset{k=1,\ldots,K}{\arg\max}\ \tau(x_j|\theta_k), \tag{5}$$

where $\tau(x_j|\theta_k) = \pi_k\, p(x_j|\theta_k) / \sum_{l=1}^{K} \pi_l\, p(x_j|\theta_l)$. In the normal mixtures, each observation $x_j \in X$ can be classified based on component-wise Mahalanobis distances $d_\Sigma^2(x_j|\,\mu_k, \Sigma_k) = (x_j - \mu_k)\,\Sigma_k^{-1}(x_j - \mu_k)^T$ as

$$z_j = \underset{k=1,\ldots,K}{\arg\min}\ d_\Sigma^2(x_j|\,\mu_k, \Sigma_k). \tag{6}$$

The confidence levels required for the ordering (4) can be computed from the Mahalanobis distances $d_\Sigma^2(x_j|\,\mu_{z_j}, \Sigma_{z_j})_{j=1,\ldots,N}$ as the value of cumulative density function of a χ_d^2 distribution. However, this approach may be computationally demanding for large samples as the Mahalanobis distances have to be computed K times for each of the observations. Besides, this approach can be used only in normal mixtures. The approximate confidence levels can be obtained by using the stochastic density quantile algorithm [24] or numerical integration of the sample space Ω, both of which are computationally efficient due to a compact representation of the sample space and both enable the use of confidence level ordering with any type of mixture distributions. Here we will compute the confidence levels by numerical integration (Algorithm 1).

Given the confidence level ordering $\tilde{v}(X|\Theta)$ and the corresponding subsample $\tilde{X}_H = \{x_{\tilde{v}_j}\}_{j=1}^{H}$ the objective of parameter estimation is to maximize the log-likelihood function $L(\Theta|\tilde{X}_H)$. In the following section we propose a robust method for maximizing the log-likelihood objective function based on confidence level ordering.

Algorithm 1. Computation of confidence levels based on numerical integration

Input: samples $X_k = \{x_j : z_j = k\}$, $j = 1, \ldots, N$, one for each mixture component $k = 1, \ldots, K$, integration precision level ξ
Numerical integration: for each mixture component k
1) Set integration limits as the bounding box of confidence level ellipse, which corresponds to the observation of the smallest value of the probability density function (PDF), i.e., $\min_{x_j \in X_k} p(x_j|\theta_k)$
2) Compute the confidence levels on a rectangular grid using the rectangle integration method and selectively increase the grid density to reach the integration precision level ξ
3) Assign the computed confidence levels to the corresponding observations in sample $X_k \subset X$

2.4 Likelihood Maximization Based on Confidence Level Ordering

Methods based on trimming usually start with multiple random initial guesses of mixture parameters Θ and then, for each initial guess, maximize the log-likelihood in a concentration process [6], [9], [18], [19]. However, performing multiple concentration processes is computationally too demanding for large samples and high trimming fractions α because a high number of the random initial guesses is required to obtain robust mixture estimates [2]. For the application of mixture models on large samples (e.g., in image analysis) a good initial guess of the mixture parameters can be obtained by machine learning techniques, prior knowledge, etc. [20].

The concentration process is generally a two-step iterative procedure. In the first step the observations are trimmed according to their ordering, while in the second step the log-likelihood function is maximized on the trimmed subsample X_H of size $H = [N(1-\alpha)]$. The log-likelihood should increase in each iteration of the concentration process, however, unlike the log-likelihood ordering $\overline{v}(X|\Theta)$, the proposed confidence level ordering $\tilde{v}(X|\Theta)$ does not guarantee the increase of the log-likelihood. We solve this by progressively trimming the ordered subsample $\tilde{X}_H$ to size $H^* < H$ in each iteration until the log-likelihood is higher than in the previous iteration, whereas the log-likelihood is increased only if the trimmed observations correspond to negative log-likelihoods. Trimming more observations is likely to remove outliers from the parameter update step, which is based on MLE and thus sensitive to outliers. Hence, additional trimming can improve convergence of the estimates. Since for any mixture parameter estimates Θ the log-likelihood (2) is maximal for a subsample $\overline{X}_H$ obtained by trimming based on the log-likelihood ordering $\overline{v}(X|\Theta)$ (3), the progressive trimming of $\tilde{X}_H$ is performed only until the following inequalities hold

$$L(\Theta|\overline{X}_H) \geq L(\Theta|\tilde{X}_{H^*}) > L(\Theta|\tilde{X}_H)). \tag{7}$$

Therefore, whenever $L(\Theta|\overline{X}_H)$ is bounded (i.e., the FAST-TLE), the proposed log-likelihood function $L(\Theta|\tilde{X}_H)$ is bounded as well. The proposed method for likelihood maximization based on confidence level ordering is given in Algorithm 2.

Algorithm 2. Method for likelihood maximization based on confidence level ordering

Input: trimming fraction α, the initial mixture parameters $\Theta^{(0)} = \{\pi_1^{(0)}, \ldots, \pi_K^{(0)}, \theta_1^{(0)}, \ldots, \theta_K^{(0)}\}$, and a log-likelihood termination threshold ϵ.
Concentration process: for each mixture component k
1) Given the mixture parameters $\Theta^{(i-1)}$, classify each observation $x_j \in X$ to one of K components according to (6) or (5).
2) Perform confidence level ordering $\tilde{v}(X|\Theta^{(i-1)})$ according to (4) and obtain a subsample $\tilde{X}_H^{(i)} = \{x_{\tilde{v}_1}, \ldots, x_{\tilde{v}_H}\}$ of size $H = [N(1-\alpha)]$.
3) Compute log-likelihood $L(\Theta^{(i-1)}|\tilde{X}_H^{(i)})$ on the subsample $\tilde{X}_H^{(i)}$ using (2). If in the first iteration ($i = 1$) or if the log-likelihood has increased over the previous iteration, i.e., $L(\Theta^{(i-1)}|\tilde{X}_H^{(i)}) > L(\Theta^{(i-1)}|\tilde{X}_H^{(i-1)})$, then continue to step 5.
4) Progressively trim the subsample $\tilde{X}_H^{(i)}$ by taking into account the lexicographic ordering of $\tilde{v}_j, j = 1, \ldots, H$, until the trimmed subsample $\tilde{X}_{H^*}^{(i)} = \{x_{\tilde{v}_1}, \ldots, x_{\tilde{v}_{H^*}}\}$ of size $H^* < H$ satisfies condition (7). If condition (7) is satisfied, set $\tilde{X}_H^{(i)} = \tilde{X}_{H^*}^{(i)}$ and continue to step 5. Otherwise, terminate iteration and return the previous mixture parameters $\Theta^{(i-1)}$.
5) Improve the mixture parameters $\Theta^{(i-1)}$ by maximizing the log-likelihood $\Theta^{(i)} = \arg\max_\Theta L(\Theta|\tilde{X}_H^{(i)})$, e.g., by using EM on the subsample $\tilde{X}_H^{(i)}$.
6) If the relative increase of the log-likelihood is below the termination threshold ϵ, i.e., $(L(\Theta^{(i)}|\tilde{X}_H^{(i)}) - L(\Theta^{(i-1)}|\tilde{X}_H^{(i-1)}))/L(\Theta^{(i-1)}|\tilde{X}_H^{(i-1)}) \leq \epsilon$, terminate iteration and return the mixture parameters $\Theta^{(i)}$. Otherwise, continue to step 1 and start iteration $i + 1$.

The key ingredient of the proposed method is the confidence level ordering (4), which preserves the inlying observations of all the mixture components even for samples of highly unbalanced mixtures and components of heterogeneous scales. For this reason, even for a trimming fraction α that is much higher than the expected fraction of the outliers h, the proposed method enables robust estimation of mixture parameters, the characteristic which we experimentally verify in the next sections.

