page 7, line 43 to page 8, line 39 (Sect. 3.1) This section of the text was partly re-written to correct some erroneous notations and few sentences were added to better explain how the method was applied. We emphasize that our aim has been a more precise description of the used algorithm, not any change to it. Please replace the old text with the corrected text below (the corresponding latex version of the text to be replaced in the manuscript can be found in a separate attached latex file).

" We denote the set of all sites in I by S, a single site (i.e. pixel location determined by its row and column indices) by $s = s(i, j)$, with $i = \{1, \ldots, M\}$ and $j = \{1, \ldots, N\}$ and the set of labels by $L = \{L_1, \ldots, L_K\}$. The configuration of the labels is denoted by Λ. After the initial segmentation we have K classes (labels) with the class-wise means $\mu = \{\mu_1, \ldots, \mu_K\}$ and variances $\sigma^2 = \{\sigma_1^2, \ldots, \sigma_K^2\}$. These means and variances remain fixed after the initial segmentation. We adopt a first-order neighbourhood system ∂s where each site has four neighbouring sites, two in the horizontal direction and two in the vertical direction. We use a pairwise clique system. If s and r belong to the same clique of the site s, i.e. $c(s) = \{s, r\}$ then s must belong to the neighbourhood ∂r and r must belong to the neighbourhood ∂s. Hence r and s are neighboring pixels. In the first order neighbourhood each site has four cliques. The potential function associated with the clique $c(s)$ is denoted by $V_{c(s)}$ or generically V_c. The pixel value in s is denoted by f_s and $f = \{f_s | s \in S\}$. Without a subscript Λ refers to a label configuration in S. We denote by Λ_S the set of all the configurations.

Assume that we at some iteration have the label configuration Λ^* in S. In the next iteration, selecting the best new label $\hat{L}_s$ for the site s, given f and $\Lambda^*_{S\setminus s}$, is equivalent to maximizing the probability distribution of labels in s, conditioned by f_s and the current label configuration in the neigbourhood $\Lambda^*_{\partial s}$. This is possible by utilizing the Markov property in MRF (Besag 1974). The selection of the best new label for the site s can be written as:

$$\hat{L}_s^{MAP} = \mathrm{argmax}_{\mathrm{L_s} \in \mathrm{L}} \mathrm{P}(\mathrm{L_s} | \mathrm{f_s}, \Lambda^*_{\partial \mathrm{s}}), \tag{1}$$

The right hand side of Eq. 1 can be written as the product:

$$g(f_s | L_s) \exp^{-\sum_{r \in \partial s} V_{c(s,r)}} \tag{2}$$

where the first term g is the likelihood function and the second term is the potential function.

For pairwise cliques the potential function $V_c(s, r)$ is reduced to two states:

$$V_{c(s,r)}(L) = \beta \gamma(L_s, L_r), \tag{3}$$

where

$$V_{c(s,r)}(L) = \begin{cases} +\beta & if\ L_s = L_r\ ; \\ -\beta & if\ L_s \neq L_r. \end{cases} \tag{4}$$

The homogeneity of the region is controlled by the parameter $\beta (> 0)$.

We assume that g has a Gaussian distribution with the class-wise mean μ_{L_s} and variance $\sigma^2_{L_s}$, i.e. in the Gibbs form it is

$$\begin{aligned} g(f_s | L_s) &= Z_s^{-1} \times \exp^{-U(f_s|L_s)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_{L_s}} \exp\left(\frac{-(f_s - \mu_{L_s})^2}{2\sigma^2_{L_s}} \right) \end{aligned} \tag{5}$$

where Z_s is a normalizing constant and $U(f_s|L_s)$ is the likelihood energy.

At site s we compute the local energy $U_s(L)$, i.e. the logarithm of the product in Eq. (2), as:

$$-U_s(L) = \log(\sqrt{2\pi}\sigma_{L_s}) + \frac{(f_s - \mu_{L_s})^2}{2\sigma^2_{L_s}} + \sum_{r \in \partial s} \beta \gamma(L_s, L_r), \tag{6}$$

Maximizing the product in Eq. (2) over L yields the new label $\hat{L}_s$. This maximization is equivalent to minimizing $U_s(L)$.

In a similar manner we obtain the best new labelling $\hat{\Lambda}$ for the whole image by solving the local minimization of $U_s(L)$ for every $s \in S$. So the global minimum $U(\Lambda)$ is achieved by local computations. This procedure results in the MAP estimate for $\hat{\Lambda}$:

$$\hat{\Lambda}^{MAP} = \text{argmax}_{\Lambda \in \Lambda_S} P(\Lambda|f) = \text{argmin}_{\Lambda \in \Lambda_S} U(\Lambda). \tag{7}$$

These kind of functions can be optimized by various methods, one being the simulated annealing method (Kirkpatrick et al., 1983) (Cerny, 1985), where a slow decrease in the probability of accepting worse solutions occurs as the algorithm searches the solution space. The method used here is an adaptation of the Metropolis-Hastings algorithm introduced in (Metropolis et al. 1953). In the algorithm the labelling is also dependent on the control variable called temperature, T, the value of which decreases as the iteration proceeds. We denote the proposed new label by L^*. If the value of the energy function $U(L^*)$ decreases, L^* is accepted always. If the value of $U(L^*)$ increases, the label is accepted with probability $\exp(-\Delta U/T)$, where ΔU is the energy difference between the new and old configuration. ”