3 Experiments on Synthetic Samples

The purpose of these experiments was to evaluate the performance of the proposed and FAST-TLE methods on synthetic samples that were generated by drawing observations from synthetic mixtures with known parameters. The balance between component weights was controlled and the synthetic samples were contaminated by a varying fraction of outliers.

3.1 Creation of Synthetic Samples with Outliers

The synthetic samples were composed of inlying and outlying observations, each of which were generated using a specific random sampling process. The inlying observations were obtained by randomly sampling from different synthetic mixtures, which consisted of three normal (Gaussian)

components in the two-dimensional (2D) real domain. Each synthetic mixture was defined by parameters $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{3}$, where π_k were the weights, μ_k the means (locations), and Σ_k the covariances (scales) of the normal components. Generation of a new synthetic mixture involved random sampling of the values of parameters π_k, μ_k and Σ_k from uniform distributions $U(a, b)$ in the range from a to b. The weights π_k were drawn from $U(0, 1)$ and are normalized to sum up to one ($\sum_{k=1}^{3} \pi_k = 1$). The coordinates of the means $\mu_k, k = 1, 2, 3$ were drawn from $U(-1, 1)$ and labeled such that their first coordinates were in ascending order, i.e., $\mu_{1,1} < \mu_{2,1} < \mu_{3,1}$. To ensure the observations belonging to the mixture components could be reliably distinguished from each other [1], we verified that the minimal distance between the first coordinates $\mu_{1,1}, \mu_{2,1}$ and $\mu_{3,1}$ was at least 1/K; otherwise, new coordinates for the means were drawn as long as this condition was not satisfied. The covariances Σ_k were generated as $\Sigma_k = SS^T \circ A$, which ensured Σ_k was positive-definite and where $S = S_{2\times 2} = \{s_{ij} : s_{ij} \sim U(-1, 1)\}$ was a random nonsingular matrix and $A = \frac{1}{\delta K}(I_{2\times 2} + 1)$ the scaling matrix and δ a positive constant set to 5. To test if the resulting synthetic mixture was well-separated, the overlap between the three mixture components was measured by Bayes error rate (BER) and the synthetic mixture was discarded if BER was higher than $BER_{max} = 0.05$. In order to avoid mixtures with badly scaled components, the lower limit to the covariance determinant was set to $0.5 \cdot 10^{-4}$. Finally, the means and covariances were rescaled as $\mu_k = \gamma\mu_k$ and $\Sigma_k = \gamma^2\Sigma_k$ using $\gamma = 10$ so that the observations x drawn from the synthetic mixtures lied in the 2D domain of $[-20, 20] \times [-20, 20]$. The obtained synthetic samples were contaminated by a varying fraction of outliers, the coordinates of which were drawn uniformly from $U(-20, 20)$ and, then, the observations within the 95 percent confidence area of any of the three components of the normal mixture were rejected. Because the outliers were clearly separable from the inliers, trimming should effectively remove these outliers. In this way, we can assess the impact of ordering schemes used for trimming and, thus, the robustness of the mixture estimation methods with respect to varying trimming fraction α and varying outlier fraction h.

3.2 Experiment Description

Experiment A. The FAST-TLE and the proposed methods were used to estimate the three-component normal mixtures from synthetic samples contaminated by an a priori unknown fraction of outliers h. The synthetic samples of $N = 10^4$ observations were drawn from the randomly generated normal mixtures and a varying number of outliers were added to each of these samples such that their fraction h varied from 0.0 to 0.5 with a step of 0.05. The two methods were tested with α from 0.0 to 0.5 in 0.05 intervals. For each tested value of h a set of 100 random mixtures was generated and, from each, three samples of size N were drawn and, then, the outliers with uniform distribution were added to the three samples. As there were 11 different values of outlier fraction h and 300 test samples were generated for each h, we obtained a total of 3,300 test samples. Experiments were also performed on synthetic mixtures with normally distributed outliers (see Supplemental materials, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2015.2404835, available online).

Experiment B. The FAST-TLE and the proposed methods were used to estimate the three-component normal mixtures with unbalanced component weights. The degree to which the mixture weights were unbalanced was represented by the value of the minimal weight. The value of the minimal weight was drawn from 11 intervals of equal size in the range $[0.01, 0.33]$ and, for each interval, 100 normal mixtures were generated. Three synthetic samples of $N = 10^4$ observations were drawn from each of these mixtures and each sample was added 10^3 of uniformly distributed outliers ($h = 0.1$). Overall, 3,300 test samples were obtained.

In *Experiments A* and *B*, the FAST-TLE and the proposed methods were executed on all test samples and, on each sample, both methods were initialized with the same values of the mixture parameters $\Theta^{(0)} = \{\pi_k^{(0)}, \mu_k^{(0)}, \Sigma_k^{(0)}\}_{k=1}^{3}$. The initial mean values $\mu_k^{(0)}$ were randomly selected among the inliers of the kth component as observations within the 95 percent confidence ellipses, while the initial covariance matrices $\Sigma_k^{(0)}$ were set to $0.3\, I_{2\times 2}$ and the initial mixture weights were set to $\pi_k^{(0)} = 1/K$. Both tested methods used the EM algorithm [3] to estimate the mixture parameters on the trimmed subsamples X_H and were executed with a maximum of 50 iterations. The final mixture parameters were used to classify the inlying observations based on MAP (5) and the performance of the two methods was evaluated by misclassification ratio (MCR), computed as the ratio between the number of incorrectly classified inlying observations and the total number of inlying observations N:

$$MCR = \frac{1}{N}\sum_{k=1}^{K}\left|\{x_j \in X : (x_j \in X_k) \wedge (z_j \neq k)\}\right|. \quad (8)$$

For any combination of h and α the reported MCR values were averaged over the three samples drawn from the corresponding normal mixture.

3.3 Results

The impact of the choice of trimming fraction α on the performance of the FAST-TLE and the proposed methods on one test sample with outlier fraction $h = 0.1$ used in the experiment is demonstrated in Fig. 2. The sample shown in Fig. 2 was composed of three normal components with slightly unbalanced component weights, i.e., $\pi_1 \approx 0.34$, $\pi_2 \approx 0.55$, $\pi_3 \approx 0.11$. When the trimming fraction α was lower than the outlier fraction h, trimming could not remove all the outliers from the original sample, thus, in both of the tested methods the MLE estimator employed on the trimmed sample produced degenerate mixture estimates. Fig. 2 shows that the mixture component with the smallest mixture weight (*rightmost*) was most affected, with over-estimated covariance and the mean shifted away from its true value. The reason was that some outliers were not trimmed and thus had an adverse influence on the EM-based mixture parameter estimation. For $\alpha = h$ the

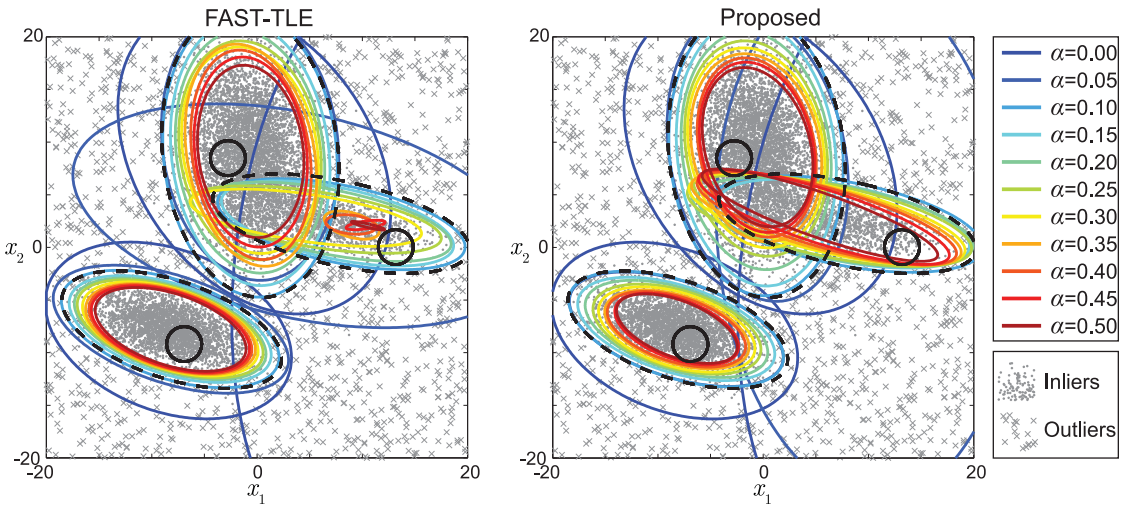


Fig. 2. Test sample consisted of inlying observations (*gray dots*) of three-component normal mixture and a fraction of $h = 0.1$ (10 percent) of uniformly distributed outliers (*gray crosses*). The ellipses show the 99 percent confidence regions of the mixture components for the initial (*black*) and true (*dashed black*) mixture parameters, while the colored ellipses show the 99 percent confidence regions of the estimated mixture components for different values of the trimming fraction α.

FAST-TLE method accurately estimated the mixture parameters, however, the estimates of the third component (the component with the smallest component weight $\pi_3 \approx 0.11$) became less accurate for $\alpha > 0.3$. The reason was that the log-likelihood ordering led to the trimming of most of the observations belonging to the third component and thus reduced the estimate of the covariance Σ_3. On the other hand, the proposed method accurately estimated all the mixture parameters for α in the range from 0.1 to 0.5. The advantage was obtained by the use of confidence level ordering (4), which preserved most of the inlying observations of all the components regardless of the component weight and regardless of the value of α. For the same test sample, Fig. 3 shows the values of MCR for different values of α that enables a parallel insight into the performance of the tested methods. Note that high MCR values correspond to false classification of inlying observations due to the degenerate values of the mixture parameters. For $\alpha = 0.1$ the MCR was the lowest for both methods and then slowly increased for the proposed method, while for the FAST-TLE method the MCR increase was considerably higher.

Experiment A. Table 1 reports median MCR (mMCR) values obtained after executing the two mixture parameter estimation methods on all of the test samples with uniformly distributed outliers. The results are presented for six intervals of increasing outlier fraction (i.e., $[0.0, 0.0]$, $[0.0, 0.1]$, $[0.0, 0.2]$, $[0.0, 0.3]$, $[0.0, 0.4]$, $[0.0, 0.5]$). The mMCR was computed separately for samples, in which $\alpha = h_{max}$, $\alpha > h_{max}$ and $\alpha < h_{max}$ and where h_{max} was the maximal outlier fraction over all of the samples in the considered outlier fraction interval. The proposed method outperformed the FAST-TLE method and was less affected at high and highly variable outlier fractions (e.g., interval $[0.0, 0.5]$). Fig. 4 shows the mMCR values for each combination of h and α. Both the FAST-TLE and the proposed methods performed poorly (high mMCR) if the trimming fraction α was lower than the outlier fraction h (i.e., values below diagonal in Fig. 4). The FAST-TLE performed best (i.e., lowest mMCR) when the trimming fraction was similar to the true outlier fraction ($\alpha \approx h$, i.e., values on the diagonal in Fig. 4), while by increasing α beyond the value of h the mMCR progressively increased. The proposed method, however, showed a good and stable performance also for the values $\alpha > h$ (i.e., values above the diagonal in Fig. 4). Similar results were obtained on test samples with normally distributed outliers (see Supplemental materials, available online).

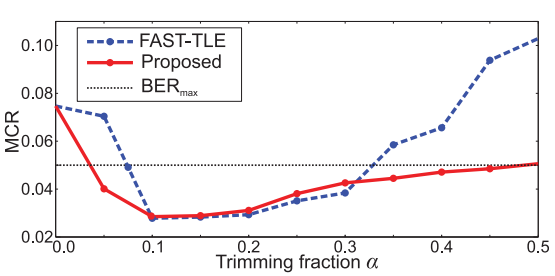


Fig. 3. Plots of misclassification ratio (MCR) for the mixtures estimated by the FAST-TLE (*dashed blue*) and the proposed methods (*red*) that at different trimming fractions α on the test samples in Fig. 2 with uniformly distributed outliers.

Experiment B. Table 2 reports the mMCR and median absolute deviations (MAD) on five intervals of the minimal weight computed over all trimming fraction values considered, $\alpha \in \{0.00, 0.05, \ldots, 0.50\}$. The proposed method outperformed the FAST-TLE method in each of the tested intervals, which corresponded to varying balance between mixture weights. Fig. 5 presents the mMCR values obtained by FAST-TLE and the proposed methods on all 11 intervals by increasing the minimal weight value. Fig. 5 confirms the hypothesis that the robustness of FAST-TLE depends drastically on balance between mixture weights, since at high values of trimming fraction α the inliers of the component with minimal weight are trimmed and, therefore, the components' parameters cannot be recovered. However, the proposed method successfully addressed this problem and gave stable results for trimming fractions higher than 0.1, which was the outlier fraction used in this experiment. The proposed method performed slightly worse in cases when the minimal mixture weight was much lower than the outlier fraction ($\pi_{min} \leq 0.05$), while at the same time the

TABLE 1
Median Misclassification Ratio (mMCR, $\times 10^{-2}$) and Median Absolute Deviation (*in parentheses*) Computed for the Mixtures Estimated by the FAST-TLE and the Proposed Methods on Test Samples with Uniformly Distributed Outliers for Six Different Intervals of the Outlier Fraction

Trimming fraction α	Method	Interval of the outlier fraction h, $[0, h_{max}]$					
		$[0.0, 0.0]$	$[0.0, 0.1]$	$[0.0, 0.2]$	$[0.0, 0.3]$	$[0.0, 0.4]$	$[0.0, 0.5]$
$\alpha = h_{max}$	FAST-TLE	2.2 (1.2)	2.7 (1.6)	3.5 (2.1)	4.3 (3.2)	6.1 (4.9)	11.3 (8.9)
	Proposed	2.2 (1.2)	**2.2** (**1.4**)	**2.4** (**1.4**)	**2.6** (**1.6**)	**2.8** (**1.7**)	**3.2** (**1.9**)
$\alpha > h_{max}$	FAST-TLE	2.8 (1.4)	3.5 (2.0)	4.3 (3.1)	5.8 (4.7)	9.8 (7.9)	–
	Proposed	**2.2** (**1.2**)	**2.2** (**1.4**)	**2.4** (**1.5**)	**2.7** (**1.6**)	**3.1** (**1.8**)	–
$\alpha < h_{max}$	FAST-TLE	–	2.7 (1.6)	3.3 (1.9)	3.9 (2.7)	4.9 (3.7)	7.6 (6.3)
	Proposed	–	**2.6** (**1.4**)	**2.7** (**1.6**)	**2.8** (**1.7**)	**3.0** (**1.8**)	**3.4** (**2.1**)

The numbers in bold correspond to the best performing method.

trimming fraction was set very high ($\alpha \geq 0.4$). The reason is that, depending on the initialization, the outliers uniformly distributed around the inliers of the minimal component will have a significant effect during the first few iterations.

4 Experiments on Brain MR Images

The analysis of brain MR images is typically approached by modeling the intensity distribution of brain structures, which can be performed through the estimation of a corresponding mixture model and which is then used for the task of brain structure segmentation, intensity nonuniformity correction, partial volume estimation, pathology detection, etc. [20], [25], [26], [27]. If the brain structures contain pathologies, then the modeling of the intensity distribution of these structures requires the use of robust mixture estimation methods, since the MR intensities corresponding to the pathological structures do not conform to the intensity distribution of the normal structures. The intensities of pathological structures such as tumor and brain lesions thus represent outliers in the intensity distribution. The fraction of outliers is directly proportional to the volume of the pathological structures that can vary significantly compared to the volume of the normal brain structures, therefore, the mixture estimation methods need to be robust to high variations of the outlier fraction. To evaluate the performance of the FAST-TLE and the proposed methods for the purpose of modeling the intensity distribution of brain structures, we created a database of MR images of brains that contained tumors of varying volumes. The mixture models estimated by the two tested methods were used to segment the normal brain structures such as the white matter (WM) and gray matter (GM) tissues, and cerebrospinal fluid (CSF), and the tumors.

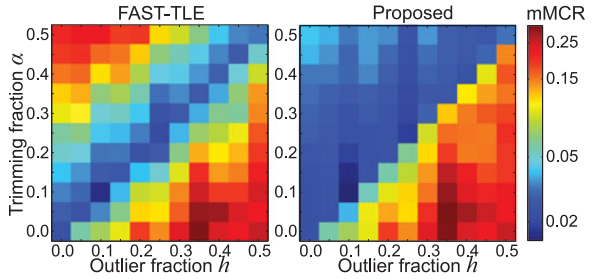


Fig. 4. Median misclassification ratio (mMCR) on test samples with uniformly distributed outliers for the mixtures estimated by the FAST-TLE (*left*) and the proposed methods (*right*).

4.1 Creation of Brain MR Image Database

A database[1] of 100 sets of MR images of brains with tumors of varying volumes was created from 20 different brain phantoms in the BrainWeb database [28]. From each of the 20 brain phantoms five sets of MR images with tumors of different volumes were created, resulting in a total of 100 sets of MR images. Tumors were generated by a realistic simulator of 3D brain tumor growth (TumorSim) [29] such that an initial tumor seed was placed at a random location inside the WM or GM tissue. The output of the TumorSim simulator that we used consisted of modified T1-, T2-weighted and FLAIR MR sequences and a modified ground truth segmentation of the brain structures and the tumor (Fig. 6). These three MR sequences are typically used for the visualization of the brain structures and for the diagnosis of the brain pathologies, since the T1w sequence generally has the highest contrast between the normal brain tissues (WM, GM), the T2w sequence highlights fluids (CSF) and inflammations, while FLAIR is complementary to T2w sequence, but suppresses the fluid-related signals. All the MR images had $256 \times 256 \times 128$ voxels with isotropic resolution $1 \times 1 \times 1$ mm^3 and contained MR-specific artifacts, such as Rician noise, partial volume and intensity inhomogeneity.

The obtained five sets of MR images per each brain phantom corresponded to five initial tumor seeds of increasing volumes and, since the volumes of the simulated tumors were proportional to the volume of the initial tumor seed, the final tumor volumes were different with mean volumes over all the 20 phantoms of 1.6, 10.6, 27.8, 44.9 and 66.0 cm^3 (Fig. 6).

4.2 Experiment Description

The aim of this experiment was to investigate the performance of the FAST-TLE and the proposed methods for the task of modeling the intensity distribution of normal brain structures in the presence of a varying amount of pathological structures. The methods were tested on 100 sets of MR images with varying tumor volumes by modeling the MR intensity distribution as a three-component normal mixture with the components corresponding to the WM, GM and CSF structures. The estimation of the normal mixture was performed on the intensities of T1w, T2w, and FLAIR

1. The code and the brain tumor MR image database are publicly available online at http://lit.fe.uni-lj.si/tools

TABLE 2
Median Misclassification Ratio (mMCR, $\times 10^{-2}$) and Median Absolute Deviation (*in parentheses*) Computed for the Mixtures Estimated by the FAST-TLE and the Proposed Methods on Test Samples with Uniformly Distributed Outliers for Six Different Intervals of the Outlier Fraction

Method	Interval of the minimal component weight $[0, \pi_{min}]$				
	$[0.01, 0.07]$	$[0.07, 0.14]$	$[0.14, 0.20]$	$[0.20, 0.27]$	$[0.27, 0.33]$
FAST-TLE	15.6 (12.9)	6.5 (4.9)	4.2 (3.0)	3.5 (2.1)	3.3 (1.6)
Proposed	**3.2** (**2.8**)	**2.7** (**1.7**)	**2.8** (**1.6**)	**2.6** (**1.5**)	**2.8** (**1.4**)

The numbers in bold correspond to the best performing method.

sequences in those voxels, which lied within the brain mask as given by the ground truth segmentation. The number of voxels captured within the brain mask varied between different phantoms and was $2 \cdot 10^6$ on average. The FAST-TLE and the proposed method were tested with different values of the trimming fraction α that varied from 0.0 to 0.5 with step 0.05. Mixture parameters were initialized based on a coarse segmentation of T1w sequence into the WM, GM and CSF structures, obtained by unsupervised dual-threshold Otsu's algorithm [30]. The initial means μ_k and covariances Σ_k of mixture components were set to corresponding sample means and sample covariances, and π_k were set to the fraction of voxels corresponding to each structure. Both of the two tested methods used the EM algorithm [3] on the trimmed subsample to estimate the mixture parameters and were executed with a maximum of 50 iterations.

The estimated mixture models were used to segment the brain MR images into WM, GM and CSF structures based on MAP (5), while the samples of trimmed observations, or outliers, $\{X \setminus X_H\}$, as obtained by the FAST-TLE and the proposed method, were further analyzed to segment the tumor. Besides the voxels corresponding to the tumor intensities, the samples of outliers may contain voxels with other atypical intensity values that deviate from the normal intensity model of brain structures due to partial volume, Rician noise [31] , intensity inhomogeneity [32], etc. Therefore, the voxels in each sample of outliers were classified as tumor or non-tumor voxels based on a heuristic hyper-intensity rule [20]. A particular voxel j was put into sample X_T (the sample of tumor voxels) if its T2w and FLAIR intensities $x_{j,T2w}$ and $x_{j,FLAIR}$, respectively, deviated from the corresponding intensity distributions of the WM voxels. At a given significance levels p_{T2w} and p_{FLAIR}, the corresponding intensity thresholds t_{T2w} and t_{FLAIR} were computed from $p_s = \int_{t_s}^{\infty} N(\mu_{WM,s}, \sigma_{WM,s})dt$, $s \in \{T2w, FLAIR\}$, i.e., the Gaussian intensity distribution of WM voxels, and were used to classify the tumor voxels as

$$X_T = \{x_j \in X : (x_{j,T2w} > t_{T2w} \wedge x_{j,FLAIR} > t_{FLAIR}) \wedge z_j \notin \{WM, GM, CSF\}\}. \quad (9)$$

The intensity thresholds were obtained at the corresponding significance levels p_{T2w} and p_{FLAIR} set to 10^{-4}. The outlier voxels that were not classified as tumors were subsequently classified as WM, GM or CSF based on the MAP and merged with their corresponding segmentations. Based on the ground truth segmentations, the accuracy of the estimated normal mixtures was evaluated as in Section 3 by computing the median misclassification ratio (mMCR) (8) over 20 phantoms for each of the five mean tumor volumes and for each setting of the trimming fraction α. The overlap between the obtained and ground truth segmentations was evaluated by Dice similarity coefficient (DSC):

$$DSC_k = \frac{2\,|Z_k \cap Z_k^{gt}|}{|Z_k| + |Z_k^{gt}|}. \quad (10)$$

The DSC was computed for each of the three normal brain structures $k \in \{WM, GM, CSF\}$ between the obtained Z_k and the ground truth segmentations Z_k^{gt}. The median of DSC (mDSC) was computed over 20 phantoms for each of the five mean tumor volumes and for each setting of the trimming fraction α.

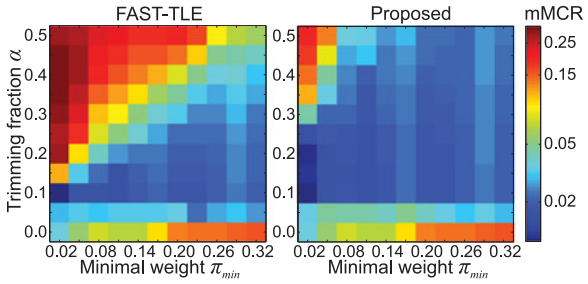


Fig. 5. Median misclassification ratio (mMCR) on test samples of synthetic mixtures with different values of the minimal mixture weight. Mixtures were estimated by the FAST-TLE (*left*) and the proposed methods (*right*) for different trimming fractions α.

4.3 Results

Table 3 reports the mMCR for the normal mixtures computed by the standard maximum likelihood estimator (MLE), the FAST-TLE and the proposed methods. The MLE used the EM algorithm and was equivalent to the FAST-TLE and the proposed methods with the trimming fraction α set to 0. The mMCR was generally higher for the MLE method compared to the other two methods that employed trimming, indicating the advantage of using robust mixture estimation methods. Nevertheless, the mMCR for the FAST-TLE method varied significantly for $\alpha > 0.3$ and was sometimes even higher than the mMCR of the MLE method, especially in the MR image sets with small tumors. On the other hand, the performance of the proposed method was not affected at high trimming fractions α and the method achieved the lowest mean mMCR. The left of Fig. 7 shows the mMCR values for each combination of mean tumor volume and α. The observed performance is consistent with the performance observed on the synthetic samples (Section 3), in which FAST-TLE achieved the lowest mMCR when the trimming fraction α was close to the true outlier

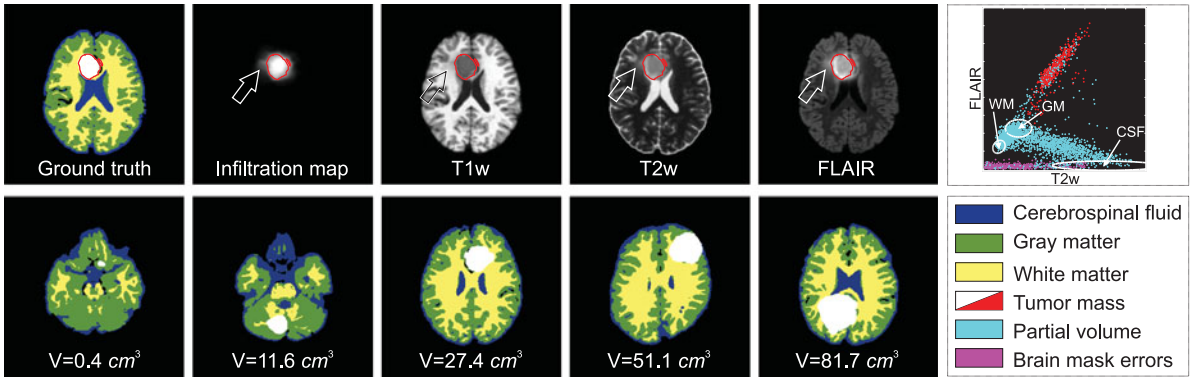


Fig. 6. *Top row from left to right*: A set of MR images, i.e., ground truth segmentation of the brain structures and tumor, tumor infiltration map, T1-, T2-weighted and FLAIR sequences with delineated tumor (*red line*), T2w–FLAIR distribution of outliers with superimposed ellipses of mixture components based on ground truth. The tumor also affected the MR intensities outside its boundary (*arrows*). *Bottom row from left to right*: Axial cross-sections of ground truth segmentations of the brain MR images with increasing tumor volume. The cross-sections are shown for different brain phantoms and each of the axial cross-sections was taken at the largest axial cross-sectional area of the tumor.

TABLE 3
Median Misclassification Ratio (mMCR, $\times 10^{-1}$) and Its Median Absolute Deviation (*in Parentheses*) for the Three Component Normal Mixtures Estimated on the MR Image Sets by the FAST-TLE and the Proposed Methods for Different Values of Trimming Fractions α

Trimming fraction α	Method	Median MCR (mMCR) for different mean tumor volumes					Overall mMCR
		$1.6cm^3$	$10.6cm^3$	$27.8cm^3$	$44.9cm^3$	$66.0cm^3$	
0.0	MLE	3.4(0.6)	3.9(0.4)	4.1(0.5)	3.4(0.5)	3.7(0.5)	3.7(0.6)
0.1	FAST-TLE	3.3 (0.7)	3.0 (0.5)	3.3 (0.6)	3.1 (0.5)	3.6 (0.3)	3.2 (0.6)
	Proposed	**3.1** (**0.7**)	**2.9** (**0.7**)	**3.2** (**0.8**)	**2.9** (**0.4**)	**3.5** (**0.3**)	**3.1** (**0.6**)
0.2	FAST-TLE	**2.3** (**0.4**)	2.4 (0.5)	**2.4** (**0.3**)	2.9 (0.5)	3.1 (0.5)	2.6 (0.5)
	Proposed	2.3 (0.5)	**2.2** (**0.5**)	2.6 (0.4)	**2.7** (**0.6**)	3.1 (0.5)	2.6 (0.5)
0.3	FAST-TLE	2.2 (0.4)	**2.1** (**0.5**)	2.2 (0.5)	**2.0** (**0.5**)	2.4 (0.5)	2.2 (0.5)
	Proposed	**2.2** (**0.2**)	2.2 (0.5)	**2.1** (**0.3**)	**2.7** (**0.7**)	**2.3** (**0.5**)	**2.2** (**0.4**)
0.4	FAST-TLE	3.2 (1.2)	2.6 (1.1)	4.0 (0.9)	2.9 (1.3)	3.1 (0.7)	3.4 (1.2)
	Proposed	**2.2** (**0.4**)	**1.7** (**0.3**)	**1.9** (**0.4**)	**2.2** (**0.7**)	**1.9** (**0.4**)	**2.0** (**0.5**)
0.5	FAST-TLE	4.5 (0.5)	4.5 (1.0)	4.8 (0.7)	4.2 (0.6)	4.2 (0.7)	4.5 (0.7)
	Proposed	**2.1** (**0.3**)	**1.7** (**0.3**)	**2.0** (**0.6**)	**1.7** (**0.5**)	**1.7** (**0.4**)	**1.9** (**0.4**)

The values in bold correspond to the best performing method in terms of mMCR.

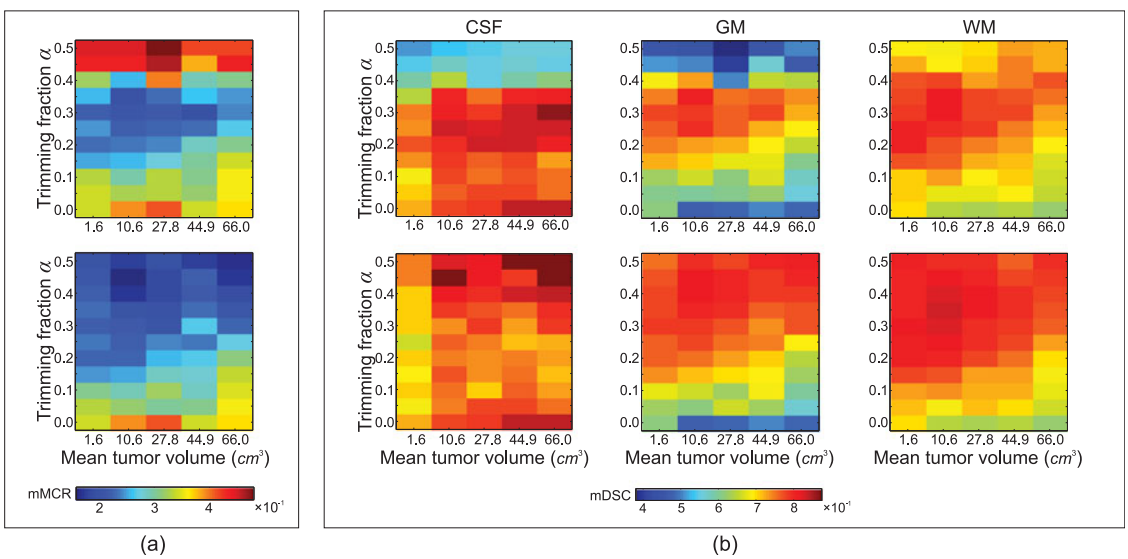


Fig. 7. Performance of the MR intensity mixture estimation and MR image segmentation: (a) median misclassification ratio (mMCR) and (b) median Dice similarity coefficient (mDSC), respectively. Low mMCR corresponds to accurate mixture estimation, while high mDSC corresponds to accurate segmentation of MR images. Results for the FAST-TLE (*top*) and the proposed (*bottom*) method are shown for different mean tumor volumes and trimming fractions α, and for the white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF).

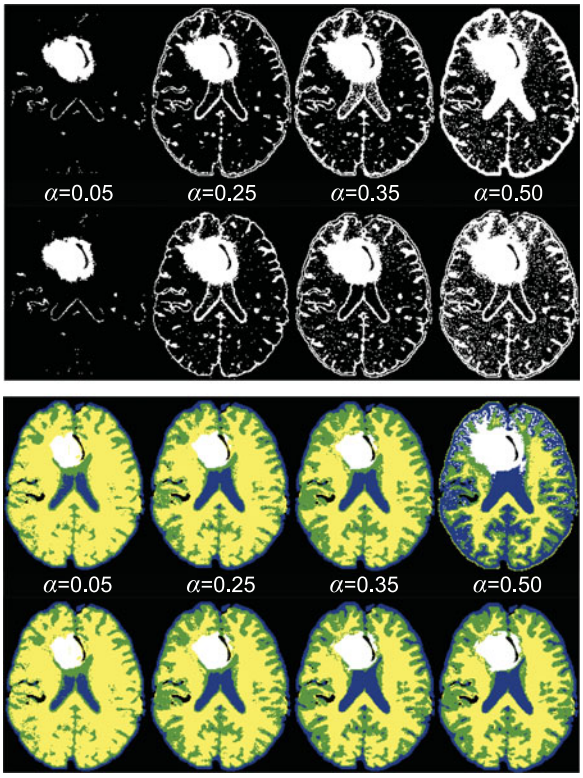


Fig. 8. Analysis of MR images shown in Fig. 6 and their cross-sections: outlier voxels (*top*) and segmentation (*bottom*) of the normal brain structures and tumors for FAST-TLE (*upper rows*) and the proposed (*lower rows*) methods for different trimming fractions α. The color-coding of the segmentations is the same as in Fig. 6.

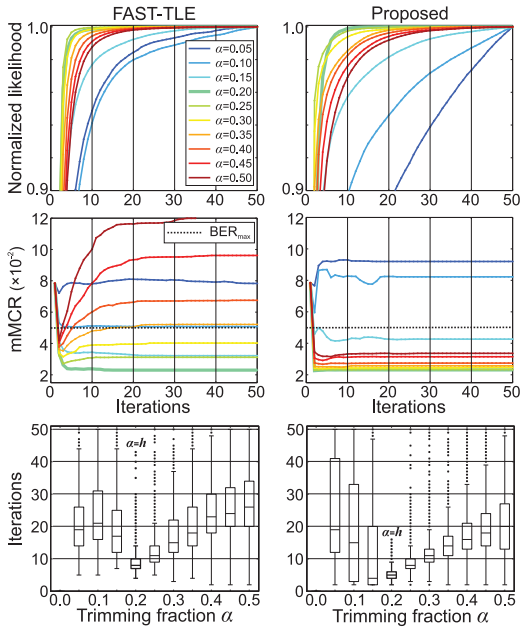


Fig. 9. The normalized mean likelihood curve (*top*), median misclassification ratio (mMCR) (*middle*) w.r.t. iterations, and the number of iterations required to convergence (*bottom*) of the FAST-TLE and the proposed methods for different trimming fractions α computed on synthetic samples contaminated by a fraction $h = 0.2$ of outliers.

fraction h, while the proposed method had low and stable mMCR when $\alpha > h$. As the exact outlier fraction h is not known for the MR image sets, as in data sets of many other applications, the ability to set the trimming fraction α arbitrarily high (up to 0.5) without trading performance has an important advantage for the practical use of mixture estimation methods.

The accuracy of the estimated mixtures has an important impact on the segmentation of normal brain structures. The performance of segmentation reported by mDSC is shown on the right of Fig. 7. The FAST-TLE achieved best mDSC in the range of trimming fraction α from 0.25 to 0.35, while for higher values of α, the mDSC for the CSF and GM structures decreased significantly. The main reason is that the log-likelihood ordering (3) resulted in trimming most of the voxels belonging to the CSF (Fig. 8), while the mixture component corresponding to CSF then modeled the CSF-GM interface and, thus, the resulting segmentation of both the CSF and GM became inaccurate. The proposed method achieved stable mDSC for α in the range from 0.3 to 0.5 and resulted in accurate segmentation of the normal brain structures and the tumors. Fig. 8 shows the segmentations with respect to α obtained by the FAST-TLE and the proposed method.

5 Performance Characteristics

The aim of this section is to investigate and compare the properties of the FAST-TLE and the proposed methods such as convergence and execution times.

The convergence properties of the two tested methods were investigated on the synthetic samples (Section 3). In this experiment 10 different synthetic mixtures were used with a fraction $h = 0.2$ of uniformly distributed outliers; thus, each sample consisted of $1.2 \cdot 10^4$ observations. The FAST-TLE and the proposed methods were tested for trimming fractions $\alpha \in \{0.05, 0.10, \ldots, 0.5\}$ and executed for 50 iterations. All 50 iterations were performed regardless of the stopping criteria, however, the iteration count was recorded when stopping criteria was met. Also, the accuracy of mixture estimation was evaluated in each iteration by computing the mMCR. The experiment was repeated 10 times, each time with a different, randomly selected initial parameters as described in Section 3. The normalized mean likelihood, recorded iteration counts and mMCR curves with respect to the 50 iterations performed for each method are shown in Fig. 9. When α was equal to h (i.e., $\alpha = 0.2$), both methods achieved the highest performance (low mMCR) and were most efficient (low iteration count). In general, the likelihood convergence of the proposed method was slower than that of FAST-TLE, however, the proposed method typically required a lower number of iterations, especially for $\alpha = h$, as the mMCR always stabilized after 20 iterations (Fig. 9). The reason for lower iteration count was also that an additional stopping criterion was incorporated into the proposed method that put an upper bound on the log-likelihood objective function (step 4 in Algorithm 2). As already observed in Section 3.3, the mMCR of the FAST-TLE method increased drastically when the trimming fraction α differed from the outlier fraction h, while the proposed method achieved low and stable mMCR for $\alpha \geq h$.

The execution times of the two tested methods were measured on synthetic samples of different sizes that varied from 10^3 to 10^5. For each sample size, observations were

TABLE 4
Characteristics of the FAST-TLE and Proposed Method, Which Was Tested with Three Implementations of the Confidence Level Ordering

Criterion	Sample size N	FAST-TLE method	Proposed method		
			Mahalanobis distances	Numerical integration	Stochastic quantile density
Mean time [s]	10^3	0.20	0.35	0.22	0.43
	10^4	0.56	3.66	0.50	1.05
	10^5	5.06	61.74	3.74	5.33
Iteration count	10^3	11	5	5	5
	10^4	23	8	9	12
	10^5	43	15	15	22
mMCR*	10^3	0.15	–	0.006	0.017
	10^4	0.14	–	0.002	0.006
	10^5	0.17	–	0.001	0.008

The mMCR of each method is computed w.r.t. the proposed method with Mahalanobis distance based implementation of confidence level ordering. The measurements are means over 10 experiment repetitions, using different outlier and trimming fractions, and different initial mixture parameters (cf. text for details).
**with respect to Mahalanobis distance based implementation*

drawn from three-component normal mixtures and the uniformly distributed outliers were added to the sample in fractions varying from 0 to 0.5 with step of 0.05. The methods were tested for trimming fraction α varying from 0.05 to 0.5 with step of 0.05. The experiment was repeated 10 times, each time with a different, randomly selected initial parameters as described in Section 3. The proposed method was tested with three different implementations of the confidence level ordering (4). The first was based on computing the component-wise Mahalanobis distances, the second was the numerical integration (Algorithm 1) and the third was the stochastic density quantile algorithm [24]. The last two implementations resulted in approximate confidence levels, therefore, their effect on the accuracy of the mixture estimates with respect to the Mahalanobis distance based implementation was measured by mMCR. Besides the execution time, iteration count was also recorded. The results are summarized in Table 4.

For the proposed method with Mahalanobis distance based implementation of the confidence level ordering, the increase in sample size N resulted in a significant increase of the execution time. On the other hand, the numerical integration and stochastic density quantile implementations were very efficient with execution times comparable to the FAST-TLE method. The approximation error introduced by the numerical integration and stochastic density quantile implementation used in the proposed method was below 0.006 and 0.017 in terms of mMCR, respectively. The mMCR between the FAST-TLE and the proposed methods (the Mahalanobis distance based implementation) was much higher, indicating a significant difference between the log-likelihood (3) and the confidence level (4) ordering schemes. Besides, the implementation incorporating numerical integration or stochastic density quantile algorithm can be easily adapted to mixtures of arbitrary forms of component distributions.

6 Discussion

In applications of mixture estimation, samples are always contaminated by the outliers and, more importantly, the a priori estimation of the fraction of outliers in a sample is often difficult, if at all possible in advance. The results of mixture estimation on samples contaminated by outliers indicate that the proposed method performs more robustly than the FAST-TLE method. The two tested methods rely on trimming of the observations to achieve robust mixture estimation, therefore, the trimming fraction is the most critical performance-affecting parameter of these methods. The experiments on synthetic and realistic samples contaminated with outliers revealed that mixture estimates obtained by the FAST-TLE method were accurate only when the value of the trimming fraction was set to a value which was similar to the actual outlier fraction. The proposed method can be applied by setting the trimming fraction to the highest expected value of the outlier fraction in the class of samples considered and, thereby, obtain accurate mixture estimates from the samples contaminated with a fraction of outliers that is either much lower or reaches up to the preset trimming fraction.

One of the advantages of trimming versus a competing strategy, which is to treat outliers as a separate component, is that trimming can capture outliers of varying distributions; e.g., diffuse, concentrated uni-modal, or multi-modal, or even a combination of diffuse and concentrated outlier distributions. This was verified by the results on synthetic samples of normal mixtures with the outliers distributed either uniformly (Section 3) or normally (Supplemental materials, available online), and by results on realistic samples of 100 brain tumor MR image data sets (Section 4). On realistic samples the distribution of outliers was both diffuse and concentrated in multiple modes (Fig. 6 *top right*), hence, trimming has the ability to capture outliers with rather arbitrary distributions.

The benefit of the proposed method is also in its ability to estimate mixtures from samples, in which the components are represented by a different and highly unbalanced number of observations. Such samples are often obtained from high density data sensors such as imaging devices, in which structures of different sizes need to be compactly represented by a mixture model. In medical image analysis, for example, the mixture models are often used to represent the intensity model of anatomical structures. The intensity

models are typically used for the tasks like image segmentation, registration, and intensity nonuniformity correction. We performed experiments on 100 sets of MR images of brains with tumors of varying volumes, in which the task was to model the intensities of the normal brain structures structures such as white matter, gray matter and cerebrospinal fluid. These structures differed significantly in volume and, therefore, the number of voxels representing the intensity of each structure also varied significantly. For example, the mean fractions of the normal brain structures, i.e., WM, GM and CSF, were 0.27, 0.49 and 0.24, respectively. The outlying intensities represented the tumor intensities and the intensities of its surrounding structures (tumor infiltration) and common MR image artifacts such as partial volume effect, Rician noise, and intensity inhomogeneity. Therefore, neither the form nor the fraction of outliers could be accurately predicted in advance and an accurate estimation of the MR intensity model in the form of a mixture is a challenging problem. Robust mixture estimation methods such as FAST-TLE have already been applied to MR image segmentation [20], [33], [34], however, as demonstrated by our experiments, their optimal application to MR image sets requires case-specific tuning of the trimming fraction value. The proposed method demonstrated stable performance of mixture estimation over a broad range of trimming fraction values (0.3-0.5) and outperformed the FAST-TLE method in segmentation of normal brain structures in the presence of tumors of varying volumes.

For the task of image segmentation, a spatial association between voxels is an important prior assumption that can be incorporated through statistical image modeling based on Markov random field (MRF) framework [35]. In MRF frameworks, the components of mixture estimates represent the likelihood of structures used in a Bayesian inference. We used the mixture estimates obtained by the FAST-TLE and the proposed method in the MRF framework, which generally improved the DSC of resulting segmentations, however, the patterns of DSC with respect to tumor volume and trimming fraction α were similar to those in Fig. 7b. Hence, we arrived at the same conclusion as previously without employing the spatial association of voxels, which is, the mixture estimates obtained by the proposed method, compared to the FAST-TLE, result in consistent DSC values for a larger range of trimming fraction α (0.3–0.5) and, thus, in more robust mixture estimates.

The robustness of mixture estimation depends on the way mixture parameters are initialized. To analyze this dependence, the experiments on the synthetic samples were also conducted with the mixture parameters initialized by k-means clustering and by ground truth mixture parameters. Although the final results varied slightly based on the initialization used, the patterns of mMCR, and thereby the conclusions, for both the FAST-TLE and the proposed methods were similar to those in Fig. 4. To improve the robustness of mixture estimates a widely used solution is to take several (random) initial mixture parameters, compute the respective mixture estimates and then select the mixture corresponding to the highest objective function [2], [9], [23]. Performing several mixture estimations may not be feasible for large samples due to increased execution time, while for samples of unbalanced mixtures and/or high value of trimming fraction a drastically higher number of the initial mixture parameters would be required for robust mixture estimation [2]. In practical applications, however, the mixture parameters can usually be initialized based on application-specific prior knowledge, as demonstrated on brain MR images [20], [36].

Although the FAST-TLE and the proposed methods do not explicitly model the outlier distributions, they assign, through the normalized likelihood or confidence level based ordering of observations, respectively, an outlier membership value to each of the observations. The proposed ordering based on confidence levels assigns a weight in range from zero to one to each observation, which can be used as a fuzzy outlier membership map in advanced image segmentation frameworks that account for spatial association of observations [35].

In conclusion, we proposed a novel method for robust estimation of unbalanced mixtures on samples contaminated with outliers. The main advantage of the proposed method is its robustness because it can estimate unbalanced mixture models irrespective of the fraction of outliers, as long as the trimming fraction is higher than the outlier fraction, which can be as high as 50 percent. The breakthrough was achieved by selecting the outliers based on confidence level ordering, which also enabled the estimation of highly unbalanced mixtures. The results on synthetic and realistic samples indicated that the proposed method was robust over a broad range of trimming fractions, therefore, it can be applied for the mixture estimation on real-world samples, in which the outlier fraction cannot be estimated in advance.

Acknowledgments

This research was supported by the Ministry of Education, Science and Sport, Republic of Slovenia, under grants J2-5473, L2-5472, and L2-4072. A. Galimzianova is the corresponding author.

References

[1] G. McLachlan and D. Peel, *Finite Mixture Models* (series Wiley series in probability and statistics). Hoboken, NJ, USA: Wiley, 2005.

[2] A. S. Hadi and A. Luceño, "Maximum trimmed likelihood estimators: A unified approach, examples, and algorithms," *Comput. Statist. Data Anal.*, vol. 25, no. 3, pp. 251–272, 1997.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B Methodol.*, vol. 39, no. 1, pp. 1–138, 1977.

[4] J. Ma and S. Fu, "On the correct convergence of the EM algorithm for Gaussian mixtures," *Pattern Recognit.*, vol. 38, no. 12, pp. 2602–2611, 2005.

[5] I. Naim and D. Gildea, "Convergence of the EM algorithm for Gaussian mixtures with unbalanced mixing coefficients," in *Proc. 29th Int. Conf. Mach. Learn.*, 2012, pp. 1655–1662.

[6] N. M. Neykov and C. H. Müller, "Breakdown point and computation of trimmed likelihood estimators in generalized linear models," in *Developments Robust Statistics*. Heidelberg, Germany. Physica-Verlag, 2003, pp. 277–286.

[7] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods* (series Wiley series in probability and statistics). Hoboken, NJ, USA: Wiley, 2006.

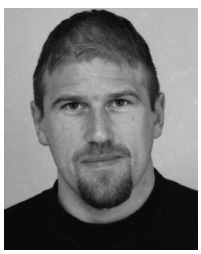
[8] G. J. McLachlan, S.-K. Ng, and R. Bean, "Robust cluster analysis via mixture models," *Austrian J. Statist.*, vol. 35, p. 157–174, 2006.

[9] N. Neykov, P. Filzmoser, R. Dimova, and P. Neytchev, "Robust fitting of mixtures using the trimmed likelihood estimator," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 299–308, 2007.

[10] M. T. Gallegos and G. Ritter, "Trimming algorithms for clustering contaminated grouped data and their robustness," *Adv. Data Anal. Classification*, vol. 3, no. 2, pp. 135–167, 2009.
[11] M. T. Gallegos and G. Ritter, "A robust method for cluster analysis," *Ann. Statist.*, vol. 33, no. 1, pp. 347–380, 2005.
[12] R. Browne, P. McNicholas, and M. Sparling, "Model-based learning using a mixture of mixtures of Gaussian and uniform distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 814–817, Apr. 2012.
[13] S. X. Lee and G. J. McLachlan, "Model-based clustering and classification with non-normal mixture distributions," *Statistical Methods Appl.*, vol. 22, no. 4, pp. 427–454, 2013.
[14] B. Franczak, R. Browne, and P. McNicholas, "Mixtures of shifted asymmetric Laplace distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1149–1157, Jun. 2013.
[15] W. Yao, Y. Wei, and C. Yu, "Robust mixture regression using the *t*-distribution," *Comput. Statist. Data Anal.*, vol. 71, pp. 116–127, 2014.
[16] J. A. Cuesta-Albertos, C. Matrán, and A. Mayo-Iscar, "Robust estimation in the normal mixture model based on robust clustering," *J. Roy. Statistical. Soc. B*, vol. 70, no. 4, pp. 779–802, 2008.
[17] M. T. Gallegos and G. Ritter, "Trimmed ML estimation of contaminated mixtures," *Sankhyā : The Indian J. Statist.*, vol. 71-A, pp. 164–220, 2009.
[18] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
[19] L. A. García-Escudero, A. Gordaliza, C. Matrán, and A. Mayo-Iscar, "A general trimming approach to robust cluster analysis," *Ann. Statist.*, vol. 36, no. 3, pp. 1324–1345, 2008.
[20] D. García-Lorenzo, S. Prima, D. Arnold, D. Collins, and C. Barillot, "Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence MRI for multiple sclerosis," *IEEE Trans. Med. Imag.*, vol. 30, no. 8, pp. 1455–1467, Aug. 2011.
[21] N. M. Neykov, P. Filzmoser, and P. N. Neytchev, "Robust joint modeling of mean and dispersion through trimming," *Comput. Statist. Data Anal.*, vol. 56, no. 1, pp. 34–48, 2012.
[22] A. Galimzianova, Ž. Špiclin, B. Likar, and F. Pernuš, "Robust mixture-parameter estimation for unsupervised segmentation of brain MR images," in *Proc. 3rd Int. Workshop Med. Comput. Vis. Large Data Med. Imag.*, 2014, pp. 84–94.
[23] L. A. García-Escudero, "A review of robust clustering methods," *Adv. Data Anal. Classification*, vol. 4, no. 2, pp. 89–109, 2010.
[24] R. J. Hyndman, "Computing and graphing highest density regions," *Am. Statist.*, vol. 50, no. 2, pp. 120–126, 1996.
[25] J. Tohka, E. Krestyannikov, I. Dinov, A. Graham, D. Shattuck, U. Ruotsalainen, and A. Toga, "Genetic algorithms for finite mixture model based voxel classification in neuroimaging," *IEEE Trans. Med. Imag.*, vol. 26, no. 5, pp. 696–711, May 2007.
[26] I. Wells, W. Grimson, R. Kikinis, and F. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 15, no. 4, pp. 429–442, Aug. 1996.
[27] M. Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J.-P. Thiran, "Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images," *IEEE Trans. Med. Imag.*, vol. 24, no. 12, pp. 1548–1565, Dec. 2005.
[28] B. Aubert-Broche, M. Griffin, G. B. Pike, A. C. Evans, and D. L. Collins, "Twenty new digital brain phantoms for creation of validation image data bases," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1410–1416, Nov. 2006.
[29] M. Prastawa, E. Bullitt, and G. Gerig, "Simulation of brain tumors in MR images for evaluation of segmentation efficacy," *Med. Image Anal.*, vol. 13, no. 2, pp. 297–311, 2009.
[30] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[31] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, no. 6, pp. 910–914, 1995.
[32] U. Vovk, F. Pernuš, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE Trans. Med. Imag.*, vol. 26, no. 3, pp. 405–421, Mar. 2007.
[33] S. Bricq, C. Collet, and J. Armspach. (2008). MS lesion segmentation based on hidden Markov chains. *Proc. Grand Challenge Workshop: Multiple Sclerosis Lesion Segmentation Challenge*, pp. 1–9. [Online]. Available: http://hdl.handle.net/10380/1450
[34] X. Tomas-Fernandez and S. Warfield, "Population intensity outliers or a new model for brain WM abnormalities," in *Proc. 9th IEEE Int. Symp. Biomed. Imag.*, 2012, pp. 1543–1546.
[35] D. García-Lorenzo, J. Lecoeur, D. L. Arnold, D. L. Collins, and C. Barillot, "Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts," *Med. Image Comput. Comput.-Assisted Intervention*, vol. 12, pp. 584–591, 2009.
[36] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.

Alfiia Galimzianova received the BSc degree and is working towards the PhD degree at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. Her research interests include robust statistics, machine learning, computer vision, and biomedical image analysis.

Franjo Pernuš received the PhD degree and is a professor at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. He is the head of the Laboratory of Imaging Technologies and his research interests involve biomedical image processing and analysis, computer vision, and their application to various biomedical and industrial problems. He is a co-founder of a high-tech company Sensum, which supplies machine vision solutions for the pharmaceutical industry.



Boštjan Likar received the PhD degree and is a professor at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. He is a member of the Laboratory of Imaging Technologies and his research interests include visual quality inspection, computer and machine vision systems, biomedical image processing, and hyperspectral imaging. He is a co-founder of a high-tech company Sensum, which supplies machine vision solutions for the pharmaceutical industry.

Žiga Špiclin received the PhD degree and is a research fellow at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. His research interests include the development of image registration, restoration and reconstruction techniques for computer vision and biomedical applications